



African Journal of Biological Sciences



An Optimized CNN Model for Lung Cancer Detection

Sulekha Das¹, *Dr. Avijit Kumar Chaudhuri², Dr. Partha Ghosh³, Swagato Sikdar⁴

¹Research Scholar, Information Technology, GCECT, Kolkata, West Bengal, India

²Associate Professor, Computer Science and Engineering, Brainware University, Barasat, West Bengal, India

³Assistant Professor Computer Science and Engineering, GCECT, Kolkata, West Bengal, India

⁴Student –CSE-AIML(2nd year), Techno Engineering College Banipur

shu7773sea@gmail.com, c.avijit@gmail.com, parth_ghos@rediffmail.com, sikdarswagato0@gmail.com

Abstract

This research study regarding the applications of Convolutional Neural Networks (CNNs) based on medical imaging data for lung cancer diagnosis is presented here. People are usually diagnosed with lung cancer at advanced stages as the initial symptoms are not noticeable. It is critical to detect the cancer in the early stages because this improves the patients' outcomes. The CNN is perfect for the mentioned tasks within medical image analysis for its feature learning capabilities - i.e., its ability to learn information automatically from raw data. Lung cancer is a kind of tumor in cases of the lung cells beginning from the lungs' air passages, including the cells aligning the air passageways. It accounts for a large portion of global cancer-related deaths. In India, Lung cancers leave behind all the other cancer types to become the leading cause of mortality. The research in this study is made up of the positive roles and efficiency of multi-modal imaging in lung cancer detection and its classification process by employing CNN techniques. This research identifies how the positioning of an effective synergy between imaging modalities and updated CNN algorithms can increase diagnostic accuracy, sensitivity, specificity, Area Under the Curve, and the kappa score.

Keywords: Convolutional Neural Network (CNN), Lung Cancer, Kappa score, Area Under the Curve(AUC)

Article History

Volume 6, Issue 5, 2024

Received: 15 May 2024

Accepted: 22 May 2024

doi:10.33472/AFJBS.6.5.2024.7143-7175

Introduction

Among all diagnosed cancers, lung cancer accounts for the largest number of cases. This type of cancer has the highest incidence of cancer-related deaths around the globe. According to GLOBOCAN 2020, lung cancer was the cause of nearly 11% of all new cases (2.2 million) and 18% of all deaths (1.8 million) worldwide in the year 2020 (Cao et al., 2020). Lung cancer has been identified as one of the serious global public health problems, and it is the second to cancer death. It emerges as the uncontrolled exponential expansion of mutated cells in the lungs most associated with cigarette smoke pollution. Lung cancer has the highest death rate found among all the types of cancer, which is made worse by the fact that it is difficult for symptoms to be recognized, especially in the early stages. Detection in a timely fashion is essential for quick and accurate treatment that leads to better health of patients. In the first stages, lasting symptoms of lung cancer are either nonexistent or too subtle to be acknowledged, and this delays the diagnostic process. Coughing and lack of breath are the most noticeable lung cancer symptoms. They, however, can very easily overlap other respiratory conditions' symptomatic presentations, thus making it quite difficult to see this disease to be the source of a particular symptom. The diagnosis of cancers mainly is dependent on biopsies, which can be cumbersome and risky, thus leading to dubiety from both patients and their doctors. Sometimes the imaging technologies could make a false positive diagnosis and thus result in patients receiving unnecessary therapies, causing more anxiety to the patient.

Combinations of multi-modal imaging, which have e.g. computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI), with CNN approaches offer high potential for tackling these problems. Multi-dimensional imaging gives a composite view using structural and functional data (structural and functional data fusion). To this end, CNNs can be utilized to relieve the processing of diverse information and, as a consequence, impart deep comprehension of lung cancer characteristics. CNNs are capable of performing better when trained on combined data, such as training the CNNs with images along with other data, so that specificity (avoiding false positives) and sensitivity (detecting true positives) are optimized, which eventually, leads to more accurate and reliable lung cancer detection as most of the false positives and false negatives are avoided. The model can thus use both spatial and general data to accomplish more precise spatial localization lesions, resulting in better

identification of tumors. Integration of multi-model imaging and CNN algorithms is a powerful tool for the early detection of the subtle changes that may signal imminent disease progression. Therefore, timely intervention is possible which may completely reverse the patient's condition. CNNs are excellent in the recognition of image data, as they can extract and analyze fine-tuning details of image features. The application of AI algorithms to multi-modal images portends an evolution of methods after the first one that may more accurately detect disorder patterns and nuances, which normally might be difficult. The multi-modal image acquisition along with the novel CNNs has presented a promising approach that affords easier lung cancer detection. These dual modalities can in combination not only make diagnostics more accurate, more sensitive, and more reliable but also allow for the identification of problems at an early stage and decision for personalized therapy approach thus improving patient outcomes.

In this article, the authors emphasize the fact that MLT has the power to work together with multimodal medical imaging to be able to detect and classify lung cancer effectively. This interdisciplinary approach which offers several advantages makes the diagnosis and treatment more accurate and personalized. The fusion of several imaging methods leads to a deeper understanding of the lungs, and thus cancer detection accuracy is raised. This is one of the features that allow the tool to have a more reliable outcome, eliminating false positives and negatives. Multi-modal image data enables pathologists to make a precise diagnosis and provide suitable treatment. Specific imaging modalities have the unique properties of revealing a tumor's physical characteristics, like its size and shape, and its metabolic activity as well, enabling a comprehensive cancer picture. Machine Learning algorithms can exploit richer information from multi-view images, hence allowing for more sophisticated analysis of lung cancer features. Thus, such an approach enables early detection of indicative cancer stages either through the recognition of subtle patterns or through the detection of the alteration of such patterns. Through AI analysis of multimodal lung pictures, lung cancer diagnosis can happen earlier. Early diagnosis is so valuable because it enables starting the treatment earlier, which consequently allows treatment to be more effective and survival rates higher. A reliable classification of the tumors will help to create a more individual approach to treatment. MLTs can help to spot biomarkers and genetic factors to ensure the choice of a single patient's precise treatment that will be based on his or her specific needs. With an overlapping context of findings from different imaging modalities the ambiguity of the diagnosis decreases. This is especially helpful when dealing with tricky cases whose features may be uncertain on one imaging modal. The

integration of cutting-edge imaging and artificial intelligence holds the place for catalyzing revolutionary diagnoses of lung cancer. This could be the beginning of the change in the clinical diagnosis, treatment plan, and monitoring paradigm in patients with lung cancer. Continued advancements in technology, coupled with ongoing research and collaboration between medical and AI experts, will further refine and enhance the capabilities of this multimodal imaging approach. The ultimate goal of this research work is to improve patient care, offering more effective and personalized strategies for managing lung cancer.

Relevant Literature

Studies in the medical literature on lung cancer diagnosis through deep learning techniques are many that uncover the applications of CNN to improve the accuracy and efficiency of lung cancer prevention and prognosis. Below is a detailed description of the main studies and techniques used in lung cancer detection that are combined with CNN technology.

Slatore et al., 2014 published one review article in the *Journal of Thoracic Oncology*. The review outlines the role played by CT scans in the detection of lung cancer, which is said to be more sensitive than traditional radiographs. It describes the difficulties, for example, false alarms and radiation exposure.

Korevaar et al., 2016 conducted a meta-analysis to evaluate the diagnostic performance of bronchoscopy and endobronchial ultrasound for lung cancer staging. It gives us an idea of the aspects that can be adopted and the areas that may need improvement in such invasive procedures.

Nie et al., 2016 published one article in *Image Computing and Computer-Assisted Intervention—MICCAI 2016: At the 19th International Conference in Athens in 2016*. Despite the fact that they are used for a brain tumour, including glioma analysis, such research illustrates the facility of using 3D CNNs for multi-modal imaging and survival time prediction. It demonstrates the possibility of applying a similar method to diagnose lung cancer patients in the future as well.

Yan et al., 2018 present DeepLesion, a dataset supporting the training of deep learning models encompassing CNNs for lesion detection. It is highlighted that the deep learning approach can be used in the image detection of lung cancer lesions, among other medical images. Rajpurkar et al., 2018 presented CheXNeXt, a new family of deep learning algorithms based on

the CNN approach, for automated chest radiograph diagnosis. This experiment proves the applicability of CNNs by achieving high accuracy and also showing the usefulness of those networks in tasks concerning large-scale screenings. Liu et al., 2019, through their study, give a summary of CNN architecture variants being employed for computer-aided diagnosis, such as lung cancer. It examines the role of dataset characteristics and transfer learning in training CNNs for better performance in medical image analysis tasks.

Zhang et al., 2019 in their investigative study, employed the use of CNNs to predict EGFR mutation status in lung adenocarcinoma using PET images. It shows the possibility of applying this technique to skilled labour systems.

Rivera et al., 2020 performed a large-scale study using the chest X-ray records dataset to evaluate the diagnostic accuracy of chest X-rays for lung cancer detection. It thus points to the insufficiency of traditional X-rays as a diagnostic aid and underlines the requirement for more sophisticated imaging modalities.

Pantanowitz et al., 2022, reviewed how AI based CNN was used in pathology, and those data were used to make cancer diagnoses. It is elucidated here how those techniques of deep learning are interwoven with histopathological analysis for the sake of lung cancer detection.

Raymahapatra et al., 2023 developed their research article on the application of CNNs to the classification of brain tumour data for the prediction of two classes of accuracy: sparse categorical accuracy and classification accuracy. The CNN model shows a promising performance, with an accuracy of 97% in classification and 99.9% in categorical accuracy prediction. The authors met this classification accuracy by fine-tuning the parameters of the optimizer and activation functions. The study demonstrated that an increase in the hidden layer number and those specific changes to the activation function improved accuracy. This architectural change was the main contribution to the article. Comparative analysis of lung cancer prediction using CNNs involves the assessment and comparison of various different CNN architectures, training methods, and datasets for evaluation of their accuracy in predicting lung cancer from imaging data like chest X-rays or CT scans. Table 1 below is a generic structure for performing a comparative study of lung cancer CT scan image datasets.

Table 1. Summary of recent work that has been performed in lung cancer detection and in this research using machine learning and deep learning algorithms

Author	Feature/methods	Performance(%)
Taher et al., 2011	HNN	Accuracy: 98 Sensitivity : 83 Specificity : 99
Makaju et al., 2018	SVM	Accuracy: 92 Sensitivity : 100 Specificity : 50
Tekade et al., 2018	CNN	Accuracy: 95.66
Alam et al., 2018	Multiclass SVM	Accuracy (detection) : 97 Accuracy (prediction) : 87
de Carvalho et al., 2018	CNNs	Accuracy : 92.6
Rodrigues et al., 2018	MLP, k-NN, SVM	Accuracy : 96.7
Ausawalaithong et al., 2018	DenseNet	Accuracy : 74.4
Nasser et al., 2019	ANN	Accuracy: 96.67
Radhika et al., 2019	SVM,LR,DT,Naive Bayes	Accuracy: 99.2
Shakeel et al., 2019	ANN	Accuracy: 99.7

Bhatia et al., 2019	XGBoost and RF	Accuracy: 99.7
Bhatia et al., 2019	XGBoost and Random Forest	Accuracy: 84
Toğaçar et al., 2020	LR,LDA,DT,SVM,KNN,	Accuracy: 98.74 Sensitivity : 98.35 Specificity : 99.12 Precision : 99.12 F1-Score : 98.74
Shin et al., 2020	ResNet	Accuracy: 95
Rehman et al., 2021	SVM,KNN	Accuracy: 93 Sensitivity: 86 Specificity : 95.4
Masud et al., 2021	CNNs	Accuracy: 96.3
Abdullah et al., 2021	CNN, SVM, k-NN	Accuracy: 95.5
Naseer et al., 2022	CNN AlexNet architecture with the SGD optimizer	Accuracy: 97.25
Raymahapatra et al., 2023	CNN	Accuracy: 94 AUC: 94
Proposed architecture	CNN	Accuracy: 100 Recall: 100 AUC: 100 Specificity – 100

		Kappa Score - 100
--	--	--------------------------

Convolutional Neural Network(CNN)

CNN is a type of deep neural network commonly utilized in image and video recognition, as well as a range of other tasks. CNN is a class of deep learning algorithms that has the potential to learn multi-level representations of data elements using multiple network layers. The main reason behind CNN is its ability to learn automatically the features needed from raw data, especially images or video. This happens when the kernel or filter slides across the feature map, covering each area progressively. Through convolution, the network includes a range of filters or kernels, that are used to process the input image. Using these filters, data can be structured and arranged in such a way that the initial layer focuses on the local information and the deeper layer deals with the abstract features. Then the pooling layers are considered for the output of the convolutional layers. It has the effect of shrinking the spatial dimensions of the features by reducing the computation, which makes the network computationally efficient and also capable of generalizing to unknown images. The outcome of the convolutional and pooling layers is flattened to a vector, which is then fed into the fully connected layer. This layer allows the network to make predictions as accurate as they are based on learned features.

CNN in Cancer Detection:

One of the most prominent AI technologies that has demonstrated efficiency is CNN. The applications include, but are not limited to, medical image analysis, and cancer detection. Numerous studies and research papers have shown the capability of CNNs for the detection and diagnosis of various types of cancers.

The research work of He et al., 2016 was a breakthrough article that for the first time introduced the concept of residual learning, which has found its application in various tasks, such as breast cancer diagnosis. Researchers have extended this thought and created models that are more effective and accurate at the classification of mammograms.

The International Skin Imaging Collaboration (ISIC) 2017 competition was one of the initiatives to promote the automatic analysis of skin lesions using CNN-based models. In their research, Codella et al., 2018 demonstrate some of the top approaches that have been extensively utilized in the field as well as the challenges that remain unresolved.

Liu et al., 2017 presented their new deep learning algorithm, XmasNet, which uses CNNs, for the classification of prostate cancer lesions from the 3D multiparametric MRI datasets given in the PROSTATEx challenge. This study signifies the outstanding potential of deep learning in improving cancer imaging.

Pandian et al., (2022) describe a method to identify unusual changes in the lung tissue using a very accurate tool, and a high detection probability is ensured. The model that has been suggested is based on the CNN and GoogleNet deep learning algorithms, which use the VGG-16 architecture as both region-proposal and classifier networks.

CNNs have shown very good performance in many computer vision applications, including image classification, object detection, and segmentation. Through the use of automatically learned hierarchical features for learning, their suitability for analyzing visual data is enhanced. It should be highlighted that CNNs have played an important role in the triumph of deep learning in image processing and image-related tasks. The developed model of the brain's visual processing system serves as the basis for its architecture, making it an excellent tool for tasks that involve spatial hierarchies and local patterns in data.

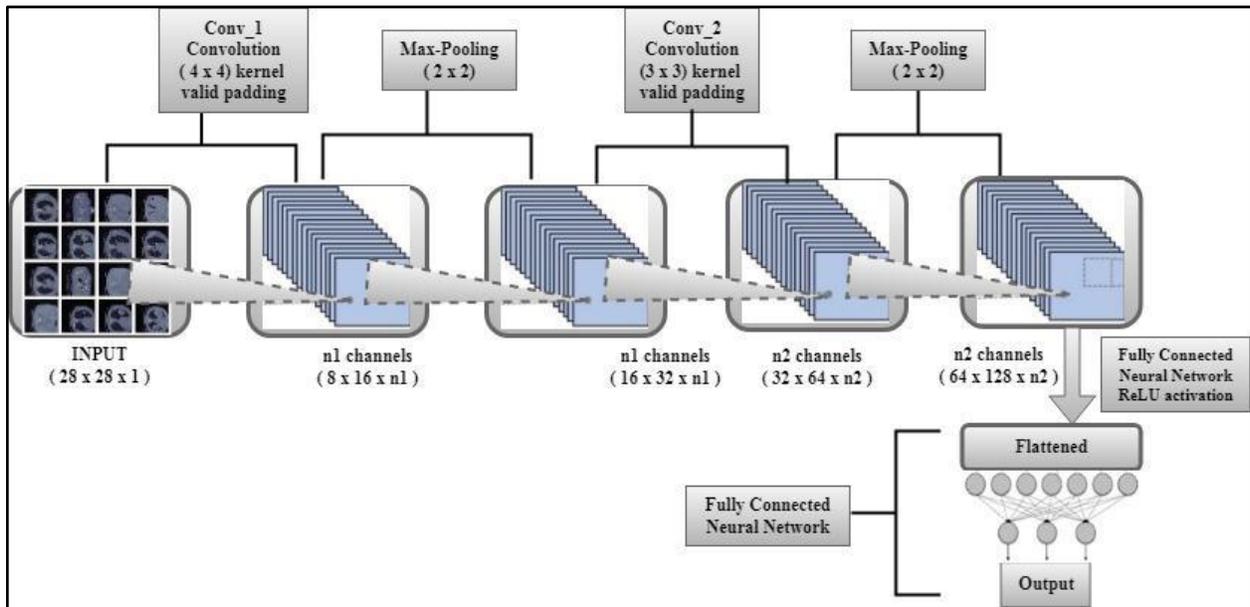


Fig. 1: Proposed CNN model for lung cancer detection

Working of the proposed CNN model

The architecture of the experimental layout of this research and model construction is presented in Fig. 1. A CNN in PyTorch is formulated from constituent layers which include convolutional, pooling, and fully connected layers, to build the model structure. These levels are necessary to build up hierarchical representations of features from input data considerations, especially for tasks like image classification. The process consists of creating an initial Python class consisting of the CNN model. All these layers are set as an initial layer where parameters are defined, such as the kernel size, the number of filters, and activation functions. For instance, the authors start with four convolution layers, which are followed by pooling layers to down sample the feature map and simply connected layers for classification. After the creation of these layers, they are passed into the forward function of the class one by one, which produces the outputs of the specific network. It is a key part of the network where data undergoes convolutional operations, followed by pooling, and then steps through fully connected layers ultimately. With this connected architecture, information flows in a predefined order during both prediction and training (backward pass) to facilitate the efficient propagation of information through the network.

The model is trained with image datasets after constructing architecture. This is done by feeding the input images through the network one after another, computing loss between predicted and actual labels, and then using techniques like backpropagation and optimization algorithms (like Stochastic Gradient Descent, Adam) to update network parameters iteratively. There is an automatic learning functionality enabled that allows the convolutional and pooling layers to learn effective feature representations from the input images. This is done through the process of optimization of parameters. One of the biggest

strengths of CNNs is their ability to learn detailed features directly from data making them suitable for achieving outstanding performance across several computer vision tasks. There were several strategies that the authors used to improve the performance of their model. Among these were adding more layers to the architecture, training the model for a greater number of epochs to develop the embeddings further, and fine-tuning hyperparameters like learning rate and regularization strength through methods such as grid search or random search. By repeating this process by experimenting and improving model architecture and training procedures, the authors improve the CNN's performance which results in more accuracy and robustness of the given vision tasks.

Importance of Kappa Statistics in cancer research

The Kappa value, or Cohen's Kappa coefficient, is a statistical index of inter-rater agreement for categorical issues. It mainly serves to assess the level of agreement between two or more raters or methods in the diverse areas, such as medical research. One of the key values of Kappa score can be related to cancer diagnosis and assessment of reliability among different diagnostic approaches or the agreement between health care providers.

The kappa statistic, which takes into account chance agreement, is defined as: $(\text{observed agreement} - \text{expected agreement}) / (1 - \text{expected agreement})$.

A high Kappa score means that a high level of agreement is higher than would be expected due to chance, while a low Kappa score indicates a poor agreement rate. Cancer diagnosis is a matter of high accuracy (to ensure a proper prognosis) as well as of reproducibility (with potentially life-or-death implications for patients). Here are some key points regarding the importance of Kappa scores in cancer diagnosis, supported by relevant literature:

Pathology that is crucial and includes the diagnosis of cancer via the visual examination of the tissue sample can be an example of interobserver agreement. Kappa scores help to determine the reliability of diagnoses made by pathologists. With such technology application in the healthcare area, clinical decision support systems could be a valuable tool in cancer diagnosis. The Kappa statistic, which measures the concordance among automated systems and human experts, can be applied for this purpose. Warren Burhenne et al., 2000 examine the correspondence between computer-aided detection and human perception in mammography. There were 67% of breast cancers detected by screening mammograms that were visible in the earlier mammograms. A blinded review of radiologist panels, with which 27% of cases were statistically evaluated

between the CAD system and radiologists, showed that the system correctly marked 77% of these cases. The sensitivity of the attending radiologists before installing the CAD was 79%, and the installation of the software did not result in a statistically significant increase in their recall rate. The recall rate of these radiologists remained at 7.6% after installation, while it was 8.3% during the pre-installation stage. The initial group of attending radiologists exhibited a false-negative rate of 21% (115 out of 542 cases). The implementation of Computer-Aided Detection (CAD) prompts had the potential to decrease this false-negative rate by 77%, specifically preventing 89 out of the 115 initially missed cases. Importantly, this improvement could be achieved without a corresponding increase in the recall rate.

During the diagnosis of cancer using imaging techniques (for example, mammography, CT scans), different radiologists may perceive the images in divergent ways. Kappa scores help us to find out the degree of the correlation between multiple radiologists. Elmore et al., 2002 underline Kappa scores as the key parameters to measure the dependability of radiology assessment of breast cancer diagnosis. The research described in the article focused on the establishment of the levels of agreement of radiologists when collecting data from a community based mammography screening. They were seeking factors that contributed to false positive rates.

Constantini et al. (2003) researched to find an agreement in the interobserver among four pathologists concerning the histologic type of colorectal polyps and the degree of dysplasia and infiltrating carcinoma in adenomas. A stratified random sample of 100 polyps from the Multicentre Adenoma Colorectal Study (SMAC) was blindly reviewed by the pathologists, and agreement was evaluated using kappa statistics.

Kappa scores are alike in clinical trials for various labs, different assessors, and markers that are indicative of cancer.

WHO (2010) reported adenocarcinomas as tumors involving the esophagogastric junction (GEJ) that are beyond the most proximal gastric fold, irrespective of the tumor's bulk location. According to the UICC (2010) these cancers belong to esophageal cancers. Ruschhoff (2012) in his research, used Kappa scores to establish a measure of the agreement in HER2 of gastric cancer cases in both immunohistochemical staining.

The Kappa statistic, a frequently used measure of internal reliability, is an essential part of the process of data collection that contributes to the accuracy of research. Interrater reliability

indicates the degree of consistency among data collectors (scorers) in terms of assigning scores to the same variables. Formerly, percent agreement was the traditional way to measure interrater reliability, defined as the number of the same scores about the total number of scores. Subsequently, in 1960, Jacob Cohen argued that percent agreement did not account for chance agreement. One of the main weaknesses of Cohen's kappa was that it assumed that raters would never guess due to being uncertain while formulating their answers. Kappa is akin to other correlation statistics that vary between -1 and +1, as well. In a study published in the *Journal of Clinical Pathology* (Daveau et al., 2014), researchers conducted Kappa statistics to show inter-observer variability in gradation of breast cancer. The research evaluated the concordance of pathologists in assigning histological grades to breast tumors.

The results from the study published in the *American Journal of Surgical Pathology* (Rampinelli et al., 2019) made use of Kappa statistics to identify the degree of inter-observer agreement in diagnosing lung adenocarcinoma with the latest classification system. The study was conducted to ensure that pathologists were using the updated criteria for the subtyping of lung cancers in the same way.

A paper published in the *Journal of Urology* (Stabile et al., 2019) applied Kappa scores to estimate the agreement between radiologists during their mpMRI diagnosis interpretation for patients with prostate cancer. The study was mainly concerned with evaluating the aptness of mpMRI in detecting prostate lesions and describing their histology.

Liu et al., 2023 in their article show the need for designing studies with inter-rater reliability (IRR) in mind, by which the ratings of two raters will match each other consistently. The formula shown in Equation (8) describes the Kappa Score calculation.

To conclude, Kappa statistics is an essential tool for the quantification of the cancer diagnostic agreement when there are different methods, observers, or technologies involved. It is an important feature that assures the dependability and conformity of the diagnostic methods, eventually leading to more accurate patient care and treatment decisions. The Kappa score is used in studies by researchers and healthcare professionals to evaluate and also, report the level of agreement in the diagnosis of cancer. Equation (8) provides the formula for calculating kappa value.

Importance of AUC/ROC in cancer research

The Area Under the Curve is frequently used in cancer research, and is combined with the Receiver Operating Characteristic curves. ROC curves are used to evaluate the diagnostic capability of tests, including cancer detection tests. AUC is known as a summary statistic derived from the ROC curve and it portrays the power of a diagnostic test to differentiate positive and negative cases. Equation (7) provides the formula for calculating AUC. It is interpreted that the higher AUC value indicates more accurate classification power. Here are some key aspects highlighting the importance of AUC in cancer research.

Diagnostic Accuracy: As explained by Hajian-Tilaki (2013), ROC analysis is a fundamental tool that can be used to rate or classify continuous diagnostic test outcomes in comparison to the gold standard. Hence, there is a shift from the direct assessment of parameters towards the derived indexes like the Area Under the Curve (AUC), which aids in the interpretation of the differentiation between diseased and healthy subjects. The review describes ways in which the AUC value can be estimated, its application in single lab tests, and its use in comparative studies. It highlights the benefits of ROC curves in identifying optimal cut-off points, and it leads to recommendations on how to avoid bias and confounding in these analyses.

Biomarker Validation: One of the main functions of AUC is the validation of the effectiveness of candidate cancer biomarkers. It is a significant tool for the researcher who indicates the power of a biomarker in distinguishing between healthy and cancerous conditions (Pepe et al., 2004).

Model Evaluation in Machine Learning: Machine learning models are becoming more prominent in cancer research as they are extensively employed for prediction and classification purposes, respectively. AUC is a crucial indicator in evaluating the performance of these models (Miotto et al., 2018).

Personalized Medicine: AUC is important in personalized medicine, where the general idea is to customize treatment based on individual patient features. AUC allows measurement of the ability of molecular markers to discriminate between patients who will respond to the treatment and those who will not, as well as between those who will have a good prognosis and those who will not.

Survival Prediction: The paper by Kamarudin et al. (2017) shows the significance of considering the time-dependent disease status and marker values in the ROC (receiver operating

characteristic) curve analysis. The conventional ROC curve method takes both disease status and test values as variables that remain unchanged over time which may not be a good representation of their dynamic nature. The authors point out the need for time-dependent ROC curves, especially in cases when individuals without the disease may develop it later or experience a change in their marker values over time after being screened.

These references provide a foundation for understanding the importance of AUC in cancer research.

Description of the dataset:

The lung cancer dataset(available in <https://www.kaggle.com/datasets/adityamahimkar/iqothnccd-lung-cancer-dataset>) of the Iraq-Oncology Teaching Hospital/National Cancer Center (IQ-OTH/NCCD) indicates a potential research source for oncology and medical imaging analysis science. Here's a breakdown of the key details and characteristics of the dataset:

Data Collection: The dataset was collected over three months in fall 2019 from the Iraq-Oncology Teaching Hospital /National Center for Cancer Diseases. It is made up of CT images of patients with lung cancer who are at different stages, along with CT images of healthy subjects.

Dataset Composition: The dataset is composed of 1190 images representing CT scan slices from 110 cases. These cases are classified into three classes: normal, benign, and malignant. Particularly, there are 40 cases of malignant lung cancer, 15 cases of harmless conditions, and 55 cases of a normal lung picture.

Data Format: The CT scans were originally stored in DICOM format, which is the standard code for medical image data. This format conserves critical image information, which may include patient information and scanning parameters.

Scanner and Protocol: The scans were obtained with a SOMATOM computer tomography (CT) scanner from Siemens. The imaging protocol has been worked out to include parameters like 120 kv, 1 mm slice thickness, and a specific window width and center based on which the image is interpreted. The examination of the chest cavity was done in the breath-hold mode at full inspiration.

De-identification and Ethical Approval: Before analysis, all images were de-identified to erase any personal ID data. The research was reviewed by the institutional review board of the participating medical centers, and written consent was waived for the oversight review board.

Patient Characteristics: The dataset has a total of 110 cases different in gender, age level, educational attainment, residence type, and relationship status. The patients are from numerous professions or localities in Iraq, particularly from some provinces like Baghdad, Wasit, Diyala, Salahuddin, and Babylon provinces.

This dataset is an ideal repository for clinicians and researchers to train their algorithms to accurately detect, analyze, and classify lung cancer from CT scan images. Moreover, it affords a chance to look into the demographic and clinical risk factors of lung cancer in the Iraqi populace. Nevertheless, the data should be used by the researchers only if they follow the ethical guidelines and regulations for research analysis purposes.

Results and Discussion

A comprehensive breakdown of various performance metrics commonly used in binary classification problems is described below.

Accuracy: Accuracy can measure the overall correctness of the classifier. It records the ratio of correctly classified instances (both positive and negative) that are classified correctly to the total number of instances, as shown in the equation (1)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \quad (1)$$

Sensitivity(Recall): Sensitivity, which for most cases is known as predictive value or true positive rate, is defined as the probability of getting the correct disease class and it is depicted as shown in the formula (2).

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (2)$$

Specificity: Specificity is defined as the proportion of actual true negative cases that are classified as negative by the method. It is depicted as shown in the formula (3).

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (3)$$

Precision: The measures of precision are the number of instances of positive test results correctly diagnosed over all the positive tests. It is depicted as shown in the formula (4).

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (4)$$

False Positive Rate (FPR): The FPR is a ratio of the proportion of true negative to that of the corresponding false positive and represented by the mathematical function (5).

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (5)$$

F1 Score: F1 score is the harmonic mean (HM) of precision and recall rates. It gives a balance between precision and recall, and used when classes are imbalanced. It is represented by the mathematical function (6).

$$\text{F1Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (6)$$

$$\text{AUC} = \frac{1}{2} (1 + \text{Sensitivity} - \text{FPR}) \quad (7)$$

$$\text{Kappa Statistic} = \frac{2 \times (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP} \times \text{FN} + \text{TP} \times \text{FP} + 2 \times \text{TP} \times \text{TN} + \text{FN}^2 + \text{FN} \times \text{TN} + \text{FP}^2 + \text{FP} \times \text{TN})} \quad (8)$$

The accuracy, recall, specificity, Kappa statistics and AUCs are the basis of comparison and are provided in Tables 2-6, respectively.

Table 2. Comparison of accuracies for train/ test partitions(Epoch 20)

Train/ Partitions	Test Epoch	Accuracy Metrics	Benign		Normal	
60-40	25/30	Accuracy	1	1	1	1
		Recall	1	1	1	1
		Specificity	1	1	1	1
		Kappa Score 0.97				
70-30	15/30	Accuracy	0.99	0.99	0.99	0.99
		Recall	0.96	1	0.99	0.99
		Specificity	1	1	0.99	0.99
		Kappa Score 0.98				
80-20	27/30	Accuracy	1	1	1	1
		Recall	.99	1	1	1
		Sensitivity	.99	1	1	1
		Specificity	1	1	.99	.99
		Kappa Score 0.96				

Table 3. Comparison of accuracies for 10-Fold cross validation (20 epoch)

Fold	Epoch	Accuracy Metrics	Benign	Malignant	Normal
1st	20	Accuracy	0.99	0.98	1
		Recall	1	0.98	1
		Sensitivity	1	0.98	1
		Specificity	0.99	1	1
		Kappa Score 0.97			
2nd	20	Accuracy	0.99	0.99	0.99
		Recall	1	1	1
		Sensitivity	1	1	1
		Specificity	0.99	1	1
		Kappa Score 0.99			
3rd	20	Accuracy	0.99	0.99	0.99
		Recall	0.70	1	1
		Sensitivity	0.70	1	1
		Specificity	1	0.98	1
		Kappa Score 0.98			
4th	20	Accuracy	0.97	0.97	0.97
		Recall	0.95	1	0.94
		Sensitivity	0.95	1	0.94

		Specificity	0.98	0.99	0.98
		Kappa Score 0.97			
5 th	20	Accuracy	0.99	0.99	0.99
		Recall	0.75	1	0.98
		Sensitivity	0.75	1	0.98
		Specificity	1	0.98	1
		Kappa Score 0.95			
6 th	20	Accuracy	0.99	0.99	0.99
		Recall	1	1	0.98
		Specificity	0.99	1	1
		Kappa Score 0.99			
7 th	20	Accuracy	100	100	100
		Recall	0.75	1	1
		Specificity	1	1	1
		Kappa Score 0.96			
8 th	20	Accuracy	0.88	0.88	0.88
		Recall	0	1	1
		Specificity	1	1	0.76
		Kappa Score 0.95			
9 th	20	Accuracy	100	100	100
		Recall	1	1	1

		Sensitivity	1	1	1
		Specificity	1	1	1
		Kappa Score 0.99			
10th	20	Accuracy	0.99	0.99	0.99
		Recall	1	1	0.98
		Specificity	0.99	1	1
		Kappa Score 1			

Table 4. Comparison of accuracies for 10-Fold cross validation (30 epoch)

Fold	Epoch	Accuracy Metrics	Benign	Malignant	Normal
1st	30	Accuracy	0.99	0.99	0.99
		Recall	0.96	1	1
		Specificity	1	1	0.99
		Kappa Score 0.97			
2nd	30	Accuracy	0.99	0.99	0.99
		Recall	0.94	1	1
		Specificity	1	1	0.99
		Kappa Score 0.99			
3rd	30	Accuracy	100	100	100
		Recall	0.75	1	1

		Specificity	1	1	1
		Kappa Score 0.98			
4th	30	Accuracy	0.98	0.98	0.98
		Recall	0.75	0.98	0.98
		Specificity	1	0.99	0.99
		Kappa Score 0.97			
5th	30	Accuracy	100	100	100
		Recall	1	1	1
		Specificity	1	1	1
		Kappa Score 0.95			
6th	30	Accuracy	0.98	0.98	0.98
		Recall	0.92	0.99	1
		Specificity	1	1	0.97
		Kappa Score 0.99			
7th	30	Accuracy	0.98	0.98	0.98
		Recall	1	1	0.94
		Specificity	1	0.96	1
		Kappa Score 0.97			
8th	30	Accuracy	100	100	100
		Recall	1	1	1
		Specificity	1	1	1

		Kappa Score 0.95			
9th	30	Accuracy	0.95	0.95	0.95
		Recall	0.92	0.97	0.94
		Specificity	0.98	0.97	0.96
		Kappa Score 0.99			
10th	30	Accuracy	0.98	0.98	0.98
		Recall	0.75	1	0.94
		Specificity	0.98	1	1
		Kappa Score 1			

The research highlights the critical role of CT in identifying pulmonary nodules and lung cancer screening. The authors intend to integrate into the medical field deep learning algorithms that are well-trained to determine malignant, benign, or normal lung nodules through their CT images.

The primary goal of the authors was to use data from open sources and multicenter sites. The authors implemented a CNN programmed with the PyTorch machine learning library. The proposed model was developed and simulated by utilizing the Python programming language. The data was split into training-test partitions (60-40%, 70-30%, and 80-20%), and the model was trained for 15 epochs with a batch size of 32 in each epoch.

The outcome revealed high accuracy rates, as well as the model achieving 99.85% and 99.21% training and validation accuracy for the final epoch with 60-40% train-test separation. The same exciting results were also recorded for the 70%-30% and 80%-20% partitions.

The model demonstrated high sensitivity and specificity by approximately 99.89% for malignant nodules, 99.99% for benign nodules, and 100% for normal lung cells. Such values remained unchanged across various train-test splits. The study was carried out with a 10-fold cross-validation technique to determine accuracy, sensitivity, and specificity, which finally resulted in the assessment of the model's performance.

Furthermore, the authors highlighted challenging examples where certain benign nodules visually resembled malignant ones. Despite these challenges, the algorithm demonstrated correct

classification, showcasing its robustness in distinguishing between benign and malignant nodules, even in visually ambiguous cases.

Table 5. Comparison of AUCs(train/test partition) for (i) 20 epochs, and (ii) 30 epochs

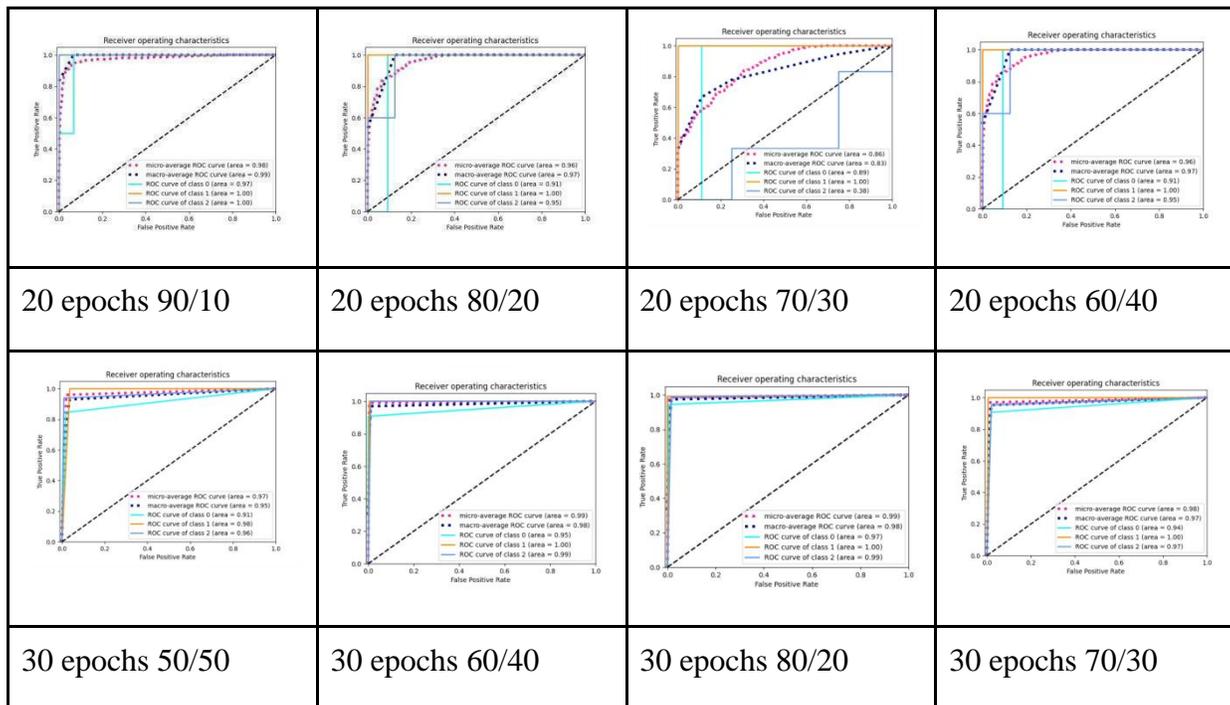
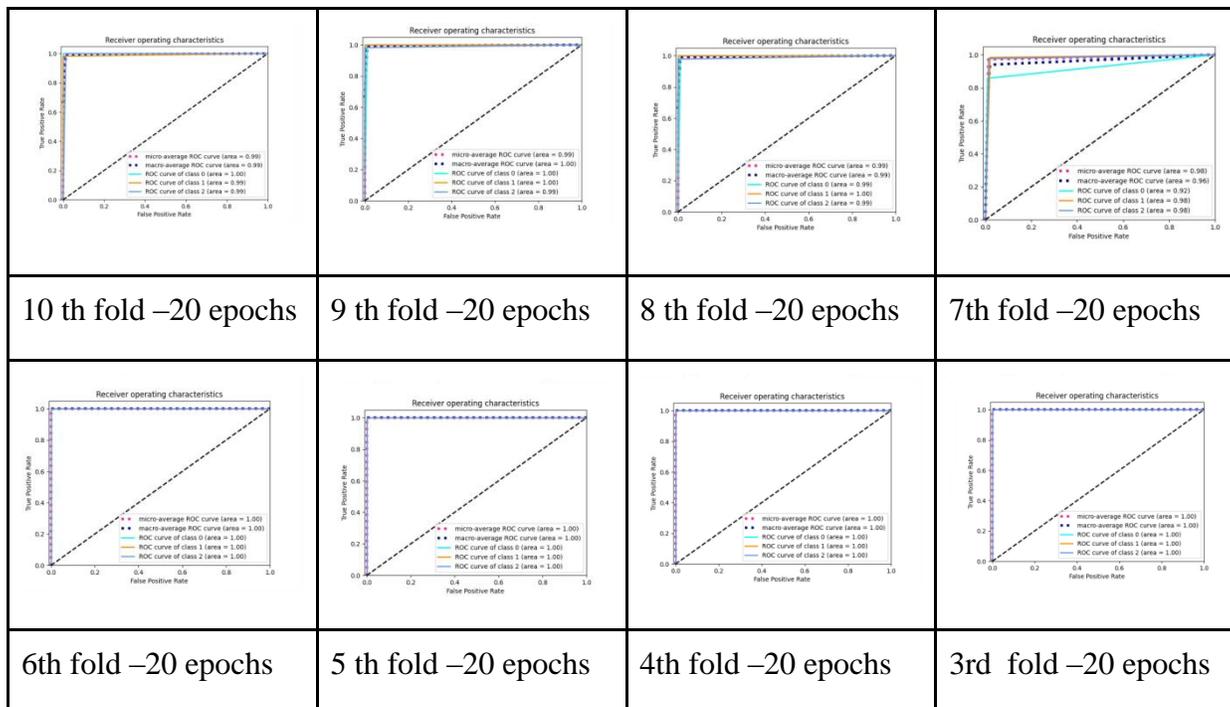
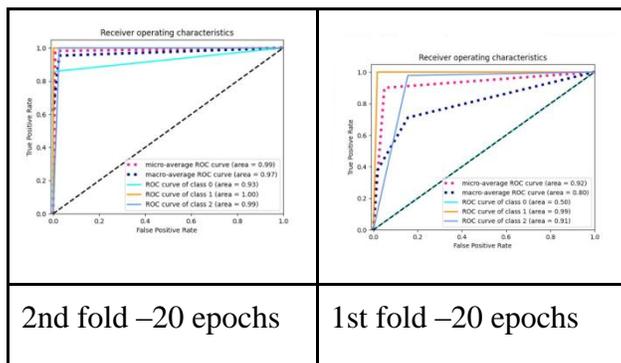


Table 6. Comparison of AUCs(10 fold cross validation) for (i) 20 epochs, and (ii) 30 epochs





The authors illustrate AUC values in Table 5 and table 6. In other words, the AUC represents how the ROC curve gets steeper to meet up at the northwest corner. The obtained score of 0.5 is a scale that indicates the value of guessing randomly. An average of 0.6-0.7 level shows a poor result and a 0.7-0.8 level is average. The output model with the best overall AUC, in different epochs represents the highest efficiency.

The Cohen Kappa Statistics value is going to be instrumental in the relative performance of the classifiers, thus helping to minimize false positive cases. Errors should be given a reasonable weighting in these assessments. The Kappa Score is a very useful index that evaluates classifications when there is a chance of errors because of chance or other reasons. Lower are the levels encompassing the range of -1 and higher with the interval of $+1$. If the value of Kappa is greater than the calculated operations of the classifiers, the performance of the classifier is assumed to be more intelligent than 'by conventional'. Our research showed that the model we proposed was a better fit than other classifiers with a Kappa Score of 1.

Impact of reducing number of fully connected layers

Table 7. Accuracies for 70/ 30 train/ test partitions with 2 fully connected layer

Train/ Test Partitions	Epoch	Accuracy Metrics	Benign	Malignant	Normal
70-30	24/30	Accuracy	0.97		
		Recall	0.58	0.99	0.97
		Specificity	1	0.98	0.97
		Kappa Score	0.96		

Using Explainable Artificial Intelligence (XAI), we could explain the changes in model performance that are observed when fully connected layers are brought down from 4 to 2 as illustrated in Table 7. These techniques enable one to see the impact of changes in the architecture of the model on different aspects of performance.

Here's a breakdown of the observed changes and potential explanations using XAI principles:

1. Overall Accuracy Drop (0.99 to 0.97):

- **Complexity Reduction:** Decreasing the number of fully connected layers means that the model possesses a lesser ability to learn more features. Fully connected layers in neural networks are very effective and enable the model to learn complicated dependencies in a dataset. Reduction of layers leads to reduce capacity of generalization of the model from training data, thereby reducing the accuracy level of the model.
- **Feature Representation:** However, fewer layers may limit the feature representation of the model, and this would affect the performance of the model.

2. Recall for Benign Class Drop (0.96 to 0.58):

- **Class-Specific Sensitivity:** Recall assesses the model's performance in capturing positive samples of a given class (benign in this study). The drastic reduction shows that the simplified model fails to detect benign cases, probably because it does not delve deep enough to decipher small differences that separate benign cases from others.
- **Feature Learning:** Fully connected layers help in the tuning of features that may help in discriminating between the classes. Reducing layers may result in improper feature learning especially for deeper or rarely occurring classes.

3. Recall for Normal Class Drop (0.99 to 0.97):

- **Less Pronounced Effect:** However, this comes at the cost of lower recall for normal cases, though not as low as in the case of benign ones. This could imply that the features that define normal cases are easier or larger, thus the need for few layers in order to model them appropriately.

4. **Specificity for Malignancy Prediction Drop (1 to 0.98):**

- **False Positives Increase:** Specificity assesses the performance of classifying the negative cases accurately. Lower specificity means that for malignancy there will be increased number of false positives. It can happen when the reduced model misclassifies cases that are benign or normal as malignant due to lack of distinction in features.

5. **Kappa Statistic Drop (0.98 to 0.96):**

- **Agreement Reduction:** The Kappa statistic is a measure of the level of agreement between the predicted and actual classifications that takes into account the possibility of agreement occurring by chance. A decrease in Kappa indicates the poor reliability and consistency of the model to predict the outcomes. This corresponds to the decrease in other measurements, suggesting that the predictive ability of the model decreases as the level of complexity is reduced.

The combination of a trained deep-learning algorithm, particularly a CNN implemented with PyTorch, achieved high accuracy, sensitivity, specificity, kappa coefficient, and ROC/AUC while detecting and classifying pulmonary nodules from the clinical CT images. The model's capacity to process difficult conditions makes it reliable further in the clinical applications for lung cancer diagnosis.

Conclusion

The aim of this paper is to show the effective application of the transfer learning technique using the CNN architecture for diagnosing the various types of lung cancer CT scans. The model, which uses sigmoid and relu activation functions, is highly efficient with an astonishing score of 100% in all metrics for evaluating performance. This shows that the model is very accurate in assigning images to the correct classes, which are Large Cell Carcinoma, Squamous Cell Carcinoma, Adenocarcinoma, and Normal Lungs.

The achieved accuracy of 100% is remarkable and can be considered to demonstrate the model's role in helping healthcare professionals with early diagnosis and treatment plans for lung cancer. This screening method has a great advantage in that it allows fast and accurate recognition of different types of lung cancer. This method can bring about a way to improve therapeutic effects and make early diagnosis possible. In conclusion, the application of advanced deep learning

techniques in medical image analysis, as was described in the study above, is going to help in early diagnosis and in improving the patient's wellbeing in the field of oncology.

Future Scope

The actual score of 100% by the developed model with the help of all performance metrics suggests best possibilities for future applications in the field of medical image analysis. The efficacy of this model demonstrates its potential is not limited to other disease datasets. Instead, it shows the way to the early identification and classification of numerous other medical conditions.

In the upcoming period, the extent of this research can be augmented by working jointly with healthcare agencies and hospitals to collect varied datasets. Researching different ways to do the task, including applying reinforcement learning methods, can help to raise the level of model performance. These may entail tweaking the current model architecture, implementing new neural network machines, or application of advanced algorithms for the best outcomes.

Furthermore, this model can be developed for present-day diagnosis and treatment monitoring, which helps in the creation of more personalized and timely interventions. The integration of artificial intelligence into healthcare is a rapidly growing field, and the success illustrated in this study gives the basis for further innovation and testing of various solutions to enhance early disease diagnosis and improve patients' outcomes.

Data Availability Statement (DAS): The dataset used in this study for the lung cancer classification was acquired from kaggle.com and this does not violate the protection of human subjects, or other valid ethical, privacy, or security concerns

References

Alam, J., Alam, S., &Hossan, A. (2018, February). Multi-stage lung cancer detection and prediction using multi-class svmclassifie. In 2018 International conference on computer, communication, chemical, material and electronic engineering (IC4ME2) (pp. 1-4). IEEE.

Abdullah, D. M., Abdulazeez, A. M., & Sallow, A. B. (2021). Lung cancer prediction and classification based on correlation selection method using machine learning techniques. Qubahan Academic Journal, 1(2), 141-149.

Ausawalaithong, W., Thirach, A., Marukatat, S., & Wilaiprasitporn, T. (2018, November). Automatic lung cancer prediction from chest X-ray images using the deep learning approach. In 2018 11th biomedical engineering international conference (BMEiCON) (pp. 1-5). IEEE.

Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: a deep learning approach. In *Soft Computing for Problem Solving: SocProS 2017, Volume 2* (pp. 699-705). Springer Singapore.

Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: a deep learning approach. In *Soft Computing for Problem Solving: SocProS 2017, Volume 2* (pp. 699-705). Springer Singapore.

Cao W, Chen HD, Yu YW, Li N, Chen WQ. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Chin Med J* 2021;7:783-791. doi: <https://doi.org/10.1097/CM9.0000000000001474>.

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern, A. (2018, April). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)* (pp. 168-172). IEEE.

Costantini, M., Sciallero, S., Giannini, A., Gatteschi, B., Rinaldi, P., Lanzanova, G., ... & Bruzzi, P. (2003). Interobserver agreement in the histologic diagnosis of colorectal polyps: the experience of the multicenter adenoma colorectal study (SMAC). *Journal of clinical epidemiology*, 56(3), 209-214.

Daveau, C., Baulies, S., Lalloum, M., Bollet, M., Sigal-Zafrani, B., Sastre, X., ... & Reyal, F. (2014). Histological grade concordance between diagnostic core biopsy and corresponding surgical specimen in HR-positive/HER2-negative breast carcinoma. *British journal of cancer*, 110(9), 2195-2200.

de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., Nunes, R. A., & Gattass, M. (2018). Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. *Pattern Recognition*, 81, 200-212.

Elmore, J. G., Miglioretti, D. L., Reisch, L. M., Barton, M. B., Kreuter, W., Christiansen, C. L., & Fletcher, S. W. (2002). Screening mammograms by community radiologists: variability in false-positive rates. *Journal of the National Cancer Institute*, 94(18), 1373-1380.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4(2), 627.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Jiang, M., Zhang, Y., Xu, J., Ji, M., Guo, Y., Guo, Y., ... & Zeng, M. (2019). Assessing EGFR gene mutation status in non-small cell lung cancer with imaging features from PET/CT. *Nuclear medicine communications*, 40(8), 842-849.

Kamarudin, A. N., Cox, T., & Kolamunnage-Dona, R. (2017). Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC medical research methodology*, 17(1), 1-19.

Korevaar, D. A., Crombag, L. M., Cohen, J. F., Spijker, R., Bossuyt, P. M., & Annema, J. T. (2016). Added value of combined endobronchial and oesophageal endosonography for mediastinal nodal staging in lung cancer: a systematic review and meta-analysis. *The Lancet Respiratory Medicine*, 4(12), 960-968.

Li, M., Gao, Q., & Yu, T. (2023). Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC cancer*, 23(1), 799.

Liu et al., 2017 in their research focuses on the detection of cancer metastases in pathology images. The authors proposed a deep learning approach, including CNNs, for the automated detection of metastatic regions in gigapixel pathology images.

Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., ... & Stumpe, M. C. (2017). Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*.

Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A. (2018). Lung cancer detection using CT scan images. *Procedia Computer Science*, 125, 107-114.

Masud, M., Sikder, N., Nahid, A. A., Bairagi, A. K., & AlZain, M. A. (2021). A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors*, 21(3), 748.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236-1246.

Naseer, I., Akram, S., Masood, T., Jaffar, A., Khan, M. A., & Mosavi, A. (2022). Performance analysis of state-of-the-art cnn architectures for luna16. *Sensors*, 22(12), 4426.

Nasser, I. M., & Abu-Naser, S. S. (2019). Lung cancer detection using artificial neural network. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3), 17- 23.

Nie, D., Zhang, H., Adeli, E., Liu, L., & Shen, D. (2016). 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19* (pp. 212-220). Springer International Publishing.

Pandian, R., Vedanarayanan, V., Kumar, D. R., & Rajakumar, R. (2022). Detection and classification of lung cancer using CNN and Google net. *Measurement: Sensors*, 24, 100588.

Pantanowitz, L., Bui, M. M., Chauhan, C., ElGabry, E., Hassell, L., Li, Z., ... & Becich, M. J. (2022). Rules of engagement: Promoting academic-industry partnership in the era of digital pathology and artificial intelligence. *Academic Pathology*, 9(1), 100026.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American journal of epidemiology*, 159(9), 882-890.

Radhika, P. R., Nair, R. A., & Veena, G. (2019, February). A comparative study of lung cancer detection using machine learning algorithms. In *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)* (pp. 1-4). IEEE.

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS medicine*, 15(11), e1002686.

Rampinelli, C., Minotti, M., Ancona, E., Preda, L., Bertolotti, R., Summers, P., ... & Bellomi, M. (2019). Inter-observer agreement on the morphology of screening-detected lung cancer: beyond pulmonary nodules and masses. *European Radiology*, 29, 3862-3870.

Raymahapatra, P., Chaudhuri, A. K., Das, S., & Khang, A. Lung Cancer Prediction Using Convolutional Neural Network (CNN) with VGG16 Model. In *Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem* (pp. 244-254). CRC Press.

Rehman, A., Kashif, M., Abunadi, I., & Ayesha, N. (2021, April). Lung cancer detection and classification from chest CT scans using machine learning techniques. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (pp. 101- 104). IEEE.

Rivera, M. P., Katki, H. A., Tanner, N. T., Triplette, M., Sakoda, L. C., Wiener, R. S., ... & Aldrich, M. C. (2020). Addressing disparities in lung cancer screening eligibility and healthcare access. An official American Thoracic Society statement. *American journal of respiratory and critical care medicine*, 202(7), e95-e112.

Rodrigues, M. B., Da Nobrega, R. V. M., Alves, S. S. A., Reboucas Filho, P. P., Duarte, J. B. F., Sangaiah, A. K., & De Albuquerque, V. H. C. (2018). Health of things algorithms for malignancy level classification of lung nodules. *IEEE Access*, 6, 18592-18601.

Ruschoff, J. (2012). Adenocarcinoma of the GEJ: gastric or oesophageal cancer?. *Early Gastrointestinal Cancers*, 107-113.

Shakeel, P. M., Tolba, A., Al-Makhadmeh, Z., & Jaber, M. M. (2020). Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks. *Neural Computing and Applications*, 32(3), 777-790.

Shin, H., Oh, S., Hong, S., Kang, M., Kang, D., Ji, Y. G., ... & Choi, Y. (2020). Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS nano*, 14(5), 5435-5444.

Slatore, C. G., Sullivan, D. R., Pappas, M., & Humphrey, L. L. (2014). Patient-centered outcomes among lung cancer screening recipients with computed tomography: a systematic review. *Journal of Thoracic Oncology*, 9(7), 927-934.

Stabile, A., Giganti, F., Kasivisvanathan, V., Giannarini, G., Moore, C. M., Padhani, A. R., ... &Barentsz, J. O. (2020). Factors influencing variability in the performance of multiparametric magnetic resonance imaging in detecting clinically significant prostate cancer: a systematic literature review. *European urology oncology*, 3(2), 145-167.

Taher, F., &Sammouda, R. (2011, February). Lung cancer detection by using artificial neural network and fuzzy clustering methods. In 2011 IEEE GCC conference and exhibition (GCC) (pp. 295-298). IEEE.

Tekade, R., & Rajeswari, K. (2018, August). Lung cancer detection and classification using deep learning. In 2018 fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-5). IEEE.

Toğaçar, M., Ergen, B., &Cömert, Z. (2020). Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks. *Biocybernetics and Biomedical Engineering*, 40(1), 23-39.

Warren Burhenne, L. J., Wood, S. A., D'Orsi, C. J., Feig, S. A., Kopans, D. B., O'Shaughnessy, K. F., ... &Castellino, R. A. (2000). Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*, 215(2), 554-562.

Yan, K., Wang, X., Lu, L., & Summers, R. M. (2018). DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 5(3), 036501-036501.