Technical Paper

Using Convolutional Neural Network to Enhance Classification Accuracy of Cancerous Lung Masses from CT Scan Images

Mohammad Mahdi Nakhaie[1], Sasan Karamizadeh[1,2*], Mohammad Ebrahim Shiri Ahmad Abady[3] and Kambiz Badie[4]

1. Ershad Damavand Institute of Higher Education, Tehran, Iran.
2. ICT Research Institute, Tehran, Iran.
3. Amirkabir University of Technology, Tehran, Iran.
4. E-Content & E-Services Research Group, IT Research Faculty, ICT Research Institute, Karegar, Tehran, 14155-3961, Tehran, Iran.

Article Info	Abstract
	Lung cancer is a highly serious illness, and detecting cancer cells early significantly enhances patients' chances of recovery. Doctors regularly examine a large number of CT scan images, which can lead to fatigue and errors. Therefore, there is a need to create a tool that can automatically detect and classify lung nodules in their early stages. Computer-aided diagnosis systems, often employing image processing and machine learning techniques, assist radiologists in identifying, and categorizing these nodules. Previous studies have often used complex models or pre-trained networks that demand significant computational power and a long time to execute. Our goal is to achieve accurate diagnosis without the need for extensive computational resources. We introduce a simple convolutional neural network with only two convolution layers, capable of accurately classifying nodules without requiring advanced computing capabilities. We conducted training and validation on two datasets, LIDC-IDRI and LUNA16, achieving impressive accuracies of 99.7% and 97.52%, respectively. These results demonstrate the superior accuracy of our proposed model compared to state-of-the-art research papers.

1. Introduction

Lung cancer, one of the most prevalent forms of cancer, has garnered significant attention from medical professionals and researchers alike. Currently, it stands as the leading cause of cancer-related deaths worldwide. Lung cancer often manifests in its early stages as lung nodules, and chest CT scans serve as the primary method for its detection. One of the key challenges contributing to the high mortality rate of lung cancer is the absence of noticeable symptoms during its initial stages. By the time clinical symptoms become evident, the disease has often progressed to advanced stages, potentially spreading to other parts of the body, significantly reducing the chances of successful treatment and cure [1]. Despite the daunting 5-year survival rate of less than 5% for locally advanced lung cancer, early detection when the tumor is small and asymptomatic can substantially increase the 5-year survival rate to over 60%. This encouraging statistic has led to numerous lung cancer screening trials [2]. The widespread adoption of CT screening techniques has significantly escalated the workload for radiologists. The manual analysis of extensive CT scans has become an exceptionally laborious and time-consuming task [3]. The application of artificial intelligence (AI) techniques holds significant importance in designing classifier systems that can extract features from images and distinguish between those displaying signs of disease [4]. Computer-aided detection systems accelerate the detection speed [5]. Furthermore,

utilizing these techniques can prevent screening errors, alleviate the diverse workload on clinicians, extend patients' survival time, and enhance their overall quality of life [6]. The human body consists of trillions of cells that undergo growth and division to generate new cells following the body's needs, a process known as cell division. Typically, cells naturally die off, and are replaced with new ones as they age or become damaged. However, when this natural process is disrupted, damaged cells can begin to proliferate, leading to the development of tumors. These tumors can be categorized as either malignant or benign [7]. Lung cancer can be classified into two primary types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) [8].

There are several basic problems in the diagnosis and classification of lung cancer:

One of the primary concerns faced by medical professionals is the early detection of lung cancer in patients. This is crucial because the majority of cases are diagnosed at advanced stages, making treatment considerably more challenging [8]. The process of screening through CT scans presents significant obstacles for radiologists, as millions of individuals are anticipated to undergo these scans annually, resulting in a substantial volume of imaging data [9]. Identifying nodules in CT scans does not automatically indicate the presence of lung cancer, given the intricate relationship between nodule characteristics and the disease itself [2]. Consequently, their categorization is a formidable task due to the diverse nature of lung nodules and their visual resemblance to adjacent tissues.

This paper introduces a novel model utilizing deep learning algorithms aimed at enhancing the accuracy of the classification of cancerous cells. The key innovations highlighted in this paper are:

1. Improving classification accuracy compared to previous papers
2. Using a model with two convolution layers instead of complex models or pre-trained networks
3. Reduction of training time and no need for advanced hardware for training due to using a model with two convolution layers

The rest of the paper is organized as what follows. Section 2 contains the Literature Review, Section 3 contains the Materials and Methodology, Section 4 contains the Results, and Section 5 contains the Conclusions.

2. Literature Review
Al-Shabi *et al*. [10] established a 2D convolutional neural network model called "Gated Dilated" to classify lung nodules as either malignant or benign using CT images. The Gated Dilated 2D CNN employed a unique approach, utilizing multiple dilated convolutions instead of the traditional max-pooling layers. This allowed it to capture features at various scales effectively. The model included a context-aware sub-network that directed features towards specific expanded convolutions. In their research paper, the authors conducted a thorough performance comparison of the Gated Dilated 2D CNN against several other CNN-based models. These models included a conventional CNN with the same layer count and channel configurations, the Gated Dilated 2D CNN without dilation, the Gated Dilated 2D CNN without gating, a multi-crop CNN, and ResNet50 and DenseNet161, both trained using two different transfer learning methods; they utilized (LIDC-IDRI) dataset in their paper.

The performance of the Gated Dilated 2D CNN model was impressive, achieving an overall accuracy of 93%, a sensitivity of 92%, and an area under the receiver operating characteristic curve (AUC) of 95% on the test data. Notably, it slightly outperformed all the other CNN-based models employed in the study. The results also indicated a substantial improvement in accuracy when detecting medium-sized nodules with diameters ranging from 5 to 12 mm.

Shen *et al*. [11] developed an interpretable hierarchical semantic 3D Convolutional Neural Network (CNN) for evaluating the malignancy of lung nodules in CT scans. They used raw cube data centered around the nodules from the LIDC-IDRI dataset on the TCIA platform. This specific convolutional neural network (CNN) has two distinct output levels. The initial stage generates crucial diagnostic semantic characteristics, while the subsequent stage computes the ultimate prediction score, indicating the malignancy of the identified nodule. These intermediate outputs provide insights into how the model interprets the raw data and improves the accuracy of the final malignancy prediction by using jump connections within the model architecture. The drawback of this paper lies in its limited reliability, as it achieved an accuracy of only 84%, which is relatively low. In contrast, our proposed model demonstrated a significantly higher accuracy of 99.7% on the LIDC-IDRI dataset, making it a more dependable and robust solution.

Ali et al. [12] developed a transferable texture 2D Convolutional Neural Network (CNN) to enhance the classification of lung nodule status through single CT images. They used the LIDC-IDRI dataset on the TCIA platform in their paper.

One noteworthy aspect of their model was the integration of an energy layer within its architecture. This energy layer served the purpose of extracting texture features from the convolutional layer. The inclusion of this energy layer brought about several advantages including the reduction of model parameters. As a result, it led to a decrease in computational complexity and memory requirements for the model.

The transferable texture in 2D CNN delivered its most impressive performance when evaluated on the LIDC-IDRI dataset. It achieved remarkable results with an overall accuracy (ACC) of 97%, a sensitivity (SE) of 96%, and an area under the receiver operating characteristic curve (AUC) of 99% on the test data. These outcomes prove the model's proficiency in accurately categorizing lung nodule status, especially when applied to the LIDC-IDRI dataset.

Lin *et al*. [13] introduced a 2D convolutional neural network (CNN) combined with Taguchi parametric optimization for the automated classification of lung nodules as either malignant or benign using CT images.

This paper employed the Taguchi method, a statistical technique using orthogonal arrays to optimize process parameters. A total of 36 experiments were conducted, considering 8 control factors at different levels. The goal was to identify the best combination of parameters and improve the model's performance.

Liu *et al*. [14] introduced a three-dimensional (CNN)-based multi-model ensemble learning approach for identifying malignant and benign lung nodules in CT scans.

They introduced a 3D (CNN)-based multi-model ensemble learning approach for the identification of malignant and benign lung nodules in CT scans. They conducted their study using data from the LIDC-IDRI dataset available on the TCIA platform.

Their approach involved the creation of multiple independent neural networks, each designed to simulate different expert behaviors. These networks were then combined using ensemble learning techniques to consolidate their results. Ensemble learning using a multi-model approach with 3D CNNs featured three distinct architectural variations. These were based on VGGNet, ResNet, and Inception Net. Within each of these architectural categories, there were three substructures, each designed to accept different input sizes. In this paper, the authors achieved an Area Under the Curve (AUC) of 94%, whereas our proposed model attained a remarkable AUC of 99%. Notably, the Convolutional Neural Network

(CNN) employed in our study exhibits lower complexity compared to the CNN utilized in their research.

Kareem *et al*. [15] proposed a CAD-based system for the classification of lung cancer into either benign (non-cancerous) or malignant (cancerous) using the IQ-OTH/NCCD dataset. Here, images were first pre-processed by applying three image-processing techniques of image enhancement, segmentation, and feature extraction using Gabor and GLCM filters, which were then classified using an SVM classifier. Their proposed method obtained 89.88% accuracy using a polynomial kernel. Machine learning-based methods perform effectively with fewer datasets; however, they necessitate manual feature extraction and selection processes, leading to heightened computational time and complexity. These challenges have been addressed through the utilization of various deep learning (DL) models, particularly Convolutional Neural Networks (CNNs), which have proven instrumental in image classification, detection, and segmentation tasks, overcoming the aforementioned limitations.

Zhao *et al*. [16] developed a multi-scale, multi-task 3D convolutional neural network methodology for classifying lung nodules as malignant or benign based on CT scans. In this paper, the LIDC-IDRI dataset is used. They conducted multi-scale volume extraction, generating two different-scale volumes from the data. To enhance training, they introduced a novel loss function as part of their model training. The result shows that the accuracy for this model was 94 percent.

Training 3D Convolutional Neural Networks (CNNs) demands substantial computational resources. One potential solution to this challenge is the utilization of models like 2D CNNs, which can mitigate the computational demands while still delivering effective results.

Halder *et al*. [17] A framework for classifying lung nodule status from CT images was developed. This framework utilized adaptive morphology-based operations along with the Gabor filter. It was coined the "two-path morphological 2D CNN." They selected a dataset consisting of 2,600 lung nodule slices, evenly divided between 1,300 benign and 1,300 malignant cases. These slices were obtained from the LIDC-IDRI dataset available on the TCIA platform. Morphology-based operations: Various morphology-based operations were applied to filter and process the lung nodules. Gabor filter: The Gabor filter was employed to capture texture variations present in the lung nodules.

Two-path morphological 2D CNN: The core of

their framework consisted of a two-path 2D Convolutional Neural Network (CNN), where both paths utilized the VGGNet architecture.

Jena *et al*. [18] introduced a neural network approach with multiple-layer latent variables for the classification of lung nodules as malignant or benign using CT images. In their methodology, the researchers selected images from the LIDC-IDRI dataset, available on the TCIA platform, for their experiments. To enhance image quality, they applied Gaussian and Wiener filters for noise reduction. They employed a region-growing segmentation technique to identify the region of interest (ROI), merging adjacent pixels to create larger regions. From these segmented lung nodules, they manually extracted essential features like area, perimeter, entropy, and intensity. To manage feature dimensionality, a deep Gaussian mixture model was utilized. Finally, they conducted nodule classification using a region-based 2D convolutional neural network (CNN). They got 88 percent accuracy in their paper. However, they acknowledged the need for improvements in the feature extraction step. The drawback of this paper is their achieved sensitivity of 70%. In our study, we observed a sensitivity of 99% in the LIDC-IDRI dataset, indicating a substantial improvement over their results.

Yu *et al*. [19] introduced a framework for the segmentation and classification of lung nodules in CT scans. Their approach comprised two main components: for segmentation lung module they used 3D Res U-Net and 3D ResNet50 utilized for classification. Their comprehensive methodology consisted of several key steps. Firstly, they meticulously prepared the data by standardizing CT scans to the Hounsfield Unit (HU) scale and ensuring consistent voxel spacing in various planes. Subsequently, they enhanced data quality through processes such as binarization and the removal of unwanted regions like bed frames and air. To refine the lung mask, morphological techniques were applied to fill gaps in the lung parenchyma. They then employed a specialized 3D Res U-Net model for the automated segmentation of lung nodules, using the refined lung mask as a basis. Following this segmentation, specific regions of interest (ROIs) measuring $48 \times 48 \times 16$ voxels were extracted based on the predicted lung nodule mask. Finally, they achieved 87% accuracy in their method. In this paper, the authors reported an Area Under the Curve (AUC) of 91%, whereas our proposed model achieved a significantly higher AUC of 99% on the LIDC-IDRI dataset. Notably, the Convolutional Neural Network (CNN) utilized in our study is less complex than the one employed

in their research.

Liao *et al*. [20] introduced a unique two-module 3D deep U-net-like network architecture for lung nodule analysis. The first module's purpose was to detect and identify suspicious nodules within a patient's scans. Subsequently, the second module focused on selecting the top 5 nodules based on their detection confidence levels. It then calculated the probabilities, indicating the likelihood of these nodules being cancerous. The results showed a classification accuracy of 81.42%, indicating the model's proficiency in making correct predictions, and an AUC of 0.87, reflecting its capability in distinguishing between different classes in this context. One drawback of this paper is its relatively low accuracy of 81.42%, rendering the results less reliable. In contrast, our proposed model demonstrated a significantly higher accuracy of 97.52% on the LUNA16 dataset, establishing its superior reliability and performance.

A group of triple neural networks was developed by Utkin *et al*. [21] to address the inclusion of unusual cancer cases in their research. These networks were based on a collection of five histograms that were collected from segmented nodule areas, and described different facets of the morphology, inner structure, and exterior structure of lung nodules. The model's overall classification ability was demonstrated by the approach's impressive classification findings, which included an accuracy rate of 91.8%. Additionally, the model showed a high sensitivity of 90.8%, underscoring its ability to accurately identify true positives, and a precision of 92.6%, indicating its proficiency in doing so in this particular environment.

A classification framework for differentiating between benign and malignant lung nodules was presented by Zhang *et al*. [22]. To recalibrate features and improve feature reuse, their method combined a squeeze-and-excitation network with aggregated residual transformations, known as SE-ResNeXt modules. The results of their investigation demonstrated outstanding performance, with the model attaining an exceptional Area Under the Curve (AUC) score of 0.9563 and a high accuracy rate of 91.67%, highlighting its usefulness in this classification assignment.

Yuan *et al*. [23] introduced a 3D convolutional network designed for the classification of nodules versus non-nodules. This network extracted spatial information through three distinct paths, each with varying field sizes, and then combined these pathways later in the model. Their model necessitated approximately 48 hours of training time and utilized 10,855 MB of memory. In

contrast, our model requires less training time and consumes fewer system resources due to its lower complexity.

Zhang *et al*. [24] introduced a DenseNet architecture that incorporated 3D filters and pooling kernels, complemented by data augmentation techniques, to classify malignancy versus benign status. Their findings revealed noteworthy results, with the model achieving an accuracy rate of 92.4%. Additionally, the model demonstrated a high specificity of 96.0%, indicating its ability to accurately identify true negative cases, and a sensitivity of 87.0%, illustrating its capacity to effectively detect true positives in this classification task.

Mastouri *et al*. [25] introduced a classification approach for distinguishing between nodules and non-nodules. Their method utilized bilinear CNN (BCNN) structures in conjunction with a linear Support Vector Machine (SVM) for the classification task. They leveraged pre-trained VGG16 and VGG19 structures, which were fine-tuned on their dataset, to extract relevant features, forming the basis for the bilinear architectures. The results of their study were notable, with the model achieving an accuracy of 91.99%, indicating its overall classification performance. Additionally, the model exhibited a specificity of 92.27%, emphasizing its ability to accurately identify true negatives, and a sensitivity of 91.85%, highlighting its capacity to effectively detect true positives in the nodules *vs.* non-nodules classification task.

In conclusion, none of the previously state-of-the-art methods discussed in this literature review have demonstrated significant effectiveness in detecting cancerous lung masses, primarily due to minor practical differences among them. In our approach, we employed a low-complexity convolutional architecture. This design enables the network to discern intricate details within the images, distinguishing between those containing cancerous lung masses and non-cancerous counterparts. Our primary goal is to create a robust feature extractor from images, aiming to enhance accuracy for the classification of adult lung cancerous masses.

3. Materials and Methodology

This paper presents a comprehensive research framework encompassing five essential steps for the task at hand, as shown in Figure 1. Firstly, it begins with the acquisition and preparation of input images. These images are then subjected to rigorous data pre-processing techniques in the second step to enhance their quality and consistency. Moving on to the third step, a CNN model is carefully designed and configured to extract meaningful features from the pre-processed data. Subsequently, the fourth step involves training the CNN model using a suitable dataset, allowing it to learn and adapt to the specific task. Finally, in the fifth step, the trained model is deployed for classification purposes, enabling it to accurately categorize and make predictions based on the input data. Together, these five steps form a structured framework for addressing the problem at hand efficiently and effectively.
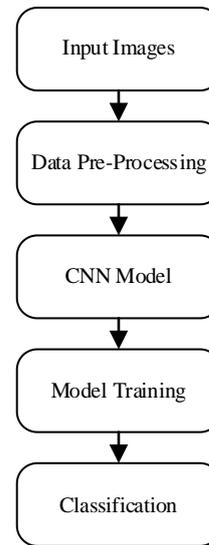

Figure 1. Research framework.

3.1. First phase: input images

We used two LIDC-IDRI and LUNA16 datasets in this paper.

CT scans remain the most effective modality for tumor identification due to their widespread availability, affordability, and superior resolution. Below we review the most important datasets that can be used for nodule classification in the lungs [26].

3.1.1. LIDC-IDRI dataset

The Lung Image Dataset Consortium and Image Dataset Resource Initiative (LIDC-IDRI) stand as the foremost and most extensively utilized resource in its field. It comprises thoracic CT scans, totaling 1018 scans derived from 1010 patients. These scans are formatted in Digital Imaging and Communications in Medicine (DICOM). Furthermore, an annotation file accompanies the dataset, containing information about lung nodules. It's worth noting that this dataset has undergone meticulous annotation by four radiologists. Within this dataset, each CT scan slice varies in thickness, ranging from 1.25 to 2.5 mm, and pixel size falls between 0.48 and 0.72 mm. In the initial annotation phase, each radiologist independently marked the CT scan images, classifying lesions as nodules with

a diameter of 3 mm or larger, nodules with a diameter of less than 3 mm or non-nodules with a diameter of 3 mm or larger. In a subsequent phase, each radiologist cross-referenced their annotations with those of their colleagues to arrive at a consensus on the findings. For each dataset entry, segmentation masks are provided for all the nodules present in the dataset [26].

3.1.2. LUNA16 dataset
The LUNA16 dataset comprises 888 CT scans, each with a slice thickness of less than 2.5 mm. These CT scan images are stored in the MHD format. It's crucial to note that this dataset is a subset of the larger LIDC-IDRI dataset and includes an annotation file detailing 1186 nodules. Importantly, these nodules have been meticulously annotated by a minimum of three radiologists. Nodules that received annotations from only one or two radiologists were excluded from the dataset [26].

3.2. Second phase: data pre-processing
In this section, from the LUNA16 dataset, we extracted 444 images of benign nodules and 444 images of malignant nodules, and the data are balanced. Figure 2 shows an example of a lung nodule image in this dataset.
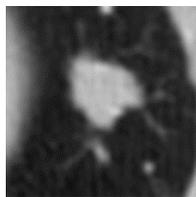

Figure 2. An example of a lung nodule image in the LUNA16 dataset.

To increase the number of data for the training proposed method, we rotated each nodule image in 3 directions, and reversed the original image. The result is shown in Figure 3.
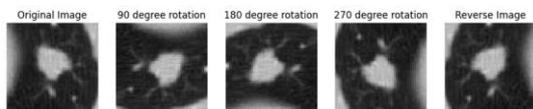

Figure 3. Augmenting data by rotating and inverting images in the LUNA16 dataset.

After this step, the number of data increased to 4440 data. Figure 4 shows that there are 2220 images of benign nodules and 2220 images of malignant nodules, so the data is balanced.

Because lung nodules are very small and cannot be easily identified on thick CT scans, we excluded scans with a slice thickness greater than 2.5 mm in the LIDC-IDRI dataset. Finally, 900 labeled CT scans were obtained after balancing the data,

including 450 benign CT scans and 450 malignant CT scans. We cut the CT scan images of the LIDC-IDRI dataset and separated the nodule part. Therefore, each image of the LIDC-IDRI dataset was divided into 100 parts by an automatic tool, and then we manually separated the lung nodule part from the other parts of the lung. The size of each of the divided images is 50x50. We can see the image sample in the LIDC-IDRI dataset in Figure 5 and the CT scan segmentation result in Figure 6. To increase the number of data by using an automatic tool written in Python, we rotated each nodule image in 3 directions, and reversed the original image. The result is shown in Figure 7.
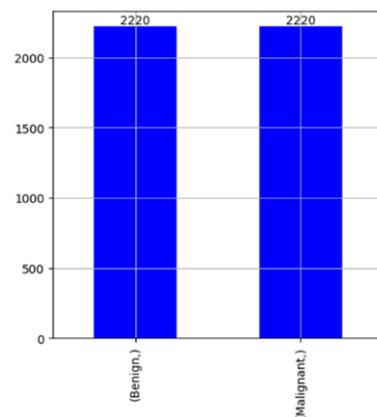

Figure 4. The number of data in both classes in the LUNA16 dataset.

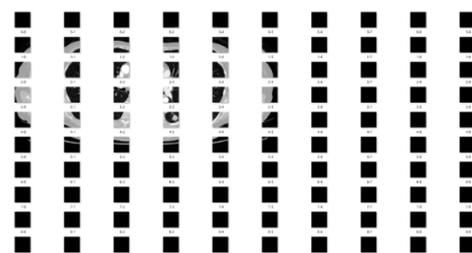**Figure 5. CT scan example of the LIDC-IDRI dataset.**

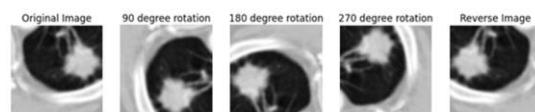**Figure 6. The result of dividing the CT scan into 100 parts.**

Figure 7. Data augmentation by rotating and inverting images in the LIDC-IDRI dataset.

After doing this step, the number of data increased to 4500 data. Figure 8 shows that there are 2250 images of benign nodules and 2250 images of malignant nodules, so the data is balanced.
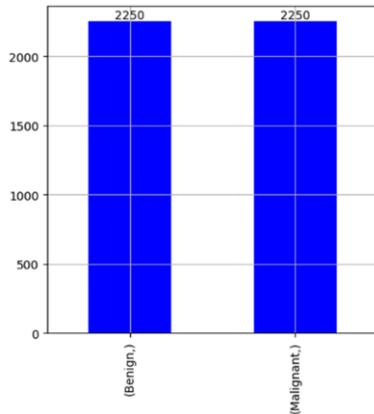

Figure 8. The number of data in both classes in the LIDC-IDRI dataset.

Further, in both datasets, we divided the data into training and validation data, and considered 70% of the data for training and 30% for validation.

3.3. Third phase: proposed model
This paper used a two-dimensional convolutional neural network with two layers. The structure of this neural network is shown in Figure 9.
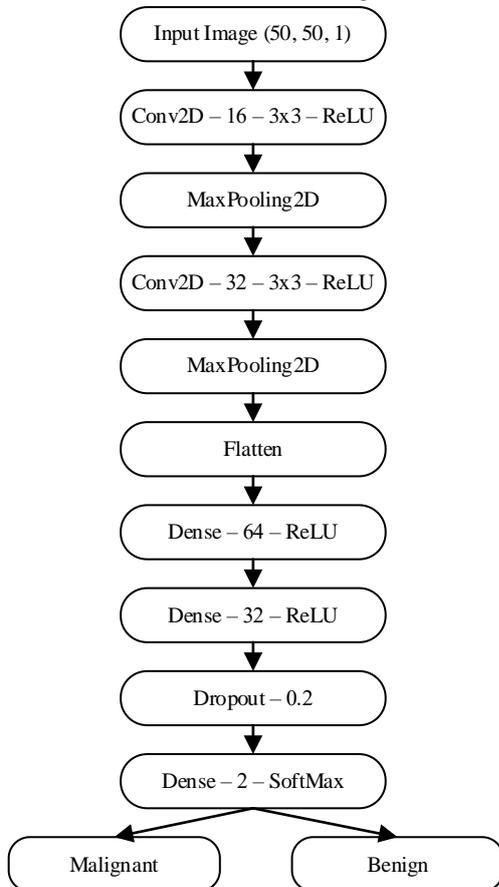

Figure 9. Structure of the proposed convolutional neural network.

This paper focuses on classifying lung cancer into two categories: Benign and Malignant, based on CT-scan slices of lung nodules. We proposed a two-layer convolutional model. The choice of a two-layer CNN was driven by the need for a balance between model complexity, available data, computational resources, and empirical validation results. This architecture demonstrated robust performance in classifying lung nodules accurately, making it a suitable choice for our specific research. The model parameters are shown in Table 1.

Table 1. Model training parameters.

Parameter	Value
Classes	Benign - Malignant
batch_size	32
Epochs	10
Optimizer	adam
Loss	sparse_categorical_crossentropy
Metrics	Accuracy

The input layer processes a grayscale 50 x 50 pixels image using a 3 x 3 convolution matrix with 16 filters, applying ReLU activation. The second layer employs bi-dimensional max-pooling, halving the input's rows and columns while preserving maximum values within 2 x 2 squares. The network then flattens the bi-dimensional data into a single row, consolidating pixel information. The flattened data is passed to a dense layer with 64 neurons, using ReLU activation, followed by another dense layer with 32 neurons and ReLU activation. A Dropout layer with a 0.2 dropout rate is added for regularization. Finally, the output layer consists of a dense layer with 2 neurons, employing SoftMax activation for classification.

Rationale behind the two-layer CNN:

The complexity of the data: Considering the complexity of lung nodule data, a more complex network might not necessarily lead to better results, especially with limited data. A simpler architecture reduces the risk of overfitting, ensuring that the model generalizes well to unseen data.

Data size and complexity: Our datasets, LIDC-IDRI and LUNA16, while comprehensive, are of moderate size. A more intricate network might require a vast amount of data to train effectively. By opting for a two-layer CNN, we strike a balance between model complexity and dataset size, ensuring efficient use of available information.

Empirical Validation: Through iterative experimentation, we observed that our two-layer CNN achieved satisfactory accuracy and convergence within a reasonable training time.

Extensive validation using both LIDC-IDRI and LUNA16 datasets confirmed the effectiveness of our model.

3.4. Fourth phase: training model

In this step, the pre-processed data was transferred to the deep learning architecture of convolutional neural networks to find the patterns in the CT scan images. Using the dropout layer helped reduce the overfitting problem. Therefore, we specified a value of 0.2 for the dropout layer.

3.4.1. Hardware used

The training and validation processes were carried out on a computer device with a Core i5-7400 CPU at 3.00 GHz main memory with a capacity of 8 GB and an NVIDIA GeForce GT-710 graphics card. We used Python version 3.7.5, TensorFlow version 2.11.0, and Keras version 2.11.1 to implement a convolutional neural network.

3.5. Fifth phase: classification

After training and validation processes, we can use this model to classify nodules. We give the image of the nodules to the model for classification.

4. Results

In this section, we present the implementation results and conduct a comparative analysis of the proposed method against various approaches, including the optimal solution, using Python.

4.1. Epochs

We used 50 training epochs; after 10 epochs, the accuracy of the model did not increase, so we set the number of training epochs to 10 and achieved good accuracy.

Table 2. Increasing the model accuracy and decreasing the model loss in each training epoch in the LIDC-IDRI dataset.

Epochs	Loss	Accuracy	Val_loss	Val_accuracy
Epoch 1/10	0.2786	0.8790	0.0251	0.9963
Epoch 2/10	0.0533	0.9813	0.0170	0.9956
Epoch 3/10	0.0262	0.9898	0.0586	0.9815
Epoch 4/10	0.0165	0.9949	0.0158	0.9963
Epoch 5/10	0.0137	0.9959	0.0121	0.9978
Epoch 6/10	0.0164	0.9940	0.0127	0.9970
Epoch 7/10	0.0055	0.9990	0.0170	0.9948
Epoch 8/10	0.0061	0.9984	0.0182	0.9963
Epoch 9/10	0.0035	0.9990	0.0125	0.9970
Epoch 10/10	0.0053	0.9984	0.0165	0.9970

Curves of increasing model accuracy and decreasing model loss according to Table 2 in the LIDC-IDRI dataset is shown in Figures 10 and 11.
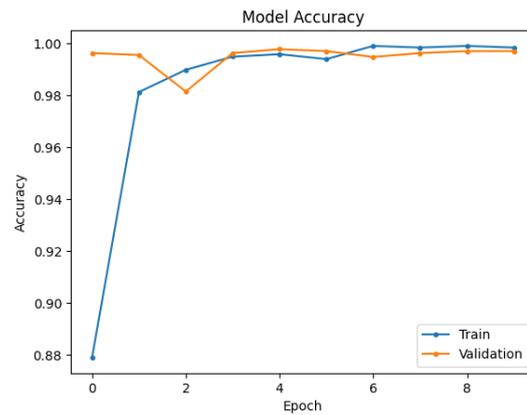

Figure 10. The curve of increasing the model accuracy in each training epoch in the LIDC-IDRI dataset.
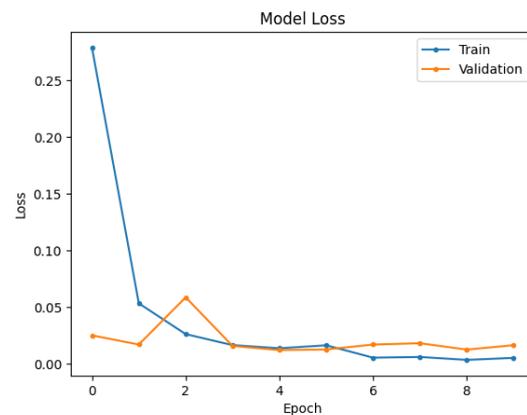

Figure 11. The curve of decreasing the model loss in each training epoch in the LIDC-IDRI dataset.

Table 3. Increasing the model accuracy and decreasing the model loss in each training epoch in the LUNA16 dataset.

Epochs	Loss	Accuracy	Val_loss	Val_accuracy
Epoch 1/10	0.3137	0.8739	0.1424	0.9414
Epoch 2/10	0.1108	0.9607	0.1241	0.9557
Epoch 3/10	0.0774	0.9730	0.1179	0.9602
Epoch 4/10	0.0711	0.9743	0.0992	0.9632
Epoch 5/10	0.0638	0.9765	0.1037	0.9587
Epoch 6/10	0.0557	0.9797	0.0910	0.9655
Epoch 7/10	0.0535	0.9781	0.0905	0.9647
Epoch 8/10	0.0451	0.9855	0.1172	0.9595
Epoch 9/10	0.0413	0.9868	0.0925	0.9707
Epoch 10/10	0.0361	0.9878	0.0864	0.9752

Curves of increasing model accuracy and decreasing model loss according to Table 3, the LUNA16 dataset is shown in Figures 12 and 13.
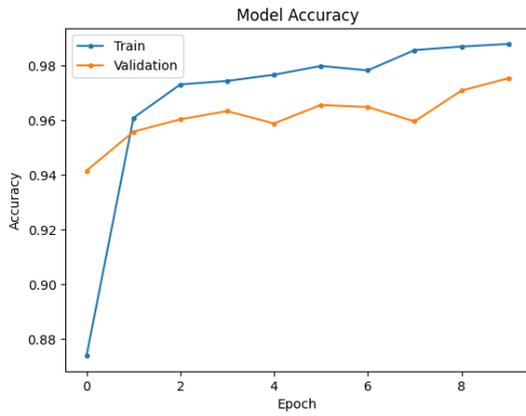
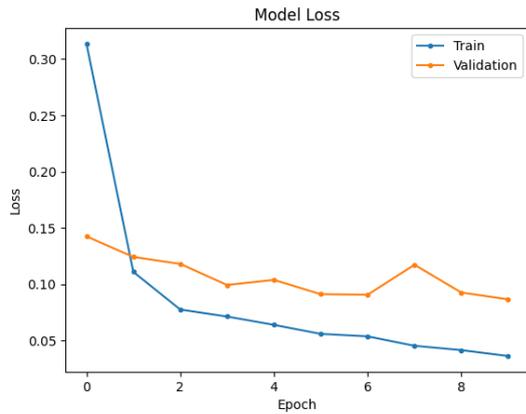**Figure 12. The curve of increasing the model accuracy in each training epoch in the LUNA16 dataset.**

Figure 13. The curve of decreasing the model loss in each training epoch in the LUNA16 dataset.

4.2. Performance metrics

The proposed model's performance was assessed using a range of metrics including accuracy, recall, precision, F1-score, and the ROC curve. A confusion matrix was employed for each dataset to facilitate these evaluations.

Accuracy, defined by Equation (1), represents the proportion of correctly predicted examples out of the total number of examples [27].

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (1)$$

Recall or sensitivity, defined by Equation (2), indicates the true positive rate and is calculated as Tp divided by $(Tp + Fn)$ [27].

$$\text{Re} \, call(Sensitivity) = \frac{Tp}{Tp + Fn} \quad (2)$$

Specificity, outlined in Equation (3), represents the true negative rate and is calculated as Tn divided by $(Tn + Fp)$ [27].

$$Specificity = \frac{Tn}{Tn + Fp} \quad (3)$$

As expressed in Equation (4), precision is the positive predictive value, denoting the number of samples correctly predicted as positive out of the total samples predicted as positive [27].

$$\text{Pr} \, ecision = \frac{Tp}{Tp + Fp} \quad (4)$$

The F1-score is given by Equation (5). It helps assess the model's balance between precision and recall, considering both false positives and false negatives [27, 28].

$$F1 = 2 \times \frac{\text{Pr} \, ecision \times \text{Re} \, call}{\text{Pr} \, ecision + \text{Re} \, call} = \frac{2TP}{2TP + Fp + Fn} \quad (5)$$

Furthermore, the ROC curve is utilized to visually demonstrate the performance of each model as shown in Equation (6). In the ROC curve, the vertical axis represents the True Positive Rate (TPR), while the horizontal axis represents the False Positive Rate (FPR). The Area Under Curve (AUC) is a metric that quantifies the performance of the ROC curve. It is defined as the area under the curve formed by connecting points on the ROC curve, denoted as $\{(x_1, y_2), (x_2, y_2),..., (x_m, y_m)\}$. A higher AUC value indicates better model performance in distinguishing between classes [28].

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i).(y_i + y_{i+1}) \quad (6)$$

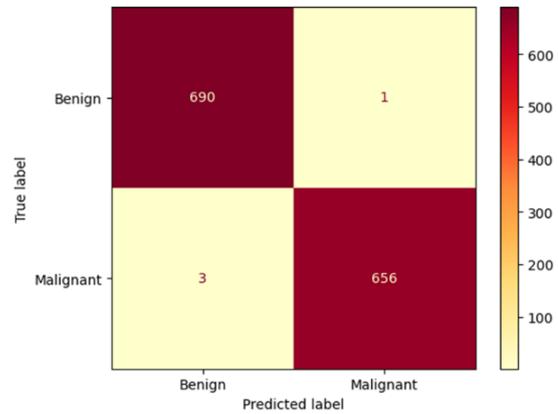## 4.2.1. Evaluation of model performance on the LIDC-IDRI dataset

Figure 14. Confusion matrix of the proposed model in the LIDC-IDRI dataset.

According to Figure 14, 690 samples belonged to the benign class, and were accurately identified as such, while 656 samples were from the malignant class and were correctly identified as members of that class. 1 sample was a member of the benign class but was wrongly diagnosed as a member of the malignant class. 3 samples were members of the malignant class but were wrongly diagnosed as members of the benign class. In this classification, we had two classes, and the classification results for the LIDC-IDRI dataset are detailed in Table 4.

The ROC curve of the LIDC-IDRI dataset is shown in Figure 15.

Table 4. Model evaluation result on the LIDC-IDRI dataset.

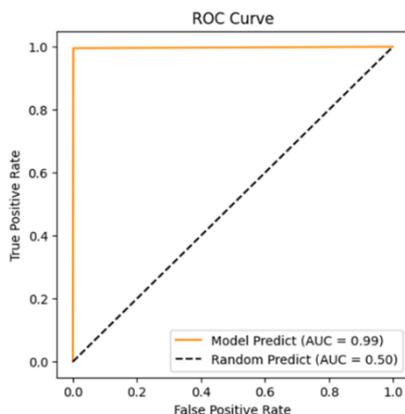| Performance metrics	Result
Accuracy	0.9970
Recall	0.9954
Specificity	0.9985
Precision	0.9984
F1-score	0.9969

Figure 15. Receiver Operating Characteristic (ROC) Curve on the LIDC-IDRI dataset.

According to Figure 15, the black line represents the random guess. A random guess would give a point along a black diagonal line. Points above the diagonal line represent good classification results (better than random) and points below the diagonal line represent bad classification results (worse than random). The orange line shows that our model does classification well. Table 5 shows the comparison between the proposed method and state-of-the-art papers that have employed the same dataset for classification purposes.

Table 5. Performance comparison of state-of-the-art models on the LIDC-IDRI dataset.

Paper	Method	ACC (%)
Al-Shabi *et al.* [10]	Gated dilated 2D CNN	93
Shen *et al.* [11]	Hierarchical semantic 3D CNN	84
Ali *et al.* [12]	Transferable texture 2D CNN	97
Lin *et al.* [13]	Taguchi optimized 2D CNN	99
Liu *et al.* [14]	Multi-model 3D CNN	90
Kareem *et al.* [15]	Support Vector Machine (SVM)	89
Zhao *et al.* [16]	Multi-scale multi-task 3D CNN	94
Halder *et al.* [17]	Two-path morphological 2D CNN	96
Jena *et al.* [18]	Region-based 2D CNN	88
Yu *et al.* [19]	3D ResNet50	87
Proposed model		**99.7**

As mentioned in Table 5, proposed method compared to other models, achieved better accuracy.

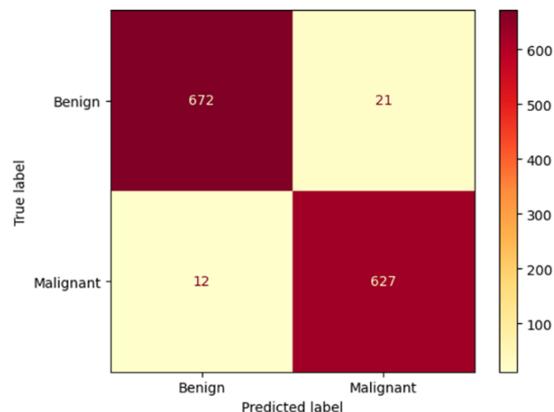### 4.2.2. Evaluation of model performance on LUNA16 dataset

Figure 16. Confusion matrix of proposed model in the LUNA16 dataset.

According to Figure 16, 672 samples belonged to the benign class, and were accurately identified as such, while 627 samples were from the malignant class and were correctly identified as members of that class. 21 samples were members of the benign class but were wrongly diagnosed as members of the malignant class. 12 samples were members of the malignant class but were wrongly diagnosed as members of the benign class.

In this classification, we had two classes, and the classification results for the LUNA16 dataset are detailed in Table 6.

Table 6. Model evaluation result on the LUNA16 dataset.

Performance metrics	Result
Accuracy	0.9752
Recall	0.9812
Specificity	0.9696
Precision	0.9675
F1-score	0.9743

The ROC curve of the LUNA16 dataset is shown in Figure 17.

According to Figure 17, the black line represents the random guess. A random guess would give a point along a black diagonal line. Points above the diagonal line represent good classification results (better than random) and points below the diagonal line represent bad classification results (worse than random). The orange line shows that our model does classification well. Table 7 shows the comparison between the proposed method and state-of-the-art papers that have employed the same dataset for classification purposes. As mentioned in

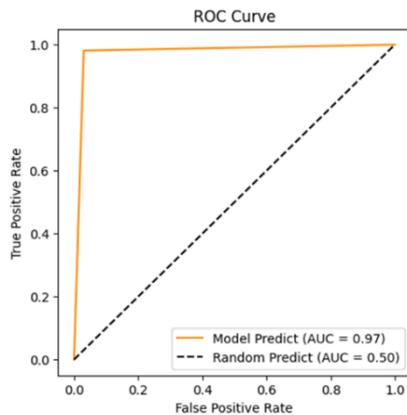Table 7, proposed method compared to other models, achieved better accuracy.

Figure 17. Receiver Operating Characteristic (ROC) Curve on the LUNA16 dataset.

Table 7. Performance comparison of state-of-the-art models on the LUNA16 dataset.

Paper	Method	ACC
Liao *et al.* [20]	A two modules 3D deep U-net-like network	0.8142
Utkin *et al.* [21]	An ensemble of triple neural networks	0.918
Zhang *et al.* [22]	Squeeze-and-excitation network along with aggregated residual transformations (SE-ResNeXt) modules	0.9167
Yuan *et al.* [23]	A 3D convolutional network	0.956
Zhang *et al.* [24]	DenseNet architecture comprising 3D filters and pooling kernels	0.924
Mastouri *et al.* [25]	Bilinear CNN (BCNN) structures combined with a linear SVM	0.9199
Proposed model		**0.9752**

As shown in Tables 8 and 9, a more complex network might not necessarily lead to better results, but we were able to obtain better accuracy with a simpler model. In addition, the training time of our model is much less than the complex models and it uses less computing resources. Therefore, the proposed model has higher reliability.

Table 8. Comparison of our proposed model with other models in LIDC-IDRI dataset.

Models	Accuracy	Training time (ms)
ResNet-50	98.6	12
LeNet-5	97.8	11
VGG-11	98.1	9
Proposed model	**99.7**	**8**

Table 9. Comparison of our proposed model with other models in LUNA16 dataset.

Models	Accuracy	Training time (ms)
ResNet-50	93.8	11
LeNet-5	90.4	10
VGG-11	92.67	9
Proposed model	**97.52**	**7**

5. Conclusion

Lung cancer is a serious disease that requires the attention of doctors and researchers to find solutions for early and accurate diagnosis of this disease to save patients' lives because early diagnosis of this disease helps to cure it and prevents the death of the patient. A CT scan is very important in diagnosing this disease in its early stages, but manual analysis of CT scans is a very tedious and time-consuming task, which puts a heavy burden on radiologists and doctors. Therefore, the use of computer-aided diagnosis systems is essential for automatically classifying CT scan images. In the recent years, complex architectures have been used to design, train, and develop artificial intelligence applications. The problem is that they require powerful hardware and take a lot of time to run. In this paper, we proposed a light and accurate algorithm that can classify nodules with high accuracy without the need for advanced hardware. We separately trained the proposed model on two LIDC-IDRI and LUNA16 datasets, achieving 99.7% and 97.52% accuracy, respectively. The classification results show that our proposed model achieved higher accuracy than similar papers and can help doctors perform the classification more efficiently and accurately than in manual mode.

References

[1] C. Gu, C. Dai, X. Shi, Z. Wu, and C. Chen, "A cloud-based deep learning model in heterogeneous data integration system for lung cancer detection in medical industry 4.0" *Journal of Industrial Information Integration*, vol. 30, November 2022.

[2] M. Astaraki, Y. Zakko, I. T. Dasu, Ö. Smedby, and C. Wang, "Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features" *Journal of Medical Physics*, vol. 83, pp. 146-153, March 2021.

[3] W. Cao, R. Wu, G. Cao, and Z. He, "A Comprehensive Review of Computer-Aided Diagnosis of Pulmonary Nodules based on Computed Tomography Scans" *IEEE Journals & Magazines*, vol. 8, pp. 154007-154023, August 2020.

[4] J. Civit-Masot, A. Bañuls-Beaterio, M. Domínguez-Morales, M. Rivas-Pérez, L. Muñoz-Saavedra, and J. M. Ro. Corral, "Non-small cell lung cancer diagnosis aid

with histopathological images using Explainable Deep Learning techniques" *Computer Methods and Programs in Biomedicine*, vol. 226, November 2022.

[5] W. Huang and L. Hu, "Using a Noisy U-Net for Detecting Lung Nodule Candidates" *IEEE Journals & Magazines*, vol. 7, pp. 67905-67915, May 2019.

[6] M. Li, X. Ma, C. Chen, Y. Yuan, S. Zhang, Z. Yan, C. Chen, F. Chen, Y. Bai, P. Zhou, X. Lv, and M. Ma, "Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes based on Histopathological Images" *IEEE Journals & Magazines*, vol. 9, pp. 53687-53707, April 2021.

[7] S. Mehmood, T. M. Ghazal, M. A. Khan, M. Zubair, M. T. Naseem, T. Faiz, and M. Ahmad, "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing" *IEEE Journals & Magazines*, vol. 10, pp. 25657-25668, February 2022.

[8] M. Entezari, M. Ghanbarirad, A. Taheriazam, M. Sadrkhanloo, A. Zabolian, M. A. Shekhi Beig Goharrizi, K. Hushmandi, A. R. Aref, M. Ashrafizadeh, A. Zarrabi, N. Nabavi, N. Rabiee, M. Hashemi, and S. Samarghandian, "Long non-coding RNAs and exosomal lncRNAs: Potential functions in lung cancer progression, drug resistance and tumor microenvironment remodeling" *Biomedicine & Pharmacotherapy*, vol. 150, June 2022.

[9] X. Cui, S. Zheng, M. A. Heuvelmans, Y. Du, G. Sidorenkov, S. Fan, Y. Li, Y. Xie, Z. Zhu, M. D. Dorrius, Y. Zhao, R. N. J. Veldhuis, G. H. de Bock, M. Oudkerk, P. M. A. van Ooijen, R. Vliegenthart, and Z. Ye, "Performance of a deep learning-based lung nodule detection system as an alternative reader in a Chinese lung cancer screening program" *Journal of Radiology*, vol. 146, January 2022.

[10] M. Al-Shabi, H. K. Lee, and M. Tan, "Gated-Dilated Networks for Lung Nodule Classification in CT Scans" *IEEE Journals & Magazines*, vol. 7, pp. 178827-178838, December 2019.

[11] S. Shen, S. X Han, D. R Aberle, A. A Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification" *Expert Systems with Applications*, vol. 128, pp. 84-95, August 2019.

[12] I. Ali, M. Muzammil, I. Ul Haq, A. A. Khaliq, and S. Abdullah, "Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network" *IEEE Journals & Magazines*, vol. 8, pp. 175859-175870, September 2020.

[13] C. Lin, S. Jeng, and M. Chen, "Using 2D CNN with Taguchi Parametric Optimization for Lung Cancer Recognition from CT Images" *journal applied sciences*, vol. 10, April 2020.

[14] H. Liu, H. Cao, E. Song, G. Ma, X. Xu, R. Jin, C. Liu, and C. Hung, "Multi-model Ensemble Learning Architecture based on 3D CNN for Lung Nodule

Malignancy Suspiciousness Classification" *Journal of Digital Imaging*, vol. 33, pp. 1242–1256, June 2020.

[15] H. F. Kareem, M. S AL-Huseiny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset" *The Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 21, pp. 1731-1738, March 2021.

[16] J. Zhao, C. Zhang, D. Li, and J. Niu, "Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules" *Journal of Digital Imaging*, vol. 33, pp. 869–878, April 2020.

[17] A. Halder, S. Chatterjee, and D. Dey, "Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification" *Biomedical Signal Processing and Control*, vol. 72, February 2021.

[18] S. R. Jena, S. Thomas George, and D. Narain Ponraj, "Lung cancer detection and classification with DGMM-RBCNN technique" *Neural Computing and Applications*, vol. 33, pp. 15601–15617, June 2021.

[19] H. Yu, J. Li, L. Zhang, Y. Cao, X. Yu, and J. Sun, "Design of lung nodules segmentation and recognition algorithm based on deep learning" *BMC Bioinformatics*, vol. 22, November 2021.

[20] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network" *IEEE Journals & Magazines*, vol. 30, pp. 3484-3495, February 2019.

[21] L. Utkin, A. Meldo, M. Kovalev, and E. Kasimov, "An Ensemble of Triplet Neural Networks for Differential Diagnostics of Lung Cancer" *IEEE Journals & Magazines*, February 2020.

[22] G. Zhang, Z. Yang, L. Gong, S. Jiang, L. Wang, and H. Zhang, "Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations" *La radiologia medica*, vol. 125, pp. 374–383, January 2020.

[23] H. Yuan, Z. Fan, Y. Wu, and J. Cheng, "An efficient multi-path 3D convolutional neural network for false-positive reduction of pulmonary nodule detection" *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. 2269–2277, August 2021.

[24] G. Zhang, L. Lin, and J. Wang, "Lung Nodule Classification in CT Images Using 3D DenseNet" *Journal of Physics*, vol. 1827, January 2021.

[25] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, "A bilinear convolutional neural network for lung nodules classification on CT images" *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. 91–101, November 2020.

[26] S. Modak, E. Abdel-Raheem, and L. Rueda, "Applications of deep learning in disease diagnosis of

chest radiographs: A survey on materials and methods" *Biomedical Engineering Advances*, vol. 5, June 2023.

[27] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases" *Computers in Biology and Medicine*, vol. 132, May 2021.

[28] H. Jiang, S. Tang, W. Liu, and Y. Zhang, "Deep learning for COVID-19 chest CT (computed tomography) image analysis: A lesson from lung cancer" *Computational and Structural Biotechnology Journal*, vol. 19, pp. 1391-1399, 2021.

استفاده از شبکه عصبی کانولوشن برای افزایش دقت طبقه‌بندی توده‌های سرطانی ریه با استفاده از تصاویر سی‌تی‌اسکن

محمد مهدی نخعی۱، ساسان کرمی زاده۱،*، محمد ابراهیم شیری احمد آبادی۲ و کامبیز بدیع۳

۱ موسسه آموزش عالی ارشاد دماوند، تهران، ایران.

۲ دانشگاه صنعتی امیرکبیر، تهران، ایران.

۳ پژوهشکده فناوری اطلاعات و ارتباطات، تهران، ایران.

چکیده:

سرطان ریه یک بیماری بسیار جدی است و تشخیص زودهنگام سلول‌های سرطانی به طور قابل توجهی شانس بهبودی بیماران را افزایش می‌دهد. پزشکان به طور مرتب تعداد زیادی از تصاویر سی‌تی‌اسکن را بررسی می‌کنند که می‌تواند منجر به ایجاد خستگی و اشتباه شود. بنابراین، نیاز به ایجاد ابزاری وجود دارد که بتواند به طور خودکار ندول‌های ریه را در مراحل اولیه شناسایی و طبقه‌بندی کند. سیستم‌های تشخیص به کمک رایانه که اغلب از تکنیک‌های پردازش تصویر و یادگیری ماشین استفاده می‌کنند، به رادیولوژیست‌ها در شناسایی و طبقه‌بندی این گره‌ها کمک می‌کنند. مطالعات قبلی اغلب از مدل‌های پیچیده یا شبکه‌های از پیش آموزش‌دیده استفاده کردند که نیاز به سخت‌افزار قوی و زمان طولانی برای اجرا دارند. هدف ما دستیابی به تشخیص دقیق بدون نیاز به سیستم محاسباتی قدرتمند است. ما یک شبکه عصبی کانولوشنال ساده را با تنها دو لایه پیچشی معرفی می‌کنیم که قادر به طبقه‌بندی دقیق گره‌ها بدون نیاز به سخت‌افزار قدرتمند است. ما فرایندهای آموزش و اعتبارسنجی را بر روی دو مجموعه داده LIDC-IDRI و LUNA16 انجام دادیم که به ترتیب به دقت‌های ۹۹٫۷ درصد و ۹۷٫۵۲ درصد رسیدیم. این نتایج، دقت برتر مدل پیشنهادی ما را در مقایسه با مقالات خوب گذشته نشان می‌دهد.

کلمات کلیدی: سرطان ریه، یادگیری عمیق، LIDC-IDRI، LUNA16، چرخانده شده.