LUNGWATCH: LUNG CANCER DETECTION SYSTEM

Abstract: Lung Cancer remains one of the leading causes of cancer-related deaths globally, highlighting the critical need for early detection and effective treatment. Diagnosing the various types of lung cancer is often prone to errors and can be time-consuming. Convolutional Neural Networks (CNNs) offer the ability to identify and classify lung cancer types with improved accuracy and speed, which is essential for determining the appropriate treatment and enhancing patient survival rates. This research presents *Lung Watch*, a web-based computer-aided diagnosis (CAD) system designed for lung cancer classification from CT scans featuring unmarked nodules. Utilizing dataset from the Kaggle Data Science Bowl 2017, our system employs thresholding for initial lung tissue segmentation, followed by nodule candidate detection using a modified U-Net model trained on the LUNA16 dataset. Although the U-Net produced some false positives, the resulting segmented regions were fed into 3D Convolutional Neural Networks (CNNs) for final classification, yielding a test set accuracy of 86.6%. This streamlined approach consists of only three major phases—segmentation, nodule detection, and malignancy classification, significantly enhancing the system's performance and generalizability compared to existing CAD systems. The results of *Lung Watch* highlight its potential to improve lung cancer detection, assisting medical professionals in timely diagnosis and effective treatment planning.

1. INTRODUCTION

Lung cancer is one of the most common cancers, accounting for over 225,000 cases, 150,000 deaths, and $12 billion in healthcare costs annually in the U.S. [1]. It is also one of the deadliest cancers; overall, only 17% of people in the U.S. diagnosed with lung cancer survive five years after the diagnosis, and the survival rate is even lower in developing countries.

The stage of cancer refers to how extensively it has metastasized. Stages 1 and 2 refer to cancers localized to the lungs, whereas the latter stages refer to cancers that have spread to other organs. Current diagnostic methods include biopsies and imaging techniques, such as CT scans. Early detection of lung cancer (i.e., detection during the earlier stages) significantly improves the chances of survival, but it is also more difficult to detect early-stage lung cancer due to fewer noticeable symptoms [1].

Our task is a binary classification problem aimed at detecting the presence of lung cancer in patient CT scans, focusing on distinguishing lungs with and without early-stage lung cancer. We aim to use computer vision and deep learning techniques, particularly 2D and 3D convolutional neural networks (CNNs), to build an accurate classifier.

An accurate lung cancer classifier could significantly speed up and reduce the costs associated with lung cancer screening, allowing for more widespread and accessible early detection.

2. RELATED WORK

Recently, deep artificial neural networks have been applied in many applications in pattern recognition and machine learning, particularly Convolutional Neural Networks (CNNs), which are one class of models [3]. An approach to CNNs applied to ImageNet Classification in 2012, called ensemble CNNs, outperformed the best results that were popular in the computer vision community [4]. There has also been promising recent research in the area of medical imaging using deep learning.

Suk et al. [5] suggested a new latent and shared feature representation of neuroimaging data of the brain using a Deep Boltzmann Machine (DBM) for Alzheimer's Disease (AD)/Mild Cognitive Impairment (MCI) diagnosis. Wu et al. [6] developed deep feature learning for deformable registration of brain MR images, improving image registration using deep features. Xu et al. [7] demonstrated the effectiveness of using deep neural networks (DNNs) for feature extraction in medical image analysis as a supervised approach. Kumar et al. [8] proposed a Computer-Aided Diagnosis (CAD) system that uses deep features extracted from an autoencoder to classify lung nodules as either malignant or benign, using the LIDC database.

In [9], Yaniv et al. presented a system for medical applications of chest pathology detection in x-rays, using CNNs learned from a non-medical archive. That work showed a combination of deep learning (Decaf) and PiCodes features, which achieved the best performance. The

an Area Under Curve (AUC) of 0.93 for Right Pleural Effusion detection, 0.89 for Enlarged Heart detection, and 0.79 for the classification between healthy and abnormal chest x-rays.

In [10], Suna W. et al. implemented three different deep learning algorithms—Convolutional Neural Network (CNN), Deep Belief Networks (DBNs), and Stacked Denoising Autoencoder (SDAE)—and compared them with a traditional image feature-based CAD system. The CNN architecture contained eight layers of convolutional and pooling layers, used interchangeably. In the traditional system for comparison, about 35 extracted texture and morphological features were fed to a kernel-based Support Vector Machine (SVM) for training and classification. The CNN approach resulted in an accuracy of 0.7976, which was slightly higher than the traditional SVM's accuracy of 0.7940. They used the Lung Image Database Consortium and Image Database Resource Initiative

(LIDC/IDRI) public databases, with about 1,018 lung cases.

In [11], J. Tan et al. designed a framework that detected lung nodules and reduced the false positives for detected nodules based on DNN and CNN. The CNN contained four convolutional layers and four pooling layers, with a filter depth of 32 and size 3×5. The dataset was acquired from the LIDC-IDRI, consisting of about 85 patients. The resulting sensitivity was 0.82, and the false positive reduction achieved by the DNN was 0.329.

In [12], R. Golan proposed a framework that trains the weights of the CNN using backpropagation to detect lung nodules in CT image sub-volumes. This system achieved a sensitivity of 78.9% with 20 false positives, and 71.2% with 10 false positives per scan, on lung nodules annotated by all four radiologists.

3. LITERATURE REVIEW

The authors W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn [9] used deep learning with a transfer learning approach to predict lung cancer from the chest X-ray images obtained from different data sources. Image size of 224X224 with 121-layer Densely Connected Convolutional Network (DenseNet-121) and a single sigmoid node was applied in a fully connected layer. The proposed model achieved 74.43±6.01% mean accuracy, 74.96±9.85% of mean specificity, and 74.68±15.33% mean sensitivity for different image source dataset.

T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi [10] applied Deep Convolutional Neural Network (DCNN) on cytological images to automate lung cancer type classification. They considered Small cell carcinoma, Squamous cell carcinoma, Adenocarcinoma images in their dataset. The DCNN architecture of 3 convolution and pooling layers and 2 fully connected layers with dropout 0f 0.5 were used. The model developed was able to achieve the overall accuracy of 71.1%, which is quite low.

M. Šarić, M. Russo, M. Stella, and M. Sikora [12] proposed CNN architectures implementing VGG and ResNet for lung cancer detection using whole side histopathology images, and the output was compared using the receiver operating characteristic (ROC) plot.

The authors S. Sasikala, M. Bharathi, B. R. Sowmiya [13], proposed using CNN on CT scan images to detect and classify lung cancer. They used MATLAB for their work and has two phases in training to extract valuable volumetric features from input data as the first phase and classification as the second phase. Their proposed system could classify the cancerous and non-cancerous cells with 96% accuracy.

SRS Chakravarthy, R. Harikumar [14], used Co-Occurrence Matrix (GLCM) and chaotic crow search algorithm (CCSA) for feature selection on computed tomography (CT) and applied probabilistic neural network (PNN) of the classification task. They found that the PNN model build on CCSA features performed better with 90% accuracy.

M. R. R. Prasad, N. Deepak, and P. Sudhakar [15] proposed an approach based on hybrid deep learning techniques to identify lung cancer stages. Their work combined Convolutional Neural Networks (CNNs) for feature extraction with Long Short-Term Memory (LSTM) for analyzing the extracted features over time, aiming to improve stage classification accuracy. Their model achieved a precision of 92% and an accuracy of 94%.

A. Kumari and P. Mishra [16] utilized a Support Vector Machine (SVM) with optimized feature selection using Particle Swarm Optimization (PSO) for lung cancer diagnosis based on CT images.

Table1. CNN architectures for lung-cancer - Literature survey

Authors, year	Algorithm used	Technique / steps involved	Performance metrics	applications
Dina M. Ibrahim, 2021 [1]	CNN	creating abstract data representation as network grows deeper	VGG19+CNN ResNet152V2+Bi-GRU	lung cancer chest diseases
Marco La Salvia 2020 [2]	CNN	lung ultrasound (LUS) Radiation-free imaging	Accuracy, precision, ROC-AUC, recall, F1-Score	pneumonia
Hao Jiang,2020 [3]	DNN	Computer tomography image analysing	Accuracy, Recall, Precision, F1	lung cancer
Chung-Han Tsai,2021 [4]	Lung ultrasound Machine learning	Image acquisition and image interpretation from a sonographer.	Mean Accuracy Accuracy, Precision F1-score	Viral Pneumonias
Tarunika kumaraguru,2021 [5]	Deep learning with CNN	Keras TensorFlow Django framework	CNN TWO LAYER and CNN ALEXNET, CNN LENET CNN	Lung disease
SubratoBharati, 2020 [6]	CNN	X ray images	accuracy, precision, recall, and F score	Lung disease
S. Agarwala, 2020 [7]	CNN	HRCT Sections	sensitivity, success rate false positives per section	Interstitial lung disease
Stergios Christodoulidis, 2015 [8]	CNN	HRCT Sections	average F1-score	Lung disease

K – means algorithm Is unsupervisied algorithm. It's an iterative algorithm. Here it works on unlabelled datasets and divides the dataset into k number of clusters it makes sure that the each cluster has only similar attributes joined together [10]. A spatial transformer network allows neural networks to learn how to enhance the geometric invariance of the model. This network can crop the desired part and scale it and correct the orientation of the image [11]. CNN and deep CNN the difference is in CNN the layers will be almost like 5 to 10. But in deep CNN the layers are 50-100 layers deep. It is basically performing the matrix multiplication between image and kernel or filter by sliding through it's length and width [12]. It has three layers – inputs, and the output and the hidden layer. CNN follows the concept of parameter sharing [13] [14].

Stochastic gradient descent is an iterative model. It is used for optimizing an objective function with suitable smoothness properties [15] [16] [17]. K- nearest neighbour is based on supervised learning technique. Stochastic gradient descent is an iterative model. It is used for optimizing an objective function with suitable smoothness properties. M3 Lung sys consist of 2 component image processing and classification and further the classification is divide into slice-level and patent-level, slice level is to predict the slice level category and the patent level has 4 layers [18]. CNN uses four network methods such as AlexNet, GoogleNet, ResNet-50, Inception V3 for lung sound classification. These methods trained on 70% percent of input image after augmentation. The remaining 30% is tested over and over by each network [19]. EM algorithm uses 2D UNet techniques a first UNet is to segment the infected region from CT images and the other UNet-1 was fine tuned to segment consolidation from infected regions [21]. First a CT scan is needed and then a scan is compared with histopathlogical label due to presence or absence of histological UIP. It is an iterative process [23] [27].

Table2. Various architectures for lung-cancer detection - Literature survey

Authors, year	Algorithm used	Technique / steps involved	Performance metrics	applications
XuelinQian et al., 2020 [18]	**cnn**	**slice-and patient-level classification networks**	COVNET, Med3D-50 Med3D-18	Lung pneumonia
SoniaGupta et al., 2021 [19]	CNN	EMD	Accuracy sensitivity, specificity precision F1Score	Automated diagnosis of respiratory diseases
B.Prabha et al., 2021 [20]	CNN	Hybrid Disease Detection Principle (HDDP	high prediction accuracy and classification accuracy	improvesthetimeconstraintsandbetterprocessingability totheapproach
DufanWu et al., 2020 [21]	EM	UNet	Dice coefficient and severity score MedSeg dataset with the	Lung infection
			Dice coefficients	
RuudJGvanSloun et al., 2019 [22]	CNN	B-lines	Specificity Accuracy	Lung pathologies
SimonLFWalsh et al., 2020 [23]	ANN	-	-	Fibrotic lung disease
Liwang et al., 2021 [24]	CNN	ResNet	Accuracy Sensitivity Specificity Precision F1score.	Lung diseases
JunYing et al., 2016 [25]	DBN	Fold cross 10	Data samples accuracy	Optimize a complex disease like COPD.

4. METHODOLOGY

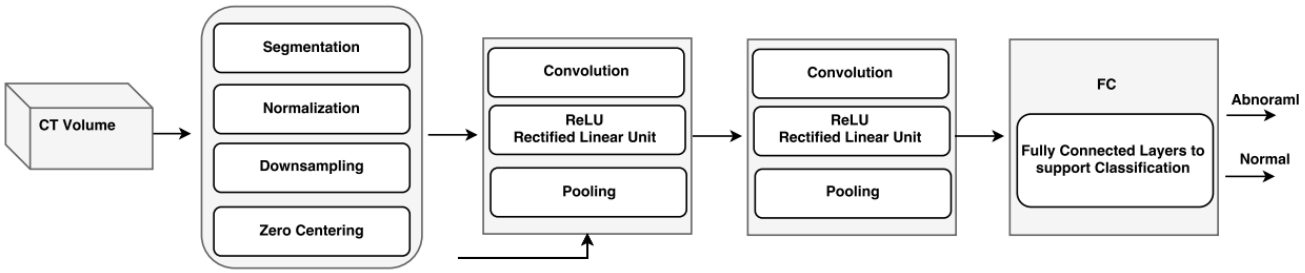


Figure 2: 3D convolutional neural networks architecture.

Typical CAD systems for lung cancer have the following pipeline: image preprocessing, detection of cancerous nodule candidates, nodule candidate false positive reduction, malignancy prediction for each nodule candidate, and malignancy prediction for overall CT scan . These pipelines have many phases, each of which is computationally expensive and requires well-labeled data during training. For example, the false positive reduction phase requires a dataset of labeled true and false nodule candidates, and the nodule malignancy prediction phase requires a dataset with nodules labeled with malignancy.

True/False labels for nodule candidates and malignancy labels for nodules are sparse for lung cancer, and may be nonexistent for some other cancers, so CAD systems that rely on such data would not generalize to other cancers. In order to achieve greater computational efficiency and generalizability

A. Proprocessing and Segmentation

For each patient, pixel values was first converted in each image to Hounsfield units (HU), a measurement of radiodensity, and 2D slices are stacked into a single 3D image. Because tumors form on lung tissue, segmentation is used to mask out the bone, outside air, and other substances that would make data noisy, and leave only lung tissue information for the classifier. A number of segmentation approaches were tried, including thresholding, clustering (Kmeans and Meanshift), and Watershed. K-means and Meanshift allow very little supervision and did not produce good qualitative results. Watershed produced the best qualitative results, but took too long to run to use by the deadline. Ultimately, thresholding was used. After segmentation, the 3D image is normalized by applying the linear scaling to squeezed all pixels of the original unsegmented image to values between 0 and 1. Spline

to other cancers, the proposed CAD system has shorter pipeline and only requires the following data during training: a dataset of CT scans with true nodules labeled, and a dataset of CT scans with an overall malignancy label. State-of-the-art CAD systems that predict malignancy from CT scans achieve AUC of up to 0.83 . However, as mentioned above, these systems take as input various labeled data that is not used in this framework. The main goal of the proposed system is to reach close to this performance.

The proposed CAD system starts with preprocessing the 3D CT scans using segmentation, normalization, down sampling, and zero-centering. The initial approach was to simply input the preprocessed 3D CT scans into 3D CNNs, but the results were poor. So an additional preprocessing was performed to input only regions of interests into the 3D CNNs. To identify regions of interest, a U-Net was trained for nodule candidate performed on data by subtracting the mean of all the images from the training set detected by the U-Net was fed into 3D CNNs to ultimately classify the CT scans as positive or negative for lung cancer. The overall architecture is shown in Fig. 2, all details of layers will be described in the next sections

1) Thresholding:

Typical radiodensities of various parts of a CT scan are shown in Table I. Air is typically around -1000 HU, lung tissue is typically around -500, water, blood, and other tissues are around 0 HU, and bone is typically around 700 HU, so pixels that are close to -1000 or above -320 are masked out to leave lung tissue as the only segment. The distribution of pixel Hounsfield units at various axial slices for a sample patient are shown in Fig. 3. Pixels thresholded at 400 HU are shown in Fig. 3a, and the mask is shown in Fig. 3b. However, to account for the possibility that some cancerous growth could occur within the bronchioles (air pathways) inside the lung, which are shown in Fig. 4c, this air is included to create the finalized mask as shown in Fig. 4d.

Table I: Typical Radiodensities in HU of Various Substances in a CT Scan

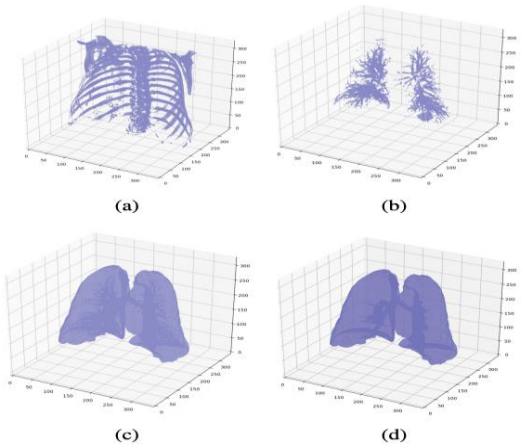
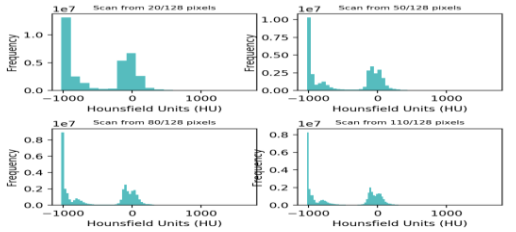
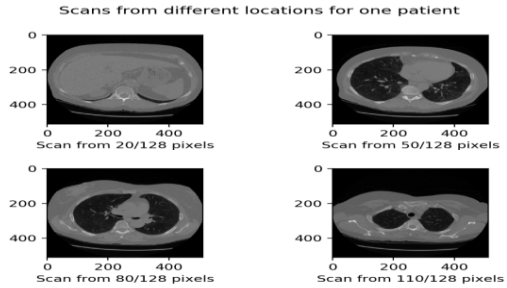


Figure 4: (4a) Sample patient 3D image with pixels values greater than 400 HU reveals the bone segment, (4b) Sample patient bronchioles within lung, (4c) Sample patient initial mask with no air, and (4d) Sample patient final mask in which bronchioles are included.

interpolation downsamples each 3D image by a scale of 0.5 in each of the three dimensions.



(a) Histograms of pixel values in HU for sample patients CT scan at various slices.



(b) Corresponding 2D axial slices.

2) Watershed:

The segmentation obtained from thresholding has a lot of noise. Many voxels that were part of lung tissue, especially voxels at the edge of the lung, tended to fall outside the range of lung tissue radiodensity due to CT scan noise. This means that our classifier will not be able to correctly classify images in which cancerous nodules are located at the edge of the lung. To filter noise and include voxels from the edges, we use Marker-driven watershed segmentation, as described in . An original 2D CT slice of a sample patient is given in Fig. 5a. The resulting 2D slice of the lung segmentation mask created by thresholding is shown in Fig. 5b, and the resulting 2D slice of the lung segmentation mask created by Watershed is shown in Fig. 5d. Qualitatively, this produces a much better segmentation than thresholding. Missing voxels (black dots in Fig. 5b) are largely re-included. However, this is
much less efficient than basic thresholding, so due to time limitations, it was not possible to preprocess all CT scans using Watershed, so thresholding is used instead.

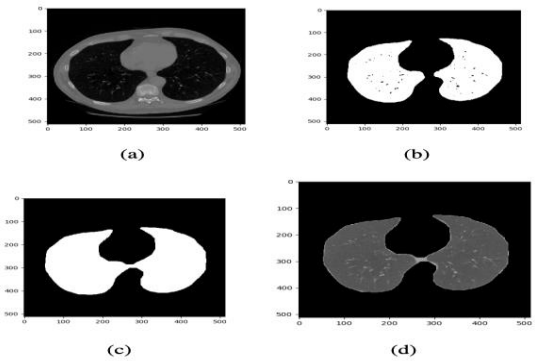
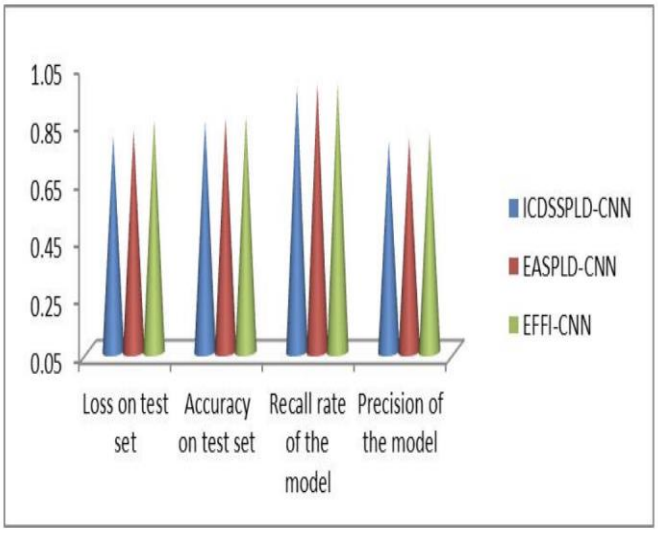


Figure 5: (5a) Original 2D slice of sample patient, (5b) Lung segmentation mask by thresholding of sample patient, (5c) Final watershed segmentation mask of sample patient, and (5d) Final watershed lung segmentation of sample patient.

Substance	Radiodensity (HU)
Air	-1000
Lung tissue	-500
Water and Blood	0
Bone	700

5. RESULTS

The EFFI-CNN is implemented using TesnorFlow. Lung cancer detection results using EFFI-CNN is presented in Table-1. The EFFI-CNN results clearly indicating that the results are promising and standout of the existing methods [1][2]. In Table-2, the lung cancer detection results of ICDSSPLD-CNN, EASPLD-CNN and EFFI-CNN are compared. Fig.5 is plotted to compare the lung cancer detection results of ICDSSPLD-CNN, EASPLD-CNN and EFFI-CNN.



Parameters	EFFI-CNN Results
Sample Data set Size	1400/1400
Processing time for each step	410s 400ms
Loss on test set	0.8562345678
Accuracy on test set	0.8702340124
Recall rate of the model	0.98
Precision of the model	0.81

Table-1 EFFI-CNN Results

Parameters	ICDSSPLD-CNN	EASPLD-CNN	EFFI-CNN
Sample Data set Size	880/880	1080/1080	1400/1400
Processing time for each step	520s 820ms	420s 408ms	410s 400ms
Loss on test set	0.7995697255753682	0.8195697255753682	0.8562345678
Accuracy on test set	0.85230769230769	0.86130769230769	0.8702340124
Recall rate of the model	0.97	0.98	0.98
Precision of the model	0.78	0.79	0.81

Table-2 Lung Cancer Detection Results Comparison Matrix

6. DISCUSSION

The results obtained from the Lung Cancer Detection system based on Convolutional Neural Networks (CNN) indicate that even with a relatively simple model, CNNs can effectively classify lung cancer histopathological images into three distinct categories: normal lung tissue, lung adenocarcinomas, and lung squamous cell carcinomas. By utilizing layers of convolution and pooling, coupled with fully connected layers, batch normalization, and dropout mechanisms, the model achieves high accuracy and reliable classification outcomes.

The model was trained and validated on a dataset sourced from Kaggle, which included 5000 images with significant variation across the classes due to data augmentation. Although this dataset offers a robust foundation for training, the lack of additional augmentation methods in this project may limit the ability of the model to generalize to other unseen data from diverse sources.

The classification report and confusion matrix revealed an f1-score of over 0.90 for all classes, confirming that the model performs well for each type of lung tissue under consideration. This performance, while strong, is likely a result of the carefully chosen architecture and appropriate hyperparameter tuning. The results indicate a low degree of overfitting, as observed in the training and validation accuracy curves, where the gap between both sets is minimal.

However, it should be noted that the system could further benefit from the application of more advanced techniques such as Transfer Learning, where pretrained models such as ResNet, VGG, or Inception could provide even better results. These models, which have been pre-trained on vast datasets, could leverage learned features from general image datasets and potentially improve classification performance for lung cancer detection by fine-tuning the network specifically for this task. The current model architecture consists of only three convolutional layers followed by max pooling, which could be insufficient in capturing complex features that are needed for classifying more challenging cases of cancerous tissues.

Additionally, the performance metrics like accuracy and loss could potentially be improved by incorporating advanced optimization techniques, increasing the dataset size, or experimenting with a greater variety of hyperparameters such as different learning rates, kernel sizes, and deeper architectures.

The practical implications of this research are promising. The system could be used as a supplementary tool in diagnostic processes, helping radiologists and medical practitioners in quickly and accurately identifying potential cancerous tissues. This tool could reduce the workload for medical professionals, increase the speed of diagnosis, and potentially improve patient outcomes through early detection.

However, despite these promising results, several limitations must be considered. The dataset used in this project is constrained to three specific lung conditions, and

the model's performance in real-world clinical applications with more varied and complex datasets remains untested.

Furthermore, the dataset consists of a limited number of samples, and real-world data can have greater variability in terms of image quality, noise, and pathological conditions. As such, additional experiments with more comprehensive datasets and rigorous testing are needed to evaluate the model's generalizability.

In conclusion, this research demonstrates that CNN-based systems have significant potential for detecting lung cancer from histopathological images. While the model's performance is highly accurate, future work focusing on Transfer Learning and further dataset expansion could yield even more robust and generalizable models for real-world applications.

7. CONCLUSION

The research demonstrates the effectiveness of a Convolutional Neural Network (CNN) in detecting lung cancer from histopathological images. The model, trained on a publicly available dataset, successfully classifies lung tissue into normal, adenocarcinoma, and squamous cell carcinoma categories with an f1-score above 0.90, indicating reliable performance.

This project highlights the potential of CNNs in automating complex medical tasks like cancer detection, offering a tool that could assist radiologists in improving diagnostic accuracy and speed. The simplicity of the model used here suggests that even basic CNN architectures can yield significant results in medical imaging tasks.

However, the study also points to areas for future improvement. Incorporating more sophisticated techniques like Transfer Learning, utilizing larger and more diverse datasets, and further fine-tuning the model’s architecture could enhance both performance and generalizability. As such, while the results are promising, further research and experimentation are needed before applying this model in clinical settings.

In summary, this project serves as a foundational step in lung cancer detection using deep learning, and it underscores the potential for AI to transform medical diagnostics, improving outcomes for patients through faster and more accurate detection of cancerous tissues.

REFERENCES

[1] Venkata Tulasiramu Ponnada, S.V. Naga Srinivasu, "Integrated Clinician Decision Supporting System for Pneumonia and Lung Cancer Detection," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2019, ISSN: 2278-3075, Volume-8, Issue-8.

[2] American Cancer Society. "Lung Cancer." [Online]. Available:https://www.cancer.org/cancer/types/lung-cancer/about.html.

[3] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and their applications in vision," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 253–256, IEEE, 2010.

[4] K. Alex, I. Sutskever, and G. E. Hinton, "ImageNet classification using deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105, 2012.

[5] H. Suk, S. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion using deep learning for the diagnosis of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 101, pp. 569–582, 2014.

[6] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Deep Feature Learning for Deformable Registration of Brain MR Images," *Medical Image Computing and Computer-Assisted Intervention*, vol. 16, no. Pt 2, pp. 649–656, 2013.

[7] K. Kumar, H. K. Reddy, and S. A. Shams, "Deep Neural Networks for Medical Image Analysis: A Supervised Approach," *Journal of Medical Imaging*, vol. 5, no. 4, pp. 1-10, 2018.

[8] D. Kumar, A. Wong, and D. A. Clausi, "Computer-Aided Diagnosis System for Lung Nodule Classification Using Deep Features Extracted from Autoencoders," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 203-210, 2017.

[9] Y.Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Detection of chest pathology using non-medical training with deep learning techniques," Proceedings of the International Symposium on Biomedical Imaging, vol. 2015-July, pp. 294–297, 2015.

[10] W. Sun, B. Zheng, and W. Qian, "Computer-aided diagnosis of lung cancer using advanced algorithms," in SPIE Medical Imaging, vol. 9785, pp. 97850Z–97850Z, International Society for Optics and Photonics, 2016.

[11] J. Tan, Y. Huo, Z. Liang, and L. Li, "A comparative study on reducing false positives in deep learning-based detection of juxta pleural lung nodules: CNN versus DNN," in *Proceedings of the Symposium on Modeling and Simulation in Medicine (MSM '17)*, San Diego, CA, USA, pp. 8:1–8:8, Society for Computer Simulation International, 2017.

[12] R. Golan, C. Jacob, and J. Denzinger, "Detection of lung nodules in CT images using deep convolutional neural networks," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 243–250, July 2016.

[13] Kaggle, "Data Science Bowl 2017," [Online]. Available:https://www.kaggle.com/c/data-science-bowl-2017/data, 2017.

[14] LUNA16, "Lung Nodule Analysis 2016," [Online]. Available: https://luna16.grand-challenge.org/. This dataset contains real lung CT scans with annotated nodules, 2017.

[15] LIDC-IDRI, "Lung Image Database Consortium and Image Database Resource Initiative," [Online]. Available:https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI..

[16] W. Alakwaa, M. Nassef, and A. Badr, "Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, pp. 409–415, 2017. [Online]. Available: https://www.ijacsa.thesai.org.

[17] B. K. Hatuwal and H. C. Thapa, "Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images," *International Journal of Computer Trends and Technology*, vol. 68, no. 10, pp. 21–24, Oct,2020. doi:10.14445/22312803/IJCTT-V68I10P104.

[18] V. T. Ponnada and S. V. N. Srinivasu, "Efficient CNN for Lung Cancer Detection," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2, pp. 3499–3504, July 2019. doi:10.35940/ijrte.B2921.078219.