Enhanced Lung Cancer Detection Using Deep Learning Ensemble Approach

Sireesha Moturi, Manvitha Ainavolu, Naga Revathi Dokku, Prathima Kasula,
Neelima Yaragani, Venkatareddy Dodda
Department of Computer Science and Engineering,
Narasaraopeta Engineering College, Palnadu Dist., Andhra Pradesh, India
sireeshamoturi@gmail.com, manvithaainavolu@gmail.com, nagarevathidokku@gmail.com,
prathimakasula5@gmail.com, neelimayaragani@gmail.com, oddavenkatareddy@gmail.com

2024 First International Conference for Women in Computing (InCoWoCo) | 979-8-3315-1894-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCOWOCO64194.2024.10863245

***Abstract*—In emerging technologies, deep learning has become the solution for most real-world problems. This is used by integrating with computer science and predicting many diseases for people at an early stage. One of the major issues is the detection of Lung Cancer either Cancerous or Non-Cancerous. Many studies have proposed the solutions to detect the Lung Nodule. However, using an ensemble approach by taking CT(Computed Tomography) Scan images gives accurate results when compared to other methodologies. In this study, we proposed the methodology by combining three deep learning models instead of one. Thereby, the performance of three Convolutional Neural Network models is combined to get accurate results. The dataset used for this is LUNA16 from Grand Challenge which is available online. This dataset consists of CT scans and an annotations file which gives the information about each CT Scan. Our Ensemble 2D CNN model gave an effectiveness of 96% which is greater than the baseline methodology.**

***Index Terms*—Deep Learning, Lung Cancer, Computed Tomography Scan, 2D Convolutional Neural Networks, Lung Nodule Analysis 2016, Ensemble Approach, Cancerous Nodules, Non-Cancerous Nodules.**

I. Introduction

Lung Cancer has become the most frequent disease in these days. Many people in the world are suffering from this problem for not identifying this disease early. The development of technology has also increased many solutions in the medical industry. However, no solution detects lung cancer through the individual symptoms. Mostly detecting through medical images is a quite difficult task to identify whether the person is prone to cancer or not. The United States has registered the greatest number of cases worldwide. In the year 2018, the GLOBOCAN estimated that nearly 2 million new cases were registered due to lung cancer.[1]

Deep learning has transformed many sectors of the economy, most importantly medical imaging, providing an automated and much more accurate means for identifying such hard-to-diagnose conditions as lung cancer. Deep learning method called Convolutional Neural Networks (CNNs) holds a highly promising ability in interpreting CT scans and determining if the lung nodule is malignant or not. This will give a promising result to find the state of a person quickly [2]. A CNN can be trained to recognize patterns in large medical image datasets that are usually invisible to the human eye. Such a CNN trained with thousands of CT scans can be trained to discern subtle differences between benign and malignant nodules, offering an essential early detection approach.

For example, a small nodule might not give any symptoms in the patient, so clinical evaluation would be pointless. Only through CT scans, can these nodules be seen by radiologists. Or interpretation of images by humans is a time-consuming subjective activity and even leads to errors. Small benign nodules, for example, are often mistaken for malignant or overlooked, leading to false positives and unnecessary treatments. Another important task in classifying these images is handling three-dimensional images. The medical images are visualized under 3D view to analyze clearly. So to handle these types of images needs to be efficient in identifying them. DL approaches introduce Neural networks which will learn the patterns of images in all dimensions. Deeper CNNs like GoogleNet, AlexNet, and ResNet [3] became advanced in analyzing the input patterns. From this, we have chosen CNNs as it has a multi-view architecture in processing input images. However, there are some disadvantages to this. Only certain features of lung nodules are identified, which leads to incorrect prognosis and incorrect classification [4].

In this study, we have improved the CNN to classify the lung Nodules. For this purpose, we have chosen LUNA16 from the Grand Challenge. It is an important task when learning the data for the model. Neural networks work like the human brain, so training the data for the model is crucial because it plays a valuable role in the model's accuracy and performance.[5] Here the ensemble approach is developed since it is tough to distinguish between a nodule and a tissue. The main goal in this is to combine the efficiencies of three CNN models and give the result accurately.

II. Related Work

A study conducted in 2018 revealed that a deep learning system could point out overlooked tumors of the lung on chest X-rays, reducing unnecessary CT scans caused by it. Previous studies documented a deep learning-based auto-detection approach to reduce ignored lung tumors on chest radiography.[6] Another study also used 2D CNNs and

Taguchi parametric optimization, but the marginal adjustment of the important parameters increased in this context to improve the percentages of identification of lung cancer from the CT images [7].

The other explained how to take the advantage of the benefits by combining several deep learning models such as CNNs and RNNs.[8] CNNs and deep neural networks were found to have great performances for malignant lung nodule classification with a log loss of 0.387732.

A variant of work introduced a bio-inspired algorithm combined with CNN for lung tumor detection. Using the WOAAPSO algorithm in this hybrid mechanism gave good results with 97.18 % accuracy, sensitivity of 97%, and specificity up to 98.66 standard [10]. The combined results provided evidence that CT over CAD enhances detection, though even the current methods are at fault-perfect to this day still requiring 100% accuracy[11]

Some reviews say that AlexNet, when combined with various classifiers, offers accuracy as high in the detection of lung cancer, while some say that VGG-16, ResNet, and Inception provide the best performance.[12] Deep learning methods, especially CNN, also performed well in classifying lung cancer subtypes, such as adenocarcinoma and squamous cell carcinoma, from pathology images. In some cases, these reached 0.97 AUC, almost the same accuracy as human pathologists could achieve [13]

One of the research study proposes the ensemble model that combines the outputs of different deep-learning architectures: CNNs and RNNs. The strengths of several models are combined using the ensemble approach in order to enhance robustness and generalizability.[14] Some literature also shows that the addition of various types of classifiers along with AlexNet improves the detection accuracy in lung cancer, although various other studies mention that models such as VGG-16, ResNet, and Inception provide the best performance; the accuracy may even reach 100% in some. On the whole, it has been observed that the CNN-based architecture Inception V3 gave very impressive results[15]

III. Proposed Method

The use of group learning methods has not been relatively easy in previous studies. In this work, we present the use of three different 2D CNN ensembles to identify lung cancer using computed tomography scans. Each of these CNNs was designed to capture a host of features by having different architectural depths and filter sizes. Each model will be trained separately with the aim of increasing the general accuracy of prediction by subsequently combining them using soft voting. A soft voting mechanism would apply to combine the predictions of each model in order to achieve high overall prediction accuracy. The method does this by averaging the probability outputs from all the models for a given instance to decide which class contains the highest average probability. This technique enhances generalization on unseen data by reducing model variance while simultaneously increasing the robustness of predictions.

The followings are the steps involved in the recommended approach:

- Data Description
- Pre-processing.
- Data Splitting.
- Creating the Model Architecture.
- Training the Model Train Dataset
- Testing the Model Using the Test Dataset

The Figure-1 shows theflowchart describes the research paradigm in this study:

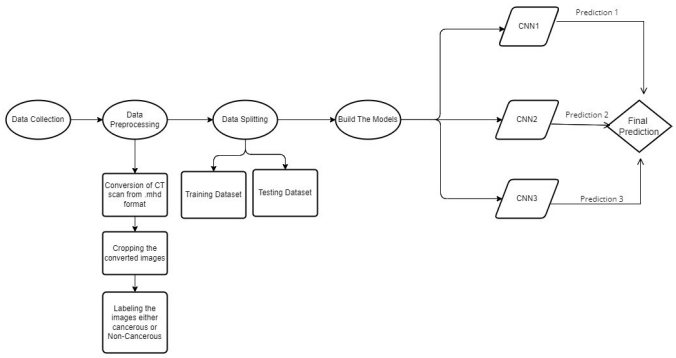

Fig. 1. Flow chart for the Proposed Methodology

A. *Experimental Setup*

All models were trained using the Adam optimizer, and the batch size was set to 32 for all models. The initial learning rate was set to 1×10^{-3}. .All experiments were conducted on an NVIDIA Tesla V100 GPU with 32 GB of RAM. On average, it took approximately 6 hours to train each model. The softmax outputs of the three models were averaged to apply the ensemble approach.

B. *Data Collection*

Lung Nodule Analysis Dataset 2016 (LUNA16) is an important source for investigators involved in lung nodule detection. This dataset is a subset of a larger database, the LIDC/IDRI, which includes 1,018 CT scans that were reviewed by four expert radiologists.[4] The LIDC/IDRI database classifies the nodules into three classes: nodules 3 mm or larger, nodules smaller than 3 mm, and non-nodules. Last, it comprises annotations indicating the centroid locations and the diameters of detected pulmonary nodules for all of its chest CT images counting to 888. To keep results consistent, users should use these subsets in 10-fold cross-validation. Each of these contains CT images along with several key files such as annotations.csv (reference annotations), sampleSubmission.csv (an example submission format), candidates.csv (basic candidate sites), candidates_V2.csv (extended candidate sites), evaluation scripts, lung segmentation data, additional annotations.csv.

The Dataset can be accessed from LUNA16(Lung Nodule Analysis 2016) Grand Challenge, it has become a widely used source for research, especially on pulmonary nodule detection and related fields.

C. Data Preprocessing

The dataset used in this study is the LUNA16 sourced from Grand Challenge 2016. This dataset contains data about lung cancer with the CT scan images in .mhd format.

- .mhd files contain image resolution, pixel spacing, and data type metadata. This information correctly interprets the associated .raw file, which contains the raw pixel data.
- This is done with libraries such as SimpleITK. The images after converting from .mhd to JPEG format are visualized, and Figure 2 shows the images

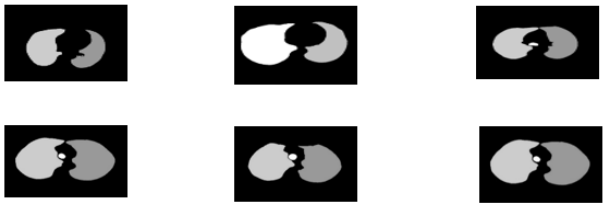

Fig. 2. CT scan images before Pre processing

- Once the image is reconstructed, the image can be saved in .JPEG format using a standard image processing library PIL (Python Imaging Library).After converting the images from MHD to JPEG format and applying normalization.
- The images in JPEG format are cropped in the region
- based on the coordinates given in the annotations.csv file, which gives information about each CT scan. Figure 3 shows the preprocessed images.

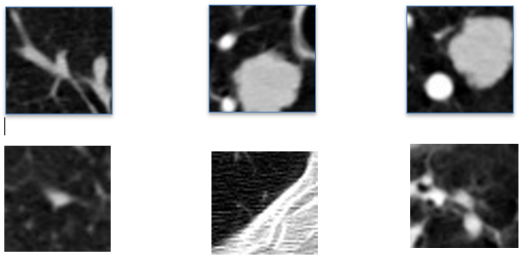

Fig. 3. Cropped CT Scan images

D. Splitting the Dataset

Creating train and test folders out of the dataset is the next step. The candidates.csv file's class column is used for the splitting process. Each CT scan is represented in this column according to its cancerous(1) or non-cancerous(0) status. The dataset is divided into two distinct sub-folders, one labeled "Cancerous" and the other "Non-Cancerous," within the data folder. The model learns about the data using the train set of data. The real data that we use to evaluate the predictions and accuracy of the model is called test data.

E. Model Architecture and Training

These are the model architecture construction and training phases, which are critical upon completion of the splitting process. The Ensemble 2D CNN algorithm is the one we have selected. We have trained three 2D CNN models in order to apply this Ensemble approach, having a few layers, and join the predictions of models in order to obtain the ensemble outcomes. The description of the three model model architectures is as follows:

- **Ensemble 2D CNN**

The building of model architectures for the three CNN networks are described below.

CNN1: In designing the first model of the convolutional neural network, a three-convolutional-block architecture was followed. In each convolution block, an additional batch normalization layer, max-p pooling layer, and dropout layer were added. Among them, the convolution layers use the ReLU activation function and size filters of 32, 64, and 128, respectively, with a 3x3 kernel size. This will enable the 'he_uniform' initializer to optimize weights for ReLU activation and set weights to small, random values. Following flattening, the model adds a fully linked layer with 128 units using ReLU as an activation function. Next comes the thick layer with softmax activation that gives class probabilities. It uses Adam optimizer for optimization and focused loss while training to handle the issue of class imbalance.

CNN2: The second model deepens the architecture by the introduction of a fourth convolution block, scaling the depth of features to be extracted from 64 to 512 filters. Each convolution block introduces a batch normalization layer, and the next layers are the max-pooling layer and a dropout layer. The dropout rate increases progressively from 0.2 to 0.5. The activation function remains ReLU for hidden layers, with the initializer as 'he_uniform' for optimizing weight distribution. The fully connected ending layer is of dimension 256 units, the activation ReLU. The softmax function will be used for the final output layer to give class probabilities. In this model, Focal loss is similar to the first CNN model.

CNN3: The third model is a bit more complex, which will define a CNN architecture with three convolutional blocks followed by batch normalization, ReLU, and dropout layers, respectively. The first two convolution blocks will contain 32 and 64 filters, each with kernel sizes 3×3 and 5×5, respectively, while the third block has 128 with kernel size 3×3.Kernel size 3×3. After that, the model is terminated by a flattening layer, which is followed by a fully connected layer of units 84, activated by ReLU. Lastly, for output, softmax activation has been used to handle binary classification problems.

Each model produced a probability vector after training over the probability of the input image being in either of the two classes. During inference, these vectors from all three models contain the operations, they are represented as

- **Convolution Operation**
  This applies small filters to the input image. It is mathematically represented as:
  $$O(i,j) = \sum_{m=U}^{F-1} \sum_{n=U}^{F-1} W(a+g, b+h) \cdot K(g,h) \quad (1)$$
  Where:
  - F is the size of the filter
  - W is the input image
  - K is the convolutional kernel
  - O denotes the output feature map
  - g, h are the indices for the kernel
  - a, b are the spatial indices of the feature map
- **ReLU Activation Function**
  This introduces non-linearity to the model and mitigates the vanishing gradient problem. It is defined as:
  $$\text{ReLU}(y) = \max(0, y) \quad (2)$$
  Where:
  - y is the argument of the ReLU function
- **Max Pooling**
  This is used to downsample the feature maps. It is denoted as:
  $$\text{Max Pooling}(p,q) = \max(\text{Window}(p,q)) \quad (3)$$
  Where:
  - p, q represent the position or index of the window in the feature map
- **Fully Connected Layer (Dense Layer)**
  This is the final stage of feature learning. It is defined as:
  $$P(y=c|z) = \frac{e^{z_c}}{\sum_{k=1}^{C} e^{z_k}} \quad (4)$$
  Where:
  - C is the number of classes
  - z_c is the logit for class c
- **Focal Loss**
  This function helps handle class imbalance during training. It is given by:
  $$FL(a_b) = -\alpha_t (1-a_b)^{\gamma} \log(a_b) \quad (5)$$
  Where:
  - a_b is the predicted probability
  - α_t is the weighting factor
  - γ is the focusing parameter
- **He Uniform Initialization**
  This is also known as He Uniform Initialization, a weighting technique, defined as:
  $$\text{Limit} = \sqrt{\frac{6}{n_{in}}} \quad (6)$$
  Where:
  - n_{in} is the number of input units to the layer
- **Soft Voting**
  The predictions from the three models are combined using soft voting, defined as:
  $$P_c = \frac{1}{M} \sum_{m=1}^{M} P_{m,c} \quad (7)$$
  Where:
  - M is the number of models
  - $P_{m,c}$ is the predicted probability for model m and class c

1) 3D CNN

: In the architecture of the third CNN, there will be a series of convolutional blocks, each of which gradually extracts increasingly intricate information from input images. Each of these has three main components: convolutional layers, dropout and batch normalization, and ReLU activation is applied at every stage.

The first convolutional block comes with a kernel size of 3×3 and 32 filters. It captures detailed low-level information such as edges and textures. The second convolutional block has the increased kernel size to 5×5, and the number of filters is increased up to 64. This additional block assists in more complex structure and pattern detection. In the final third convolutional block, it increases the number of filters to 128 with a 3×3 kernel size to focus more on even finer and detailed features.

The output of such convolutional layers is passed through a flattening layer that takes this multi-dimensional output and compresses it into a 1D vector. This vector then inputs a fully connected layer of size 84, where the ReLU activation function assists the model to find non-linear relations in data. Ultimately, whether the picture belongs to the first or the second category-for example, cancerous or not, which is the case if we want to classify tissue as either cancerous or non-cancerous-in general, depends on the output of a softmax activation function that produces probabilities for binary classification.

Each block uses dropout to help with overfitting, making certain neurons randomly switch off during training and utilizing the use of batch normalization that normalizes each layer's output in an effort to stabilize the training and speed up the convergence. Combining all these methods helps to improve the generalizing capacity of the model. It is much more helpful while dealing with intricate datasets like medical imaging where it is pretty important to identify even slighter patterns.

F. Experimental Results and Analysis

The study bases several criteria to assess the efficiency of the proposed ensemble deep learning model. These are F1-score, recall, accuracy, and precision. Each of these

measures was selected to evaluate different facets of functionality of the model ensuring a multi-faceted perception of its strengths and weaknesses.

1) Results for ensemble 2D CNN model:: The above model architectures were compiled and trained on the dataset. The results and analysis of the models were described below: **CNN1:** The model's overall effectiveness was quite high, at an accuracy of 95%, but with good precision on non-cancerous cases classification at 0.98. The model is trained for 70 epochs. In the case of the cancerous cases, this recall was higher at 0.91, meaning that more true positives were likely to be caught by the model, but at the expense of a higher rate of false positives. The F1-score for the cancerous class showed an instance balance between precision and recall at 0.82. The Figure 4 and figure 5 shows the results and ROC curves of CNN1 are shown below:

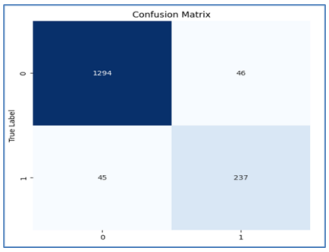

Fig. 4. Confusion matrix for CNN1

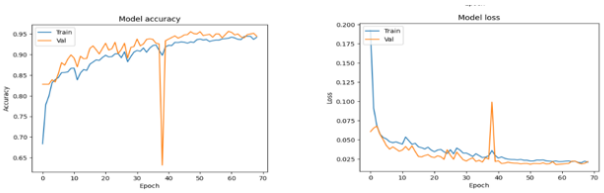

Fig. 5. Accuracy and Loss Curves for CNN1

CNN2 The CNN2 model performed very well in differentiating cancerous from non-cancerous cases, reflected by a testing accuracy of 94.91%. The high precision and recall values for both classes demonstrate that the model is good in predicting cancerous cases. However, it tends to have a bit more false negatives, 45 cases, compared with false positives, 27 cases.

The figure 6 and figure 7 shows the results and ROC curves of CNN2 are shown below:

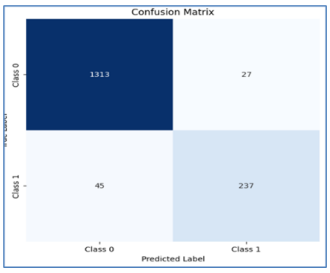

Fig. 6. Confusion matrix for CNN2

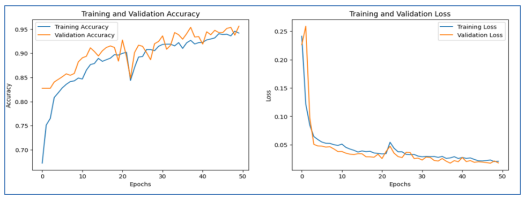

Fig. 7. Accuracy and Loss curves for CNN2

CNN3 Subsequently, the model was used in testing data to estimate the result of the CNN in line with the results obtained from the study. The first model of CNN in iteration shows us an accuracy of 91%. The model was trained with 50 epochs. Figure 8 and Figure 9 shows the the results and ROC curves of CNN3 are shown below:

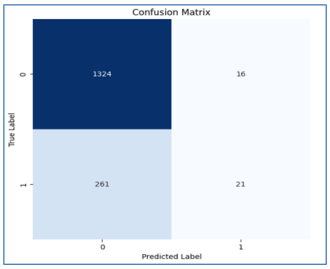

Fig. 8. Confusion Matrix for CNN3

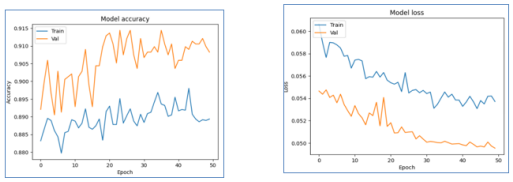

Fig. 9. Accuracy and loss curves for CNN3

Results for ensemble 2D CNN Constructing the ensemble model will combine the probabilities of those three models.Performances will combined because the models were assembled by using different input layers and activation functions.This approach has been improved the performance of the resultant model, as the miscalssification of nodules is reduced.

Three separate models had accuracy values that were around

that point. Once we combined the three models, our accuracy was 96%. This is because an ensemble has the added advantage of making strong, reliable predictions by learning a variety of patterns from the dataset without overfitting.

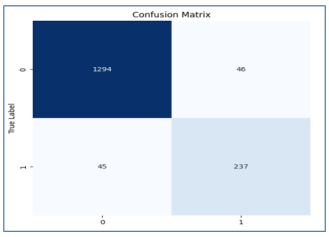

Fig. 10. Confusion Matrix for 2D CNN

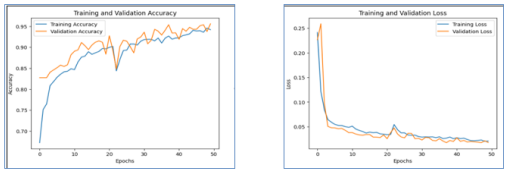

Fig. 11. Accuracy and loss Curves for Ensemble 2D CNN

Classification Report for Model 1:				
	precision	recall	f1-score	support
0	0.97	0.97	0.97	1340
1	0.87	0.86	0.87	282
accuracy			0.95	1622
macro avg	0.92	0.92	0.92	1622
weighted avg	0.95	0.95	0.95	1622

Fig. 12. Classification report for Ensemble 2D CNN

Comparison with 3D CNN model: with the 3D CNN, the accuracy was 90.63% when trained in 50 epochs, indicating that the model correctly classified 91% of the test images. Thus, this may already be a pretty strong performance for a binary classification task in medical imaging, where clinically correct predictions are absolutely essential. Figure 14 and figure 15 shows the results and ROC curves of 3D CNN:

	precision	recall	f1-score	support
0	0.95	0.98	0.97	1340
1	0.91	0.78	0.84	282
accuracy			0.95	1622
macro avg	0.93	0.88	0.90	1622
weighted avg	0.95	0.95	0.95	1622

Fig. 13. Classification Report for 3D CNN

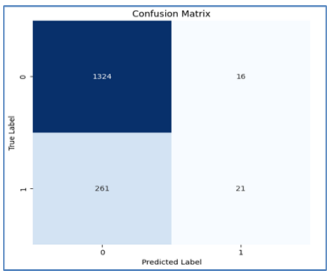

Fig. 14. Confusion matrix for 3D CNN

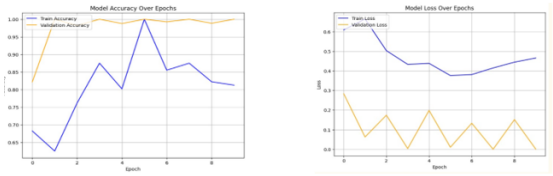

Fig. 15. Accuracy and Loss Curves for 3D CNN

TABLE I
COMPARISON WITH OTHER METHODOLOGIES

Methodology	**Algorithms Used**	**Results**
Taguchi Optimization with 2D CNN [7]	2D CNN, Taguchi Optimizer	94%
Segmentation Method [6]	CNN architecture using segmentation	73%
Adaptive Boosting Algorithm with CNN	CNN, AdaBoost Classifier	90%
2D CNN Ensemble Approach [16]	2D CNN architectures	95%
Proposed Method (**2D CNN Ensemble**)	Three 2D CNN architectures, Ensemble Approach	96%

IV. CONCLUSION

In this work, we propose an ensemble learning technique in the diagnosis of lung cancer via a CT scan image-based approach that relies on three different architectures of 2D CNNs. By using different depths and filter sizes in the creation of each CNN for capturing most of the properties of lung nodules, it ensures more reliability with respect to accuracy in classification. Compared to individual models, ensemble models worked better as classification accuracy improved to 96% by aggregating the predictions of several CNNs through soft voting. Further, it succeeded in handling class imbalance, particularly if Focal Loss is used and hard data to classify are targeted.This work demonstrates that an ensemble of deep learning methods can achieve superior accuracy and robustness in medical image analysis tasks, particularly for more complex applications like lung cancer diagnosis. Improvement in diagnostic performance in the ensemble model is then realized based on its ability to aggregate diverse feature extraction techniques applied differently by different CNNs, which also allows for a better comprehension of the input data.

V. Limitations and Future Scope

The major drawback is that the study makes use of 2D CNNs, which are not good at getting the 3D structure of lung nodules. Generally, lung nodules are three-dimensional; if one applies two-dimensional slices, they may lose important spatial information, which can propagate to affect the performance of the model. Future work in this field should include 3D CNN architectures that may very well increase classification accuracy far more effectively by analyzing volumetric data.

Future work may also explore the application of transfer learning. Transfer learning is accomplished through the employment of pre-trained networks that already learned informative features from large medical image collections, and might be applied for further improvement in model generalization performance on new datasets. The complementing properties of many different classifiers might be exploited for further improvement in model performance by exploring other ensemble methods, including stacking or boosting.

VI. Dataset Reference

Setio, A.A.A., Traverso, A., de Bel, T., Berens, M.S. LUNA16: A challenge for automatic nodule detection in CT scans. Grand Challenge, available at::https://luna16.grand-challenge.org/.

References

[1] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, p. 5569, 2022. [Online]. Available: https://doi.org/10.3390/cancers14225569

[2] M. A. Heuvelmans *et al.*, "Lung cancer prediction by deep learning to identify benign lung nodules," *Lung Cancer*, vol. 154, p. 14, 2021. [Online]. Available: https://doi.org/10.1016/j.lungcan.2021.01.027

[3] Z. Li *et al.*, "Deep learning methods for lung cancer segmentation in whole-slide histopathology images—the ACDC@LUNGp challenge 2019," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 429–440, 2020. [Online]. Available: https://doi.org/10.1109/JBHI.2020.3039741

[4] P. Huang *et al.*, "Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method," *The Lancet Digital Health*, vol. 1, no. 7, pp. e353–e362, 2019. [Online]. Available: https://doi.org/10.1016/j.lungcan.2021.01.027

[5] R. R. Subramanian *et al.*, "Lung cancer prediction using deep learning framework," *Int. J. Control Autom.*, vol. 13, no. 3, pp. 154–160, 2020. [Online]. Available: https://www.researchgate.net/publication/346700750

[6] A. Shimazaki *et al.*, "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," *Sci. Rep.*, vol. 12, no. 1, p. 727, 2022. [Online]. Available: https://doi.org/10.1038/s41598-021-04667-w

[7] C. J. Lin, S. Y. Jeng, and M. K. Chen, "Using 2D CNN with Taguchi parametric optimization for lung cancer recognition from CT images," *Appl. Sci.*, vol. 10, no. 7, p. 2591, 2020. [Online]. Available: https://doi.org/10.3390/app10025991

[8] J. Zhou *et al.*, "An ensemble deep learning model for risk stratification of invasive lung adenocarcinoma using thin-slice CT," *NPJ Digital Med.*, vol. 6, no. 1, p. 119, 2023. [Online]. Available: https://doi.org/10.1038/s41746-023-00866-z

[9] R. Tekade and K. Rajeswari, "Lung cancer detection and classification using deep learning," in *Proc. 2018 Fourth Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Pune, India, 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ICCUBEA.2018.8697352

[10] S. Vijh, P. Gaurav, and H. M. Pandey, "Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection," *Neural Comput. Appl.*, vol. 35, no. 33, pp. 23711–23724, 2023. [Online]. Available: https://doi.org/10.1007/s00521-020-05362-z

[11] S. Makaju *et al.*, "Lung cancer detection using CT scan images," *Procedia Comput. Sci.*, vol. 125, pp. 107–114, 2018. [Online]. Available: https://doi.org/10.1016/j.procs.2017.12.016

[12] N. Kalaivani *et al.*, "Deep learning based lung cancer detection and classification," in *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 994, no. 1, p. 012026, 2020. [Online]. Available: https://doi.org/10.1088/1757-899X/994/1/012026

[13] P. M. Shakeel *et al.*, "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, 2019. [Online]. Available: https://doi.org/10.1016/j.measurement.2019.05.027

[14] S. Moturi *et al.*, "Prediction of liver disease using machine learning algorithms," in *Proc. Int. Conf. Data Sci. Appl.*, Singapore, 2023, pp. 243–254. [Online]. Available: https://doi.org/10.1007/978-981-99-7817-5$_1$9

[15] A. Seva, S. T. Rao, and M. Sireesha, "Prediction of liver disease with random forest classifier through SMOTE-ENN balancing," in *2024 IEEE 13th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, 2024, pp. 928–933. [Online]. Available: https://doi.org/10.1109/CSNT60213.2024.10546170