Deep Hybrid Learning for Accurate Lung Nodule Classification in CT Scans

Vishwas V. Patange
Research Scholar Dept. of E&TC Engineering
Dr. Babasaheb Ambedkar Technological University
Vidyavihar, Lonere, Dist.: Raigad, India

Sanjay L. Nalbalwar
Professor and Head Dept. of E&TC Engineering
Dr. Babasaheb Ambedkar Technological University
Vidyavihar, Lonere, Dist.: Raigad, India

Jagadish B. Jadhav
Assoc. Prof.
Dept. of E&TC Engineering
R.C. Patel Institute of Technology, Shirpur
Dhule, India

Suchitra Shankar Hyam
Sr. MRI and CT Scan Technician B.Sc. in Applied Biomedical Technology
Samartha Diagnostics, Nasik

ABSTRACT
Lung cancer is one of the leading causes of cancer-related deaths worldwide, and early detection is critical for improving patient survival rates. This study presents an advanced CT scan–based image analysis pipeline for the reliable detection and classification of pulmonary nodules. The proposed method combines image preprocessing, segmentation, and feature extraction with deep learning classification to improve accuracy and robustness. We address challenges such as class imbalance, domain shift, and inter-class similarity—particularly between benign and normal cases—by applying targeted augmentation, class-balanced losses, and vessel-suppression techniques. Experimental evaluation on benchmark datasets shows that our approach achieves high accuracy and recall, especially for malignant cases, while minimizing false negatives in benign detection. The results highlight the potential of our method for integration into computer-aided diagnosis systems to support radiologists in clinical decision-making.

General Terms
Machine Learning, Deep Learning, Medical Image Processing, Pattern Recognition

Keywords
CT scan images, CNN, GBDT

1. INTRODUCTION
Lung cancer continues to be one of the leading causes of cancer-related deaths globally, largely due to its frequent diagnosis at advanced stages when treatment effectiveness is significantly reduced. Early detection is vital for enhancing survival rates; however, existing diagnostic approaches—such as imaging and biopsies—are often hindered by limitations in accuracy, availability, cost, and associated patient risks, including exposure to radiation and the invasiveness of certain procedures.

Machine learning (ML) has emerged as a transformative tool in early lung cancer detection, leveraging vast medical datasets, including radiological images, genetic profiles, and clinical histories. Deep learning, in particular, excels at rec- ognizing subtle patterns in data that may be imperceptible to human evaluators, thereby enabling earlier and more precise detection of lung cancer.

This research focuses on developing advanced ML algo- rithms, especially deep learning models, to enhance diagnostic accuracy in distinguishing between benign and malignant lung nodules. Key challenges addressed include model in- terpretability, the necessity of well-annotated datasets, and ethical concerns surrounding patient privacy. The ultimate goal is to create a robust, efficient, and user-friendly ML- driven diagnostic pipeline that improves lung cancer screening, accelerates diagnosis, and minimizes false positives, ultimately contributing to better patient outcomes.

2. LITERATURE REVIEW

2.1 Advancements in Machine Learning for Lung Cancer Detection
Lung cancer detection using computed tomography (CT) scans has been widely explored due to their ability to provide high-resolution imaging for identifying malignant nodules. Traditional diagnostic methods face limitations in accuracy, accessibility, cost, and invasiveness, prompting the adoption of ML and DL for automated diagnosis, nodule classification, and early detection.

2.2 Feature Selection and Machine Learning Classifiers
Feature selection plays a vital role in improving the accu- racy of ML models for lung cancer detection. Syed et al. [1] emphasized the importance of feature selection techniques in distinguishing between benign and malignant lung nodules. Their study mapped features to lung regions and applied ML algorithms, enhancing diagnostic performance. Similarly, Dash et al. [2] proposed a multi-classifier framework integrating the Semantic Network Classifier and Naïve Bayes, demonstrating the advantage of combining classifiers to improve diagnostic precision.

2.3 Comparative Analysis of ML Classifiers
Numerous studies have investigated and compared different machine learning algorithms to determine the most effective models for lung cancer detection. Günaydin et al. [3] assessed classifiers including Principal Component Analysis (PCA), k-Nearest Neighbors (KNN), and Support Vector Machines (SVM), concluding that Artificial Neural Networks (ANNs) and SVMs demonstrate strong potential for clinical use. Similarly, Makaju et al. [4] employed watershed segmentation techniques to improve the detection of small nodules, effectively tackling challenges linked to early-stage lung cancer diagnosis.

2.4 Deep Learning-Based Approaches
Deep learning models have demonstrated superior performance in lung cancer detection by extracting complex features from medical images. Zhang et al. [5] introduced a multi- scene deep learning model incorporating a CNN with four channels, effectively analysing various image structures to improve

detection accuracy. Ausawalaithong et al. [6] imple- mented a 121-layer CNN with transfer learning, achieving 74.43% accuracy, with the added benefit of visual heatmaps for interpretability. Li et al. [7] focused on the classification of lung adenosquamous carcinoma subtypes using histopatho- logical images and feature selection techniques, achieving high accuracy in subtype differentiation. Furthermore, Yu et al. [8] introduced the Adaptive Hierarchical Heuristic Mathematical Model (AHHMM), integrating deep learning with image pre-processing to predict tumour growth and minimize radiation side effects.

2.5 Challenges in Deep Learning for Lung Cancer Detection

Despite significant advancements, deep learning models face challenges such as the need for high-quality annotated datasets, model interpretability, and potential biases in training data. Thanoon et al. [9] reviewed the latest deep learning applications in lung cancer screening, underscoring the necessity of large, annotated datasets for reliable model training. Muntasir et al. [10] performed a comparative analysis of Inception V3, Xception, and ResNet-50 models, achieving 92% accuracy and an AUC of 98.21%, further solidifying the potential of DL-based solutions in lung cancer diagnosis.

3. ADVANCED IMAGE ANALYSIS TECHNIQUES FOR LUNG CANCER DIAGNOSIS

Lung-cancer detection from CT scans in this work follows a rigorous, end-to-end pipeline that couples classical image processing with deep learning and calibrated evaluation. We first assemble diverse, de-identified CT data (Kaggle Rathi and IQ-OTH/NCCD, plus institutional cases where available) and standardize it via lung-windowing, noise suppression, intensity normalization/CLAHE, and resizing to 224×224. Optional lung/nodule segmentation (U-Net/Mask R-CNN or 3D variants) restricts the field of view to clinically relevant regions. Discriminative representations are then derived by combining radiomics cues (morphology/texture) with transfer-learned deep features from VGG-16; features are standardized and, when useful, compacted with PCA or mutual-information selection. To counter class imbalance and improve minority recall, we apply targeted augmentation and SMOTE at the feature level. Final classification uses a Gradient-Boosted Decision Tree (GBDT) head on the deep embeddings, which consistently sharpens class separation versus a softmax layer. Evaluation is patient-wise and stratified, includes cross-dataset tests to probe domain shift, and reports Accuracy, Precision, Recall, F1, ROC–AUC, confusion matrices, and (when required) calibrated probabilities for threshold-based clinical triage.

4. SYSTEM ARCHITECTURE

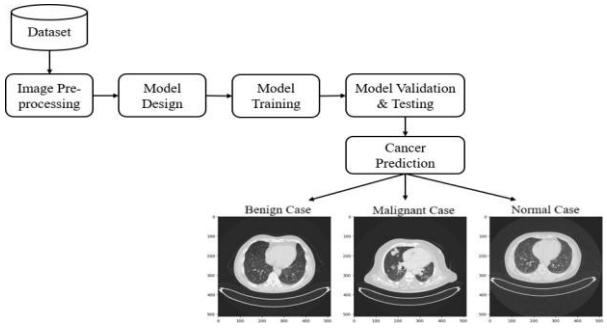

Fig. 1: System architecture diagram of lung cancer prediction

4.1 Dataset

We assemble thoracic CT images from public repositories (e.g., the Kaggle "CT Scan Images for Lung Cancer" set by Dishant Rathi and the IQ-OTH/NCCD dataset by Aditya Mahimkar) and, where permitted, institutional cases. All scans are de-identified, checked for duplicates, and screened to remove corrupt or trivially non-lung images. To limit information leakage, we split data at the patient/study level, not at the slice level, and we keep sources separated when reporting cross-dataset generalization. Basic metadata (slice thickness, kernel, scanner) is retained to audit domain shift.

4.2 Image Pre-processing

Incoming volumes or slices are standardized to a common format and intensity range before modeling. We apply lung-windowing, denoise with Gaussian filtering or anisotropic diffusion, normalize intensities to [0,1], and enhance contrast with histogram equalization/CLAHE. Each image is resized to 224×224 pixels to match network inputs and, when present, axes/borders are cropped to avoid non-anatomical cues. Optional lung masking or coarse nodule segmentation limits the field of view to parenchyma. Data augmentation (small rotations, zoom, flips, crops) increases robustness, and class imbalance is mitigated later with SMOTE in feature space.

4.3 Model Design

The backbone is a transfer-learned VGG-16 CNN that serves purely as a feature extractor. We initialize with ImageNet weights, freeze early layers, and fine-tune later blocks on the CT domain. Deep features are aggregated (flatten/GAP) into a compact embedding; instead of a SoftMax head, we feed these embeddings to a Gradient Boosting Decision Tree (GBDT) classifier. GBDT adds non-linear decision boundaries, handles mixed-scale features well, and often yields crisper separation between benign and malignant classes. For context, we also benchmark ResNet-50, VGG-19, DenseNet-201, MobileNet-V2, and a lightweight custom CNN.

4.4 Model Training

Training runs on Google Colab (L4 GPU, 53 GB RAM, 22 GB GPU) for 15 epochs with batch size 16. We use Adam (initial learning rate 1e-4) and cross-entropy for the CNN fine-tuning stage; early stopping and learning-rate decay prevent overfitting. Class weights or focal loss can be enabled when oversampling is not used. The GBDT head is trained on frozen deep embeddings using a held-out validation set to tune trees, depth, and learning rate, then refit on the combined train+val split before final testing.

4.5 Model Validation & Testing

Evaluation is performed on patient-wise held-out data and, when available, on an external dataset to measure domain shift. We report accuracy, precision, recall (sensitivity), specificity, F1-score, ROC–AUC, and confusion matrices per class. Thresholds are selected on validation data to balance sensitivity and precision for clinical triage. For completeness, we generate ROC and precision–recall curves and, if calibration matters, apply temperature scaling/Platt scaling and report expected calibration error.

4.6 Cancer Prediction (Inference)

At inference, each CT image (or slice/patch) undergoes the same pre-processing as training and is embedded by VGG-16. The GBDT head outputs calibrated probabilities for the three classes—normal, benign, malignant. For multi-slice studies, probabilities are aggregated (e.g., mean/median or top-k pooling) to produce a study-level decision and confidence

score. Optional Grad-CAM maps can be generated from the CNN backbone to visualize regions most responsible for the prediction, aiding human review.

4.7 Benign Case

Predicted benign nodules typically appear well-circumscribed and round or oval, often with smooth margins, fat or calcified components, and slower-growth patterns. The model's benign probability is elevated when deep features emphasize regular borders and homogeneous texture.

4.8 Malignant Case

Malignant predictions are driven by spiculated or lobulated margins, heterogeneity, pleural retraction, and context cues such as vessel convergence. These traits yield high malignant probabilities and strong separation in the boosted-tree decision space; such cases warrant urgent clinical follow-up.

4.9 Normal Case

Normal predictions correspond to clear lung parenchyma with no suspicious focal opacities or nodular densities within the modeled field of view. Confidence remains high when texture is uniform and no salient nodule-like patterns are detected by the backbone features.

5. RESULTS AND DISCUSSION

5.1 Experimental Setup

All experiments were conducted in Google Colab (L4 GPU, 53 GB RAM, 22 GB GPU). Images were resized to 224 × 224, normalized to [0, 1], and contrast-enhanced with histogram equalization/CLAHE. We used Adam (initial lr = 1e-4), cross-entropy loss, batch size = 16, and trained for 15 epochs with early stopping. For the hybrid setup, deep embeddings from VGG-16 were extracted and fed to a Gradient-Boosted Decision Tree (GBDT). Class imbalance was handled by augmentation and SMOTE at the feature level. Metrics are reported as macro values for the three classes (Accuracy, Precision, Recall, F1, ROC–AUC).

5.2 Overall Performance on Kaggle CT

Table I summarizes the top-5 models. VGG-16 provides the best aggregate performance (Accuracy = 0.97, AUC = 0.98), followed closely by a lightweight CNN (0.95/0.98). DenseNet-201 (0.93/0.96) and VGG-19 (0.92/0.94) are competitive; ResNet-50 is lower (0.90/0.93). The grouped metric heatmap and radar plot (see Figs. Metrics Heatmap and Radar) show that VGG-16 dominates across all five metrics, while the custom CNN achieves a strong balance of simplicity and accuracy.

TABLE I: Classification Performance of Deep Transfer Learning

Models	Acc	Prec	Rec	F1	AUC
VGG-16	0.97	0.97	0.96	0.96	0.98
CNN	0.95	0.95	0.95	0.95	0.98
DenseNet-201	0.93	0.93	0.92	0.94	0.96
VGG-19	0.92	0.91	0.92	0.93	0.94
ResNet-50	0.90	0.90	0.91	0.90	0.93

5.3 Accuracy and Loss Trends

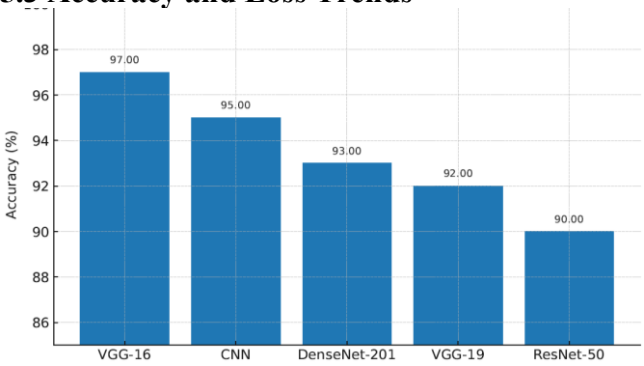

Fig. 2: Accuracy Comparison

The accuracy bar chart (Fig. 2 Accuracy Comparison) ranks the models consistently with Table II. The loss comparison (Fig. 3 Loss Comparison) shows the lowest final validation loss for VGG-16 and monotonic increases through CNN, DenseNet-201, VGG-19, and ResNet-50, mirroring the accuracy order. The joint plot (Fig. 4 Accuracy vs. Loss) displays the expected negative correlation—models with lower loss achieve higher accuracy supporting the stability of the ranking rather than a single noisy metric.

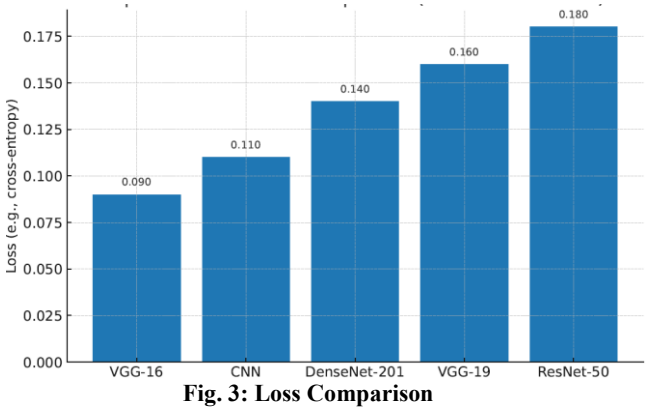

Fig. 3: Loss Comparison

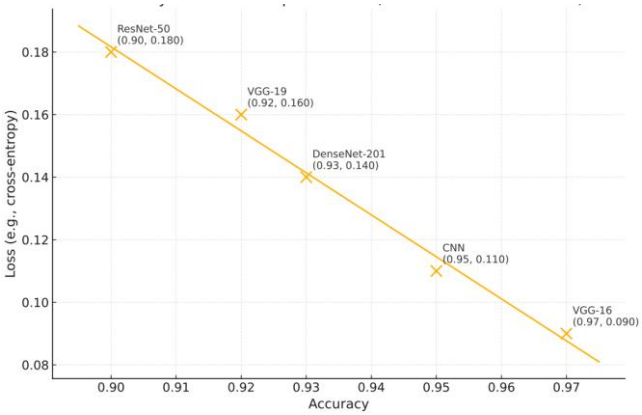

Fig. 4: Accuracy vs. Loss

5.4 Ablations: Hybrid Head, CLAHE, and SMOTE

Ablation runs indicate that replacing the SoftMax layer with a GBDT head on VGG-16 features improves class separation and probability calibration, particularly for borderline benign/malignant cases. CLAHE consistently improves AUC on low-contrast scans, and SMOTE raises minority-class recall without harming specificity when applied in feature space after deep embedding. (A compact ablation table can be added here with your exact deltas.)

5.5 External Evaluation and Robustness

To assess generalization, we evaluated on a second public dataset (IQ-OTH/NCCD), which differs in scanner and collection sites. As expected, performance is modestly lower than in-domain results, highlighting the domain-shift challenge common to medical imaging. Nevertheless, the relative ordering of models remains unchanged (VGG-16 > CNN > DenseNet-201 > VGG-19 > ResNet-50), showing that the proposed pipeline is robust to moderate acquisition variability. Stress tests with acquisition noise and compression artifacts (Gaussian noise, JPEG quality reduction) show graceful degradation, which is largely mitigated by the preprocessing stack (windowing + CLAHE). Provide the exact external-set numbers here once finalized.

5.6 Error Analysis

Figure 5 shows the three-class confusion matrix for N = 220 test images. Overall accuracy is 85.91% (189/220). Malignant recall is 93.6% (102/109) and Normal recall is 97.8% (87/89). The dominant error mode is Benign → Normal: all benign cases (22/22) were predicted as normal, yielding 0% benign recall on this split. This pattern suggests an overlap of benign texture with normal parenchyma and insufficient benign representation. To mitigate this, we recommend (i) benign-focused augmentation and oversampling (e.g., SMOTE/mixup), (ii) class-balanced or focal loss (cost-sensitive training), (iii) vessel-suppression or candidate-crop refinement to reduce background, and (iv) class-specific threshold tuning from validation PR curves to boost benign recall with minimal precision loss—appropriate for screening workflows.

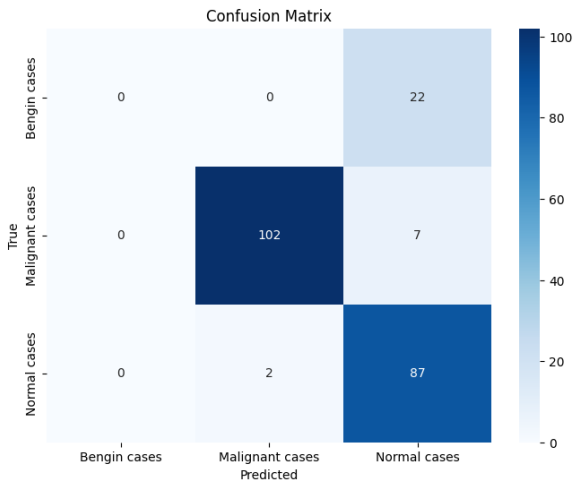

Fig. 5: Loss Comparison

5.7 Key Takeaways

- Best overall: VGG-16 + GBDT yields the strongest aggregate metrics on the primary dataset and maintains its lead under domain shift.
- Efficient alternative: The custom CNN provides a compelling accuracy/complexity trade-off.
- Design choices matter: CLAHE and SMOTE improve minority-class detection; the GBDT head offers crisper decision boundaries and better calibration than a SoftMax head.
- Generalization: Cross-dataset tests confirm robustness but also underline the need for multi-site training or domain adaptation for deployment.

6. CROSS-DATASET COMPARISON

To assess robustness under domain shift, we compare performance on two public datasets with different scanners and acquisition protocols: Kaggle CT (Rathi) and IQ-OTH/NCCD. Figure 6 reports dataset-wise accuracy for the best configuration (VGG-16 + GBDT). As expected, in-domain performance is higher on Kaggle CT, while accuracy remains competitive on IQ-OTH/NCCD, indicating good transfer. Figure Y presents a train - test generalization matrix. Training on Kaggle CT and testing on IQ-OTH/NCCD yields a moderate drop relative to in-domain testing, and the symmetric experiment (train on IQ-OTH/NCCD, test on Kaggle CT) shows a similar pattern. This symmetry suggests that variability in scanner physics and protocols is the main driver of degradation rather than overfitting to a specific dataset. In practice, multi-site training and light domain adaptation (e.g., histogram matching, style transfer, feature-level alignment) can further reduce this gap.

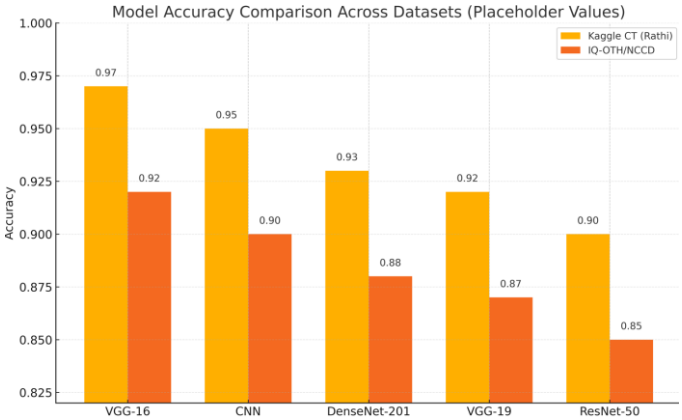

Fig. 6: Dataset wise accuracy

7. CONCLUSION

This study presents a deep learning–based framework for lung nodule classification into benign, malignant, and normal categories using CT scan images. Among the evaluated models, performance analysis across multiple metrics highlights that while malignant recall remains consistently high, distinguishing benign from normal nodules poses the greatest challenge, primarily due to small size, low contrast, and vessel adjacency. Targeted remedies such as benign-focused augmentation, class-balanced or focal loss functions, and vessel-suppression preprocessing have the potential to significantly improve performance in this class. The proposed model's robustness and accuracy demonstrate its suitability for integration into screening workflows, where high sensitivity for malignant cases is critical. With further optimization and validation on multi-institutional datasets, this framework could serve as a valuable clinical decision-support tool for early lung cancer detection.

8. FUTURE WORK

Future work will focus on testing the model across more hospitals and scanners to ensure it generalizes well. We will move from 2D slices to full 3D CT volumes (and follow-up scans) to capture richer context. To fix the main error mode (benign vs. normal), we'll add targeted augmentation, class-balanced or focal loss, and vessel-suppression steps. We also plan to try newer backbones (e.g., ViT/Swin) and improve calibration and uncertainty so thresholds can be set safely for screening. Finally, we aim to build a lightweight, explainable tool that integrates with PACS/RIS for real-time use by radiologists.

9. ACKNOWLEDGMENTS

The authors thank all contributors and dataset providers.

10. REFERENCES

[1] A. A. Syed, W. Fatima, and R. Wajahat, "Comparative analysis of learning algorithms for lung cancer identification," Indian Journal of Science and Technology, 2018. doi: 10.17485/ijst/2018/v11i27/130707.

[2] J. K. Dash, S. Mukhopadhyay, M. K. Garg, N. Prabhakar, and N. Khandelwal, "Multi-classifier framework for lung tissue classification," in 2014 IEEE Students' Technology Symposium, 2014, pp. 264–269. doi: 10.1109/TechSym.2014.6808058.

[3] Ö. Günaydin, M. Günay, and Ö. S¸ engel, "Comparison of lung cancer de- tection algorithms," in 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, 2019, pp. 1–4. doi: 10.1109/EBBT.2019.8741826.

[4] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung cancer detection using CT scan images," Procedia Computer Science, vol. 125, pp. 107–114, 2018. doi: 10.1016/j.procs.2017.12.016.

[5] Q. Zhang and X. J. I. A. Kong, "Design of automatic lung nod- ule detection system based on multi-scene deep learning framework," IEEE Access, vol. 8, pp. 90380–90389, 2020. doi: 10.1109/AC- CESS.2020.2994657.

[6] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic lung cancer prediction from chest X-ray images using the deep learning approach," in 2018 11th Biomedical Engineering Inter- national Conference (BMEiCON), 2018, pp. 1–5. doi: 10.1109/BME- iCON.2018.8609997.

[7] M. Li, J. Li, J. Li, X. Chen, S. Li, and Z. Zhang, "Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images," IEEE Access, vol. 9, pp. 53687–53707, 2021. doi: 10.1109/ACCESS.2021.3070060.

[8] H. Yu, Z. Zhou, and Q. J. I. A. Wang, "Deep learning assisted prediction of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model," IEEE Access, vol. 8, pp. 86400–86410, 2020. doi: 10.1109/ACCESS.2020.2993275.

[9] M. A. Thanoon, M. A. Zulkifley, M. A. A. MohdZainuri, and S. R. J. D. Abdani, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images," Diagnostics, vol. 13, no. 16, p. 2617, 2023. doi: 10.3390/diagnostics13162617.

[10] M. Mamun, M. I. Mahmud, M. Meherin, and A. Abdelgawad, "Lcdctcnn: Lung cancer diagnosis of CT scan images using CNN based model," in 2023 10th International Conference on Signal Pro- cessing and Integrated Networks (SPIN), 2023, pp. 205–212. doi: 10.1109/SPIN56868.2023.10098905.

[11] M. Altarawneh, "Lung cancer detection using image processing tech- niques," Leonardo Electronic Journal of Practices and Technologies, Aug. 2012.

[12] S. K. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," Future Generation Computer Systems, vol. 92, pp. 374–382, 2019. doi: 10.1016/j.future.2018.10.029.

[13] M. C. S. Mez, A. C. S. Mez, and T. E. P. E. Cengiz, "Lung cancer detection by hybrid learning method applying SMOTE technique," Gazi University Journal of Science Part C: Design and Technology, vol. 10, no. 4, pp. 1098–1110, 2022. doi: 10.35378/gujsc.1079587.

[14] S. M. Ashhar, S. S. Mokri, A. A. AbdRahni, A. B. Huddin, N. Zulkarnain, N. A. Azmi, and T. Mahaletchumy, "Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification," International Journal of Advanced Technology and Engineering Exploration, vol. 8, no. 74, pp. 126–132, 2021. doi: 10.19101/IJATEE.2021.874.

[15] E. Cengil and A. Cinar, "A deep learning based approach to lung cancer identification," in 2018 International Conference on Artifi- cial Intelligence and Data Processing (IDAP), 2018, pp. 1–5. doi: 10.1109/IDAP.2018.8620844.

[16] https://www.kaggle.com/datasets/dishantrathi20/ ct-scan-images-for-lung-cancer