EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

Performance Assessment of Deep Learning Models on Non-Small Cell Lung Cancer Type Classification

K. Ezhilraja[1,*], P. Shanmugavadivu[2,**]

[1]Research Scholar, Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Gandhigram - 624302, Tamil Nadu, India
[2]Professor, Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Gandhigram - 624302, Tamil Nadu, India (Email: psvadivu67@gmail.com)

Abstract

In recent years, lung cancer incidents are very high with equally high mortality rate. The main reason for fatal incidences is the late diagnosis and confirmation of the disease at an advanced stage. Identification of the disease at an early stage using lung Computed Tomography (CT) offers tremendous scope for timely medical intervention. The article illustrates the use of deep transfer learning-based pre-trained models for the diagnosis of Non-Small Cell Lung Cancer (NSCLC). The datasets were chosen from Chest CT Scan Images and the Lung Image Database Consortium (LIDC), containing over 3,179 images depicting three NSCLC types, namely, normal, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. The process is designed to measure the accuracy of NSCLC detection with an experimental dataset using approaches with and without pre-processing of lung images. The transfer learning models use deep learning and produce good results in prediction and classification. The image dataset was first handled by the convolutional neural networks DenseNet121, ResNet50, InceptionV3, VGG16, Xception, and VGG19. As a second phase, input images were subjected to contrast/brightness enhancement using Multi Level Dualistic Sub Image Histogram Equalization (ML-DSIHE). Enhanced images were further processed using shape-based feature extraction. Finally, those features input to CNN models and the results recorded. Among these models, VGG16 achieved the highest accuracy of 81.42% using the original dataset and 91.64% with the enhanced dataset. The performance of these two approaches was also evaluated using Precision, Recall, F1-Score, Accuracy, and Loss. It is confirmed that VGG16 gives highly reliable accuracy when trained upon enhanced images.

Keywords: Lung Cancer Classification, Lung Image Enhancement, Contrast Enhancement, Lung Feature Extraction, Lung Cancer Detection.

Received on 11 02 2024, accepted on 28 05 2024, published on 24 06 2024

doi: 10.4108/eetpht.10.6423

1. Introduction

Early detection of lung cancer is crucial in increasing the survival rate of patients. Medical imaging technologies such as CT scans and X-rays are commonly used to detect lung cancer in the early stages of lung cancer. The U.S. National Institute of Health, National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Cancer Statistics Review, 1975–2015 provides information on cancer incidence, mortality, survival, and trends in the United States. The report is based on data collected under the SEER program, which covers approximately 28% of the US population. The report includes statistics on various types of cancer, including lung cancer, and provides information on age-specific and age-adjusted incidence rates, mortality rates, and survival rates. The SEER program serves as a reliable resource for cancer

*Corresponding author. Email: kezhilraja3@gmail.com
** Corresponding author. Email: psvadivu67@gmail.com

researchers, policy makers, and health care providers, as it provides valuable statistics and data on cancer collected from the United States [23].
The application of artificial intelligence (AI) in medical imaging, highlighting the potential benefits and challenges associated with the use of AI in clinical practices has been discussed in detail [1]. The various AI techniques that have been developed for cancer detection using deep learning and computer-aided diagnosis are meticulously reported. The article exemplifies the role of AI in improving the accuracy of cancer detection and diagnosis, leading to treatment planning. The relevance of AI to the field of medical imaging and the potential of AI cancer care are given in detail. In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown promising results in the early detection of lung cancer. CNNs are powerful machine learning models which learn and extract the features from medical images automatically, making them a reliable approach for lung cancer classification [2]. Some of the popular CNN models used in the early detection of lung cancer include DenseNet121, ResNet50, InceptionV3, VGG16, Xception, and VGG19. These models work by learning from large datasets of medical images to identify patterns that distinguish cancerous tissues from non-cancerous ones. These have demonstrated the ability to attain high accuracy in classification of lung cancer images and promises lung cancer detection at an early stage, with improved accuracy.

2. Review of literature

This section provides an overview of the existing literature on lung cancer classification methods, highlighting their computational merits and limitations.
Sharma et al [24] presented a two-step algorithm for early detection of lung cancer using CT images. It performed patch extraction and nodule segmentation using Otsu and morphological operations. A deep CNN was used for nodule classification, with test accuracy of 84.13%; sensitivity and specificity as 91.69% and 73.16% respectively.
Lakshmanaprabu et al [5] proposed an automated diagnosis/ classification method for CT lung images, utilizing Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA) to analyze and classify lung nodules into malignant or benign. The use of Modified Gravitational Search Algorithm (MGSA) to optimize the ODNN led to high sensitivity, specificity, and accuracy in lung cancer classification recorded as 96.2%, 94.2% and 94.56% respectively.
Nasrullah et al [4] developed a deep learning-based model for precise diagnosis of malignant nodules in lungs. This model utilized two customized mixed link network (CMixNet) architectures for nodule detection and classification. It employed faster R-CNN and U-Net encoder-decoder architecture for nodule detection and Gradient Boosting Machine (GBM) for classification. It achieved minimum false positive values and by combining decisions on physiological symptoms and clinical biomarkers. It achieved 94% sensitivity and 91% specificity on LIDC-IDRI datasets, outperforming the existing methods.
Mhaske et al [25] developed a CAD system for efficient diagnosis of lung cancer using deep learning techniques. This system integrates CNN for precise feature extraction and Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM) for accurate classification. Additionally, CT scan images underwent segmentation based on Otsu thresholding method, rendering a reliable and practical tool in the realm of medical imaging for contemporary medical diagnosis.
Sajja et al [5] proposed a deep neural network based on GoogleNet for automated diagnosis of cancer from CT scan images. The implementation of a densely connected architecture proved to be beneficial in mitigating the computational costs and addressing overfitting concerns. The evaluation was conducted using the (LIDC) dataset, and a comparative analysis was performed against pre-trained CNN models, specifically AlexNet, GoogleNet, and ResNet50. This approach demonstrated a superior classification accuracy of 91%, outperforming those models under consideration.
Sang et al [6] proposed a framework incorporating two key components: a deep 3D Faster R-CNN for detection and a U-Net encoder-decoder with MixNet for feature learning within lung nodules. The method showcased notable improvements when compared to existing approaches, achieving a sensitivity of 94%, specificity of 90%, and an impressive Area Under the Receiver Operating Curve (AUC) score of 0.99. This evaluation was conducted on a dataset consisting of 1200 images sourced from LIDC-IDRI.
Bhandary et al [7] introduced a Deep Learning (DL) framework for pneumonia and lung cancer classification using modified AlexNet (MAN). This model improved classification accuracy by combining handcrafted and learned features. The results were compared with other pre-trained DL techniques. The classification accuracy of 97.27% was achieved on benchmark lung cancer CT images of LIDC-IDRI.

3. Proposed Work

The Chest CT Scan Image Dataset was chosen from the Kaggle Repository and also the LIDC dataset. The Kaggle dataset consists of 613 images of lung cancer patients, created in 2019. The LIDC dataset comprises 1,018 cases, with the number of nodules identified differing for each case, totalling 2,566 nodules in the dataset, according to the LIDC-IDRI dataset website. The dataset includes three NSCLC types of chest cancer viz., adenocarcinoma, large cell carcinoma, squamous cell carcinoma, alongside normal cells. The dataset was split into a training set of 60%, a testing set of 20%, and a validation set of 20%. The deep learning model's classification potential was evaluated using three performance metrics, viz. precision,

recall, and F1-score. Precision quantifies the ratio of true positives among all positive predictions generated by the model, whereas recall gauges the ratio of true positives among all positive instances present in the dataset. The F1-score, a valuable composite metric, represents the harmonic mean of precision and recall, offering a unified measure that strikes a balance between both evaluation criteria.
In this research works six pretrained models namely DenseNet121, ResNet50, InceptionV3, VGG16, Xception, and VGG19 were handpicked based on their proven performance in prediction/classification [8-23, 26-28]. These models were trained on the LIDC directly as well as on the enhanced input images using ML-DSIHE. Eventually, the performance metrics precision, recall, and F1-score were employed to assess the performance of these models with reference to two approaches. The computational steps of the experimental study are given in the next section. The corresponding workflow diagram is given is fig. 2.

3.1. Methodology of NSCLC classification using Deep Learning Models

Input: *CT scan chest image dataset*
Output: *Classification accuracy for six pre-trained model (on Original and Enhanced image)*

---

begin

Step 1: *Read CT Scan Chest Images*
Step 2: *Pre-process the images using the ML-DSIHE method*
Step 3: *Split the dataset into training, testing and validation set*
Step 4: *Train the CNN models - Xception, ResNet50, DenseNet121, VGG19, VGG16, and InceptionV3 models*
A. Original Images; and
B. Enhanced Images
Step 5: *Evaluate the performance of each model on the testing dataset on NSCLC classification*
Step 6: *Identification of the best performance CNN Model*

end

The same procedure is visually depicted in Fig 1.

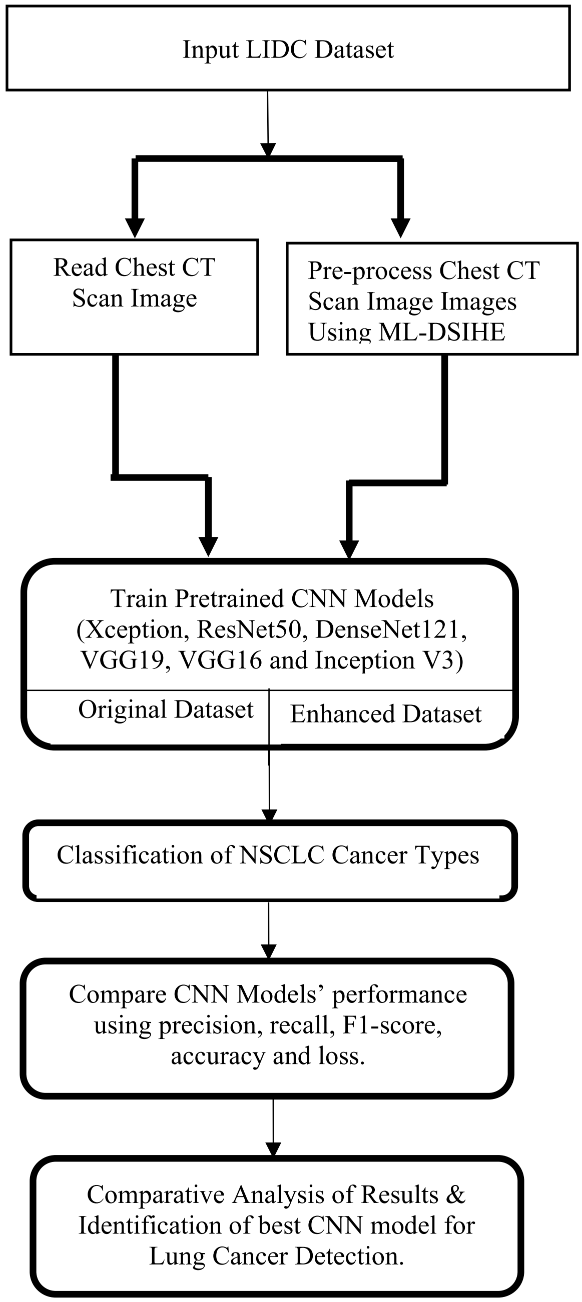


Fig 1: Workflow of the Proposed Research

5. Result and Discussion

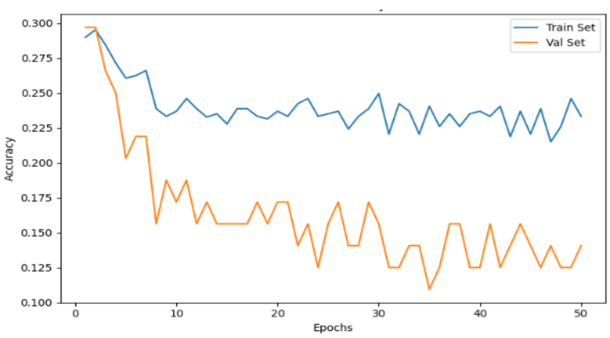
The dataset was split into a training set of 60%, a testing set of 20%, and a validation set of 20%. Table 1 depicts the Testing Accuracy (TA) scores of all those pre-trained models different Epoch (E), Learning Rate (LA), Original Dataset (OD) and Enhanced Dataset (ED). The experimental trials were conducted for two different learning rates, say 0.1 and 0.001 for four different values of epochs 5, 10, 20, and 50.

Table 1: Testing Accuracy of Pretrained CNN models for Epochs and Learning Rates

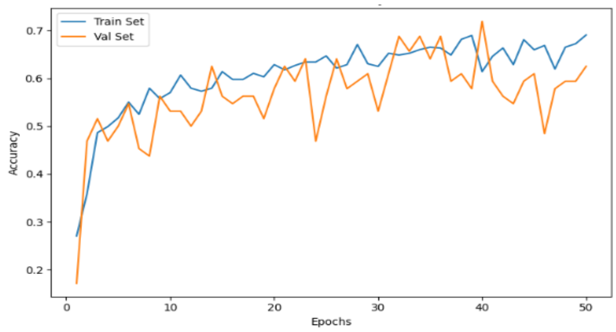
No. of Epoch	Learning Rate	Xception		DenseNet121		VGG16		ResNet50		VGG19	
		Testing Accuracy									
		OD	ED	OD	ED	OD	ED	OD	ED	OD	ED
5	0.1	23	26	42	44	48	51	48	52	51	54
5	0.001	27	29	45	48	51	55	51	56	52	57
10	0.01	25	28	43	46	52	57	52	58	56	60
10	0.001	29	31	47	50	56	65	56	61	53	61
20	0.1	26	29	46	49	53	75	53	60	57	63
20	0.001	29	32	49	52	57	78	57	65	60	65
50	0.1	28	32	48	51	60	82	60	66	62	67
50	0.001	29	34	53	56	77	89	65	71	65	72

Table 1 shows the testing accuracy of the six deep learning models for five different values of epochs and two learning rates. It was observed that the testing accuracy of all models trained on enhanced data was invariably higher when compared to the training approach on original data. Additionally, the relationship between epochs and learning rate against the accuracy of the model is illustrated in the table 1. Increasing the number of epochs and decreasing the learning rate tends to improve the testing accuracy of the models. Among the models, DenseNet121 and ResNet50 perform relatively well on the original data, whereas VGG16 and VGG19 show higher accuracy on the enhanced data. Xception shows a moderate performance on both original and enhanced data. The overall analysis of the results suggest that the performance of the deep learning models can be improved by training them on contrast /brightness enhanced data as well as by optimizing the hyperparameters values.

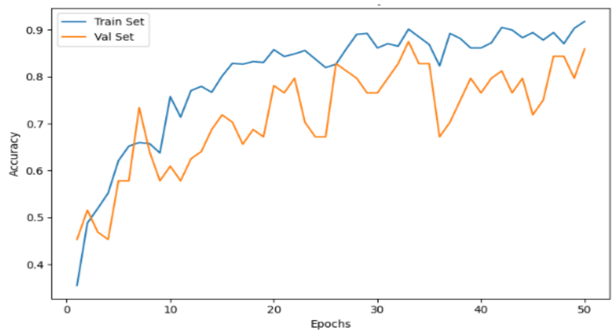
Fig. 2 and 3 show the overall accuracy and loss produced by each of the chosen models, during training and validation. If is undoubtedly clear that VGG16 stands out in terms of highest accuracy and lowest loss obtained during experimentation.



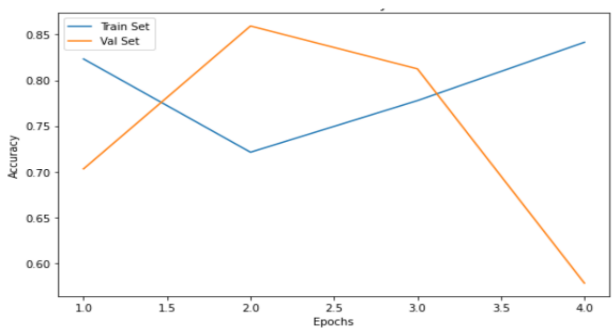
(a)



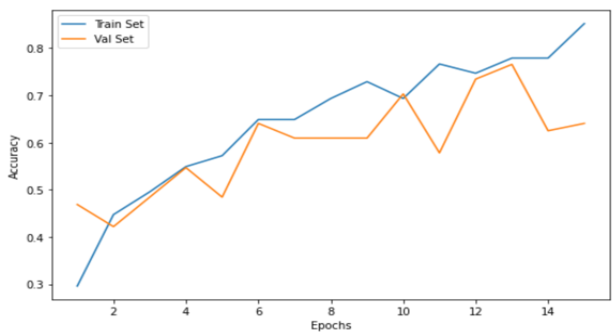
(b)



(c)

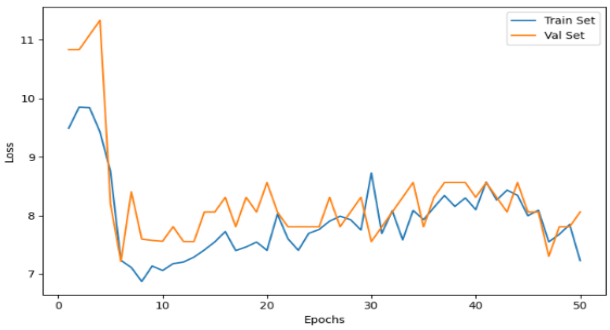


(d)

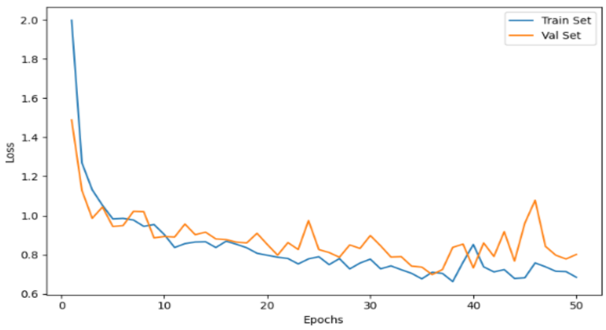


(e)

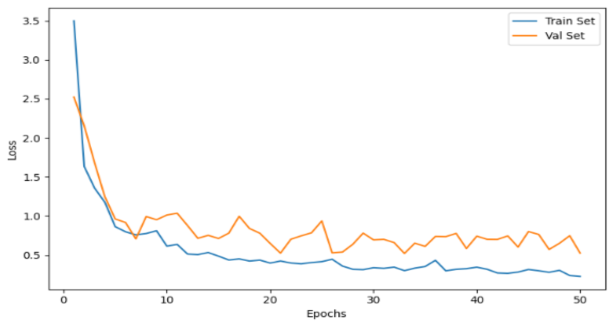
Fig 2: Training and validation accuracy of the CNN models (a) Xception; (b) Densenet121; (c) VGG16; (d) ResNet50; and (e) VGG19



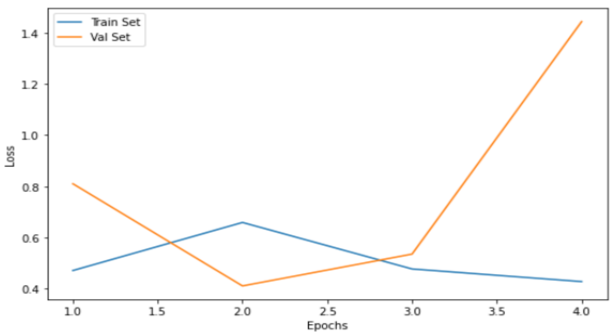
(a1)



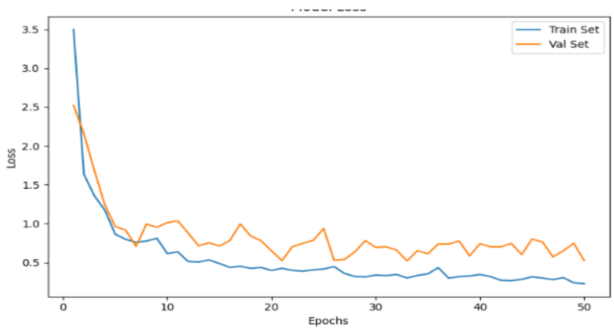
(b2)



(c3)



(d4)



(e5)

Fig 3: Training and Validation Loss

Fig 2 and Fig 3 depict the accuracy and loss values for different deep learning models.

Table 2: Performance measure of NSCLC type detection using VGG16

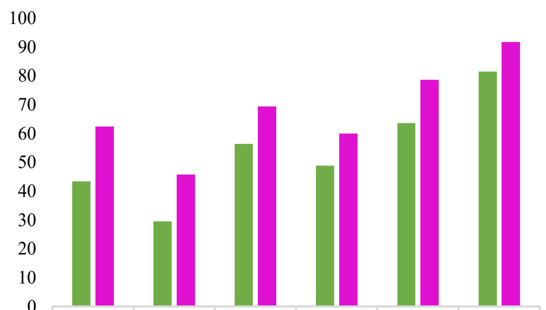
NSCLC TYPE	Precision		Recall		F1-score	
	OD	ED	OD	ED	OD	ED
Normal	0.93	0.95	0.91	0.93	0.89	0.95
Adenocarcinoma	0.94	0.96	0.81	0.85	0.85	0.92
Squamous-Cell-Carcinoma	0.93	0.95	0.92	0.95	0.89	0.94
Large-cell-Carcinoma	0.71	0.80	0.94	0.96	0.83	0.90

Table 2 shows the accuracy recorded by VGG16 with respect to lung cancer type identification in precision, recall and F1-score. The performance of VGG16 is promising on both original and enhanced datasets. VGG16 proved to be more efficient on the enhanced dataset for identifying normal cells, adenocarcinoma, squamous cell carcinoma and large cell carcinoma. The highest precision of 0.96 was obtained for adenocarcinoma identification on the enhanced data. Likewise, the maximum recall was 0.96 for large cell carcinoma on the enhanced data. The highest F1-Score for normal tissue detection was 0.96.

Table 3: Accuracy of Deep Learning Models using Original Dataset and Enhanced Dataset

Models	Accuracy %		% Increase
	OD	ED	
DenseNet121	43.40	62.40	19.00
ResNet50	29.53	45.70	16.17
InceptionV3	56.40	69.40	13.00
Xception	48.88	59.88	11.00
VGG19	63.60	78.60	15.00
VGG16	81.42	91.64	10.22

Table 3 gives the results on the accuracy of CNN models for lung cancer classification on original and enhanced datasets, during testing. The table also shows the percentage increase in accuracy for each model when using the enhanced dataset instead of the original dataset. The results indicate that all the models perform better with the enhanced dataset, with the increase in accuracy ranging from 9.1% for CNN to 19% for DenseNet121. VGG16 had the highest accuracy among all models for both datasets, but with only a 10.22% increase in accuracy. It can be seen from Table 2 that VGG16 produces the maximum accuracy of 81.42% on original data and 91.64% on the enhanced data, indicating the former approach is more effective. It is interesting to observe that all the chosen models have depicted an increase in accuracy of classification, ranging from 10.22 to 19.22 on the enhanced dataset. The results suggest that enhancing the dataset prior to training can significantly improve the accuracy of deep learning models for lung cancer classification. The same results are visualized in Fig. 4 for easy interpretation.



	DenseNet121	ResNet50	InceptionV3	Xception	VGG19	VGG16
OD	43.4	29.53	56.4	48.88	63.6	81.42
ED	62.4	45.7	69.4	59.88	78.6	91.64

Fig 4: Accuracy of Original and Enhanced Dataset

This research study has given a clear illustration that VGG16 offers the best results in NSCLC detection, with hyperparameter values of 0.001 for learning rate and 50 for epoch. It is also demonstrated that pretrained models can be ideally trained for lung cancer detection, with a precursor that the model performs better with an enhanced image.

6. Conclusion

This paper has successfully shown that deep transfer learning-based pre-trained models can be used for diagnosing NSCLC using Chest CT scan images and LIDC dataset images. The models were trained on the original dataset and on images enhanced using the ML-DSIHE method of the original CT scan image dataset. The original images were fed into five different CNNs, namely DenseNet121, ResNet50, InceptionV3, VGG16, Xception, and VGG19 for NSCLC type classification. The performance of these models significantly varied depending on the nature of the input image and values assigned for epoch and learning rate. Among the models tested, VGG16 consistently performed, achieving the highest accuracy of 81.42% original image and 91.64% on the enhanced image. These findings offer scope for fine-tuning the VGG16 model to obtain the best fit for the classification of NSCLC types.

References

[1] A. Khalid et al., "Deep Learning Applications in Cancer Diagnosis and Prognosis," Computational and Structural Biotechnology Journal, vol. 18, pp. 8-17, 2020. doi: 10.1016/j.csbj.2019.12.010.

[2] Wenya and Linda Bi, "Artificial Intelligence in Cancer Imaging: Clinical

[3] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, 374-382.

[4] Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors, 19(17), 3722.

[5] Sajja, T., Devarapalli, R., &Kalluri, H. (2019). Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning. Traitement du Signal, 36(4), 339-344.

[6] Sang, J., Alam, M. S., & Xiang, H. (2019, May). Automated detection and classification for early-stage lung cancer on CT images using deep learning. In Pattern Recognition and Tracking XXX (Vol. 10995, pp. 200-207). SPIE.

[7] Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., ... & Raja, N. S. M. (2020). Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters, 129, 271-278.

[8] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G. S., Naidich, D. P., Shetty, S., & Lungren, M. P. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 25(6), 954-961.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[10] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. "Densely Connected Convolutional Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

[11] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556 (2014).

[12] Challenges and Applications", A Cancer Journal for Clinicians, 2019

[13] Zhang, L., Wang, S., & Liu, Q. (2019). A deep learning-based model for lung nodule classification. Computerized Medical Imaging and Graphics, 77, 101639

[14] Wang, S., Shi, W., Chen, X., Jiang, R., & Song, Z. (2018). Lung nodule classification using deep residual learning based on CT images. Computer Methods and Programs in Biomedicine, 157, 19-30

[15] Song, Y., Chen, Y., Liu, Z., & Jiang, H. (2020). Deep learning-based classification of lung nodules using a convolutional neural network with 3D reconstruction. Frontiers in Oncology, 10, 57

[16] Li, X., Zhang, S., Zhang, Q., Wei, X., & Pan, Y. (2018). Lung nodule classification using deep convolutional neural networks on computed tomography images with additional color channels. Computers in biology and medicine, 102, 33-39

[17] Yao, J., Wu, Y., Wang, Z., Zhang, J., & Zhang, Y. (2019). A support vector machine classifier with feature selection for lung nodule detection in CT images. Journal of healthcare engineering, 2019

[18] Qu, J., Wang, S., Wang, Z., Tang, J., Zhang, Y., & Lv, X. (2020). A comparative study on the classification of lung nodules in CT images using deep learning methods. Journal of medical systems, 44(5)

[19] Qu, J., Wang, S., Wang, Z., Tang, J., Zhang, Y., & Lv, X. (2020). A comparative study on the classification of lung nodules in CT images using deep learning methods. Journal of medical systems, 44(5)

[20] Cui, L., Zhang, L., Guo, X., & Wang, Y. (2020). A deep learning-based approach for classification of lung nodules in CT images. BMC medical imaging, 20(1), 1-9

[21] Lee, J. H., Lee, K., Lee, S. M., & Kang, M. J. (2019). A deep learning-based approach for automatic lung nodule detection and classification on CT images. Computers in Biology and Medicine, 109, 217-236. doi: 10.1016/j.compbiomed.2019.04.011

[22] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

[23] U.S. National Institute of Health, National Cancer Institute. SEER Cancer Statistics Review, 1975–2015.

[24] Sharma, M., Bhatt, J. S., & Joshi, M. V. (2018, April). Early detection of lung cancer from CT images: nodule segmentation and classification using deep learning. In Tenth International Conference on Machine Vision (ICMV 2017) (Vol. 10696, pp. 226-233). SPIE.

[25] Mhaske, D., Rajeswari, K., &Tekade, R. (2019, September). Deep learning algorithm for classification and prediction of lung cancer using ct scan images. In 2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[26] Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[28] Islam, M. T., Raskar, M. A., Islam, M. R., & Roy, C. K. (2020). A Deep Learning Approach for Lung Nodule Classification in CT Images. In 2020 IEEE International Conference on Innovations in Information Technology (IIT) (pp. 115-120). IEEE