History of Medicine, 2024, 10(2): 80–88

DOI: 10.17720/2409-5834.v10.2.2024.10

DEEP LEARNING-BASED APPROACH FOR LUNG CANCER CLASSIFICATION FOR IMPROVED DIAGNOSIS

T. Jayarajan[1], B. Tharuni[2], Thandu Gnanadeep[2], Vittal Manikanta[2]

[1]Assist.Professor, [2]UG Scholar, [1,2]Department of CSE (AI&ML)

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Hyderabad, Telangana.

ABSTRACT

Medical imaging tools are essential in early-stage lung cancer diagnostics and the monitoring of lung cancer during treatment. Various medical imaging modalities, such as chest X-ray, magnetic resonance imaging, positron emission tomography, computed tomography, and molecular imaging techniques, have been extensively studied for lung cancer detection. These techniques have some limitations, including not classifying cancer images automatically, which is unsuitable for patients with other pathologies. It is urgently necessary to develop a sensitive and accurate approach to the early diagnosis of lung cancer. Deep learning is one of the fastest-growing topics in medical imaging, with rapidly emerging applications spanning medical image-based and textural data modalities. With the help of deep learning-based medical imaging tools, clinicians can detect and classify lung nodules more accurately and quickly. Therefore, this work implements the advanced modifications in CNN model for the detection of lung cancer from chest scan images. The proposed CNN model is able to classify the benign and malignant i.e., normal, and cancerous with higher accuracy as compared to state-of-the-art machine learning approach called support vector machine (SVM) classifier. In addition, the obtained quality metrics discloses the superiority of proposed deep CNN model for assisting the expertise in an enhanced diagnosis.

Keywords: Chest CT images, Lung cancer, Decision support system, Image preprocessing, Deep learning.

1. INTRODUCTION

Lung cancer is the primary cause of cancer death worldwide, with 2.09 million new cases and 1.76 million people dying from lung cancer in 2018 [1]. Four case-controlled studies from Japan reported in the early 2000s that the combined use of chest radiographs and sputum cytology in screening was effective for reducing lung cancer mortality. In contrast, two randomized controlled trials conducted from 1980 to 1990 concluded that screening with chest radiographs was not effective in reducing mortality in lung cancer [2, 3]. Although the efficacy of chest radiographs in lung cancer screening remains controversial, chest radiographs are more cost-effective, easier to access, and deliver lower radiation dose compared with low dose computed tomography (CT). A further disadvantage of chest CT is excessive false positive (FP) results. It has been reported that 96% of nodules detected by low-dose CT screening are FPs, which commonly leads to unnecessary follow-up and invasive examinations. Chest radiography is inferior to chest CT in terms of sensitivity but superior in terms of specificity. Taking these characteristics into consideration, the development of a computer-aided diagnosis (CAD) model for chest radiograph would have value by improving sensitivity while maintaining low FP results [4].

Many computer-aided detection (CAD) systems have been extensively studied for lung cancer detection and classification [5, 6]. Compared to trained radiologists, CAD systems provide better lung nodules and cancer detection performance in medical images. Generally, the CAD-based lung cancer detection system includes four steps: image processing, extraction of the region of interest (ROI), feature selection, and classification. Among these steps, feature selection and classification play the most critical roles in improving the accuracy and sensitivity of the CAD system, which relies on image processing to capture reliable features. However, benign, and malignant nodule classification is a challenge. Therefore, a rapid, cost-effective, and highly sensitive deep learning-based CAD system for lung cancer prediction is urgently needed.

2. LITERATURE SURVEY

The development of malignant cells in the lungs is known as lung cancer. Overall men and women's mortality rates have increased as a result of growing cancer incidence. Lung cancer is a disease wherein the cells in the lungs quickly multiply. Lung cancer cannot be eradicated, but it can be reduced [7]. The number of people affected with lung cancer is precisely equal to the number of people who smoke continuously. Lung cancer treatment was evaluated using classification approaches such as Naive Bayes, SVM, Decision Tree, and Logistic Regression. Pradhan et al. [8] conduct an empirical evaluation of multiple machine learning methods that can be used to identify lung cancer using IoT devices. A survey of roughly 65 papers employing machine learning techniques to forecast various diseases was conducted in this study. The study focuses on a variety of machine learning methods for detecting a variety of diseases in order to identify a gap in prospective lung cancer detection in medical IoT. Deep residual learning is used by Bhatia et al. [9] to identify lung cancer from CT scans. With the UNet and ResNet algorithms, we propose a series of pre-processing approaches for emphasising cancer-prone lung regions and retrieving characteristics. The extracted features are fed through several classifiers, namely Adaboost and Random Forest, and the individual predictions are ensembled to calculate the likelihood of a CT scan becoming cancerous.

Without learning inadequate human information, Shin et al. [10, 11] use deep learning to investigate the characteristics of cell exosomes and determine the similarities in human plasma extracellular vesicles. The deep learning classifier was tested with exosome SERS data from regular and lung cancer cell lines and was able to categorise them with 95% efficiency. The deep learning algorithm projected that 90.7% of patients' plasma exosomes were more similar to lung cancer cell extracellular vesicles than the mean of healthy controls in 43 patients, encompassing stage I and II cancer patients. In the ability to forecast lung ADC subtypes, researchers looked at four clinical factors: age, sex, tumour size, and smoking status, as well as 40 radiomic markers. The LIFEx software was used to extract radiomic characteristics from PET scans of segmented cancers. The clinical and radio mic variables were ranked, and a subset of meaningful features was chosen based on Gini coefficient scores for histopathological class relationships [12]. In the estimation of survival, a deep learning network with a tumour cell and metastatic staging system was used to examine the dependability of individual therapy suggestions supplied by the deep learning preservation neural network. The C statistics were employed to evaluate the performance of the model. The computational intelligence survival neural network model's longevity forecasts and treatment strategies were made easier with the use of a customer interface [13].

3. PROPOSED METHODOLOGY

The project focuses on enhancing the detection and classification of lung cancer using advanced deep learning techniques, specifically through a modified Convolutional Neural Network (CNN) model. Lung cancer remains one of the most lethal forms of cancer, and early detection is crucial for improving patient outcomes. Traditional imaging modalities like chest X-rays, MRI, PET, and CT scans are widely used but have limitations, such as the inability to automatically classify cancer images and challenges in detecting lung cancer in patients with other pathologies. To address these limitations, the project leverages the rapid advancements in deep learning, which have shown significant potential in medical imaging applications.

The core of the project is the development and implementation of a CNN model tailored for lung cancer detection from chest scan images. The project begins with a comprehensive review of current imaging techniques and the role of deep learning in medical imaging, highlighting the advantages of CNNs over traditional machine learning models like Support Vector Machines (SVMs). Following this, the methodology outlines the process of data collection, preprocessing, model architecture design, and training. Data is sourced from public datasets, preprocessed to ensure uniformity, and split into training and validation sets to facilitate robust model training. The proposed CNN model is meticulously designed with multiple convolutional and pooling layers to effectively capture and learn features from the chest scan images. The model architecture includes techniques like dropout for regularization to prevent overfitting and ensure better generalization. The training process involves optimizing hyperparameters and validating the model's performance through metrics like accuracy, precision, recall, F1-score, and the ROC-AUC curve. The results section provides a detailed comparison between the CNN model and the SVM classifier, demonstrating the superior performance of the CNN in terms of accuracy and diagnostic capabilities.

Finally, the project concludes with a discussion on the implications of the findings for clinical practice, emphasizing the potential of deep learning models to assist clinicians in making more accurate and timely diagnoses of lung cancer. The conclusion also suggests future research directions, including the integration of transfer learning and the exploration of more complex architectures to further improve detection accuracy. Overall, this project aims to contribute to the field of medical imaging by providing a more sensitive and accurate tool for early lung cancer diagnosis, ultimately improving patient care and outcomes.

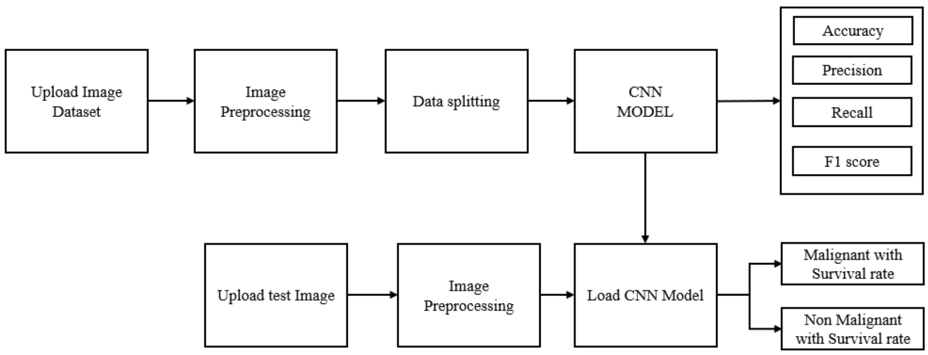

Figure 1: Block diagram of proposed lung cancer classification model.

4. RESULTS AND DISCUSSION

Figure 2 shows that Presents a collection of images sourced from a dataset categorized as Non-Malignant class. Each image visibly represents lung tissue devoid of any cancerous manifestations, serving as representative samples for this class in the dataset. Figure 3 shows that exhibits a series of images sourced from a dataset categorized as Malignant class. Each image conspicuously showcases lung tissue displaying cancerous growth or abnormalities, thereby serving as exemplars for this category within the dataset. Figure 4 shows Depicts the visual layout of a user interface (UI) tailored for the explicit purpose of detecting instances of lung cancer within images. This UI encompasses various interactive elements and functionalities aimed at facilitating the process of image-based lung cancer detection.

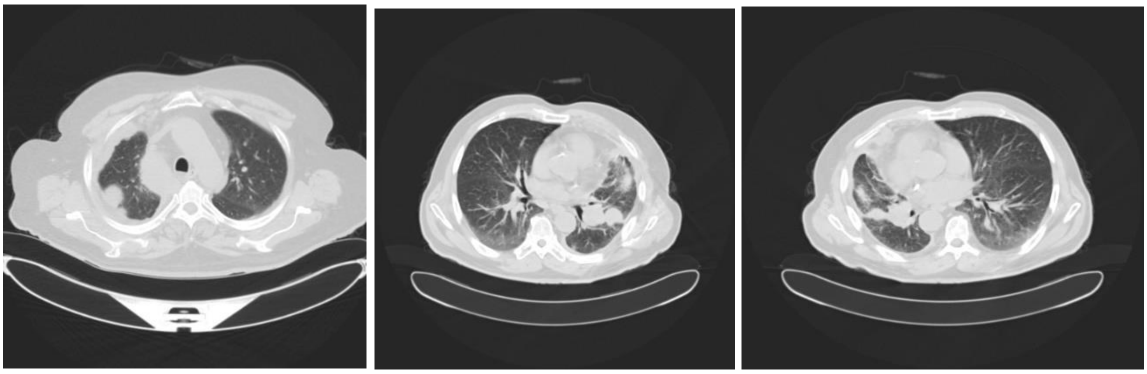

Figure 2: Sample images from dataset with Non-Malignant class.

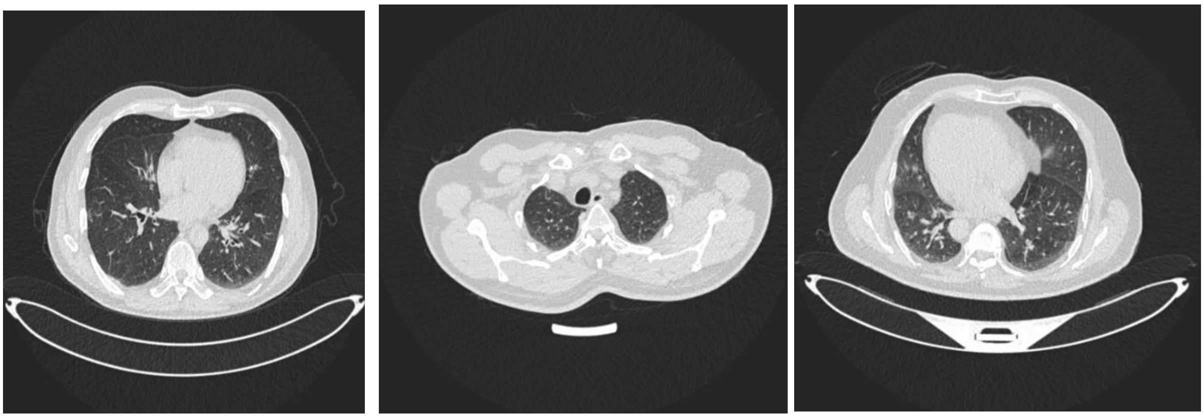

Figure 3: Sample images from dataset with Malignant class.

Figure 4: sample UI used for Lung cancer detection from images.

Figure 5 shows the visual representation of data presented within the UI subsequent to undergoing a series of preprocessing steps. These preprocessing steps typically involve image enhancements, corrections, or feature extraction techniques aimed at optimizing the images for subsequent analysis and detection tasks. Figure 6 shows the graphical depiction of the performance metrics attributed to a SVM utilized for the task of lung cancer detection. These metrics may include accuracy, precision, recall, or F1-score, indicating the efficacy of the SVM model in distinguishing between cancerous and non-cancerous lung tissue. Figure 7 is a visual representation of the confusion matrix corresponding to the performance evaluation of the SVM Classifier. This confusion matrix offers a comprehensive breakdown of the classifier's predictive performance, including true positive, false positive, true negative, and false negative classifications.

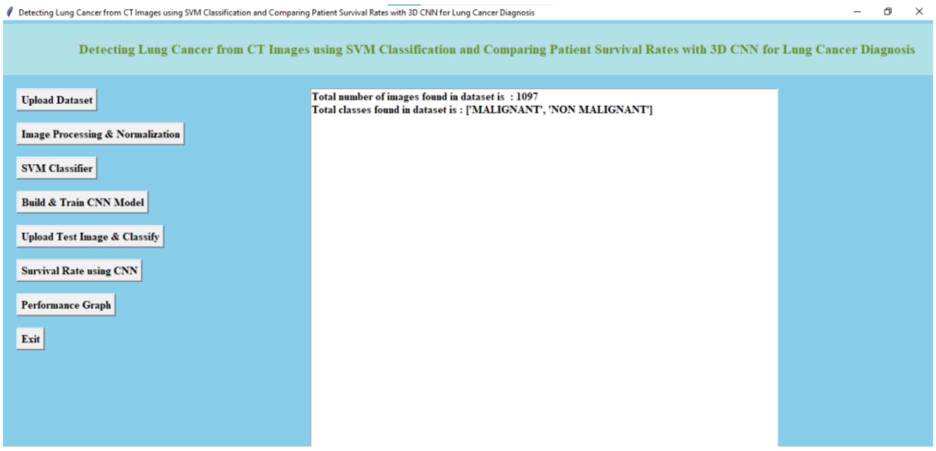

Figure 5: UI shows the data after image preprocessing

Figure 6: Performance of SVM classifier.

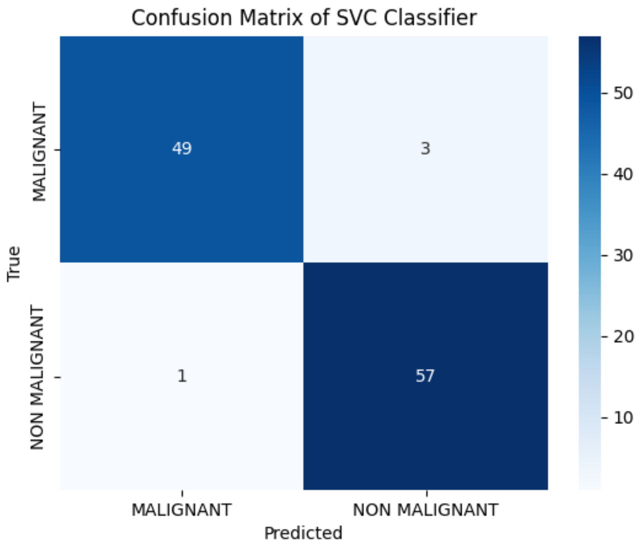

Figure 7: Confusion matrix of SVM Classifier.

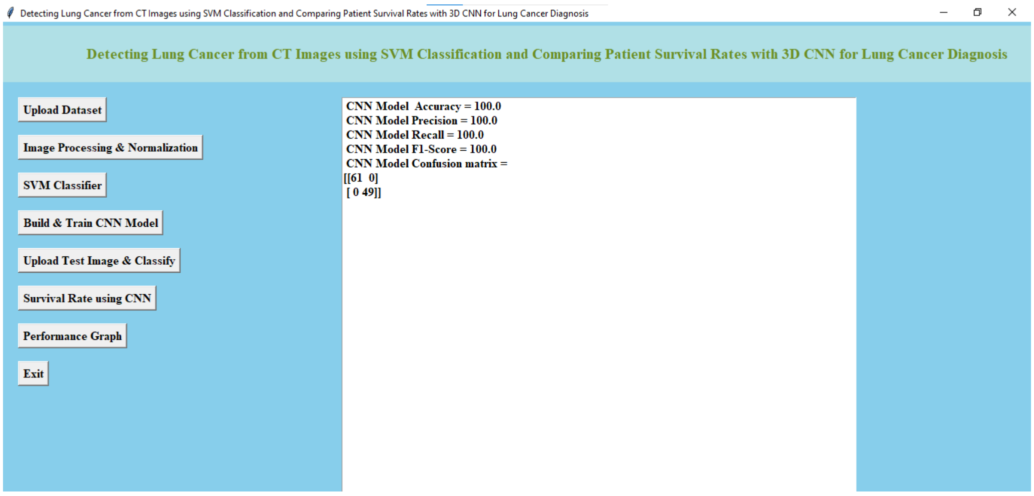

Figure 8: Performance evaluation of CNN Classifier

Figure 8 showcases a graphical representation of the performance evaluation metrics associated with CNN classifier deployed for lung cancer detection. This visual display highlights the CNN model's effectiveness in accurately identifying lung cancer instances within images. Figure 9 shows an iteration-wise graphical representation of the accuracy and loss values observed during the training process of the CNN Classifier. This graph offers insights into the model's learning progress and convergence towards optimal performance over successive training iterations. Figure 10 shows the visual depiction of the confusion matrix corresponding to the proposed CNN Model's performance evaluation. This confusion matrix furnishes detailed insights into the CNN model's classification accuracy and error rates across different classes.

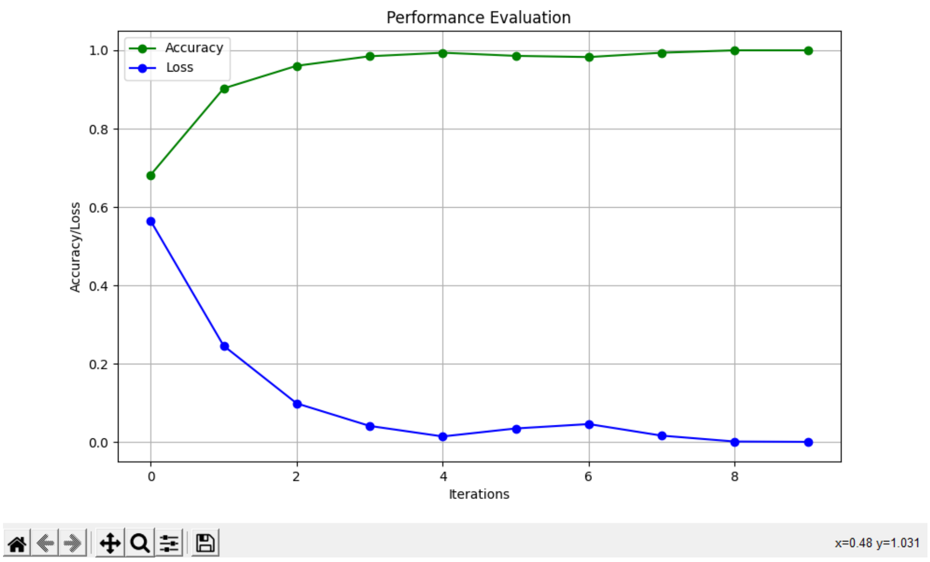

Figure 9: Iteration wise accuracy and loss graph.

Figure 11 shows the visual output generated by the CNN model, encompassing the predicted outcomes of lung cancer detection on a set of input images. This display includes examples of correctly and incorrectly classified lung tissue images. Figure 11 consists of two distinct visual components; one representing the predictive estimation of survival rates based on detected lung cancer instances, and the other presenting the general output generated by the CNN Model. These visual outputs provide valuable insights into the prognostic capabilities and overall performance of the CNN model. Figure 12 presents a comparative graphical analysis between the performance metrics of the SVM classifier and the CNN Model. This comparative graph facilitates a comprehensive evaluation of the two models' respective effectiveness in lung cancer detection tasks

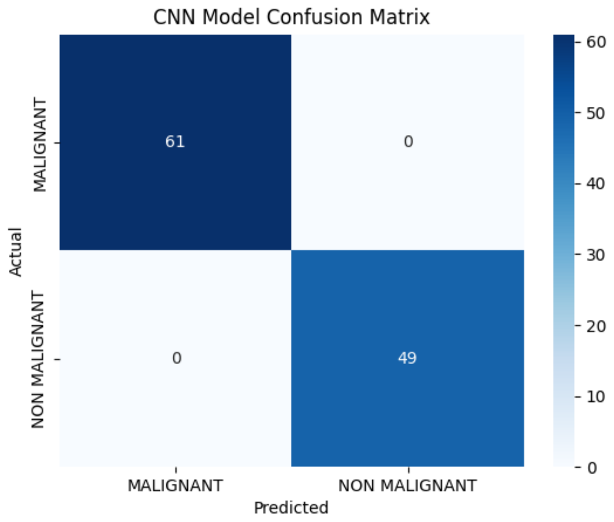

Figure 10: Confusion matrix of proposed CNN model.

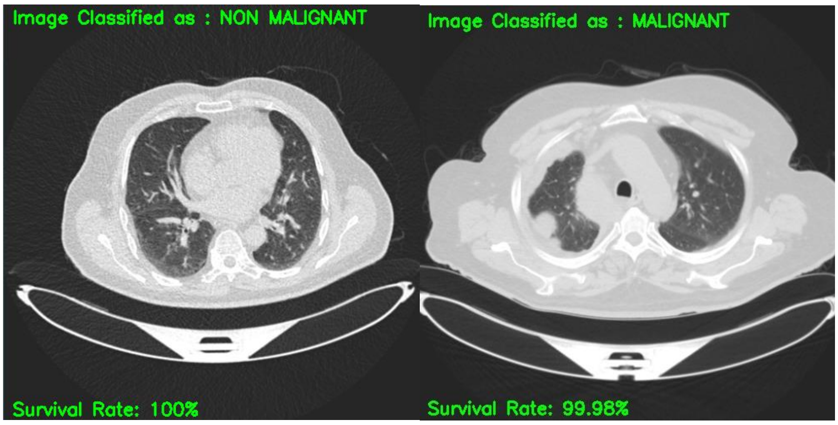

Figure 11: Prediction of survival rate and output using CNN Model.

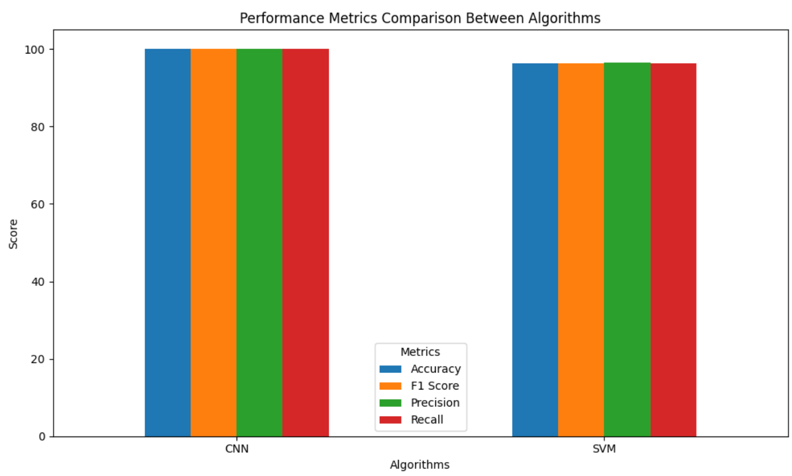

Figure 12: Performance comparison graph of SVM classifier and CNN Model.

5. CONCLUSION

In conclusion, the implementation of advanced modifications in convolutional neural network (CNN) models for the detection of lung cancer from chest scan images represents a significant advancement in early-stage diagnosis and monitoring of lung cancer. The proposed CNN model exhibits superior performance in classifying benign and malignant cases, distinguishing between normal and cancerous conditions with higher accuracy compared to conventional machine learning approaches like support vector machine (SVM) classifiers. By harnessing the power of deep learning, clinicians can benefit from more accurate and efficient detection and classification of lung nodules, enabling earlier intervention and better patient outcomes. The incorporation of deep learning-based medical imaging tools into clinical practice enhances diagnostic capabilities and supports medical professionals in making informed decisions for patient care.

REFERENCES

[1] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, H.J. Aerts, Deep learning predicts lung cancer treatment response from serial medical imaging, Clin. Cancer Res. 25 (11) (2019) 3266–3275.

[2] M.I. Faisal, S. Bashir, Z.S. Khan, F.H. Khan, An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer, December, in: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), IEEE, 2018, pp. 1–4.

[3] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Feny¨o, A. Tsirigos, Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning, Nat. Med. 24 (10) (2018) 1559–1567.

[4] D.M. Ibrahim, N.M. Elshennawy, A.M. Sarhan, Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, Comput. Biol. Med. 132 (2021), 104348.

[5] Avanzo, J. Stancanello, G. Pirrone, G. Sartor, Radiomics and deep learning in lung cancer, Strahlenther. Onkol. 196 (10) (2020) 879–887.

[6] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET- based radiomics to predict the histological subtypes of lung cancer, Clin. Nucl. Med. 44 (12) (2019) 956–960.

[7] P. R. Radhika, R.A. Nair, G. Veena, A comparative study of lung cancer detection using machine learning algorithms, 2019, February, in: IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2019, pp. 1–4.

[8] K. Pradhan, P. Chawla, Medical Internet of things using machine learning algorithms for lung cancer detection, J. Manag. Anal. 7 (4) (2020) 591–623.

[9] S. Bhatia, Y. Sinha, L. Goel, Lung cancer detection: a deep learning approach, in: Soft Computing for Problem Solving, Springer, Singapore, 2019, pp. 699–705.

[10] H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y.G. Ji, Y. Choi, Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes, ACS Nano 14 (5) (2020) 5435–5444.

[11] V. Rajasekar, B. Predi´c, M. Saracevic, M. Elhoseny, D. Karabasevic, D. Stanujkic, P. Jayapaul, Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm, Sci. Rep. 12 (1) (2022) 1–11.

[12] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET- based radiomics to predict the histological subtypes of lung cancer, Clin. Nucl. Med. 44 (12) (2019) 956–960.

[13] Y. She, Z. Jin, J. Wu, J. Deng, L. Zhang, H. Su, C. Chen, Development and validation of a deep learning model for non–small cell lung cancer survival, JAMA Netw. Open 3 (6) (2020) e205842, e205842.