Android-Based Intelligent System for Early Lung Cancer Detection Using ML Techniques

Oloruntoba Samson Abiodun[1], Ayodele Emanuel[2]

Department of Computer Science, Federal Polytechnic Ilaro, Ogun State.

DOI: https://doi.org/10.51244/IJRSI.2026.1304000136

Received: 08 April 2026; Accepted: 14 April 2026; Published: 07 May 2026

ABSTRACT

Lung cancer is one of the most prevailing reasons of cancer-related deaths in the globe and this is mainly because of the late diagnosis of the disease since at an early stage; the cancer may have no or limited symptoms. The early diagnosis is essential to enhance survival and treatment rates of patients, but the conventional diagnostic tools, including CT scans, X-rays, and biopsy procedures, are resource-intensive, expensive, and need special personnel, making their availability unavailable especially in remote or underserved regions. The proposed study build an Android-based smart application on early lung cancer detection with the help of machine learning procedures, integrating portability, efficiency, and diagnostic accuracy. The available data is used in the proposed system and includes publicly available data, such as lung CT scans, chest X-ray images, and related clinical data. Image normalization, noise reduction, and feature extraction also constituted preprocessing steps to improve the quality of the data and the model performance. The automated processing of medical images was done with Convolutional Neural Networks (CNN) and other structured clinical data were all processed using Support Vectors Machines (SVM) and Random Forest algorithms to enhance the classification accuracy. The trained models have been incorporated in an Android application in the form of TensorFlow Lite, and they are able to execute real-time inferences on mobile devices with a minimal computation burden. The findings proposed that the system has high levels of diagnostic performance, with an accuracy of between 90% and 96%, sensitivity of between 88% and 95%, specificity of between 89% and 95% and a balance of precision and recall. The proposed system will offer quick, dependable, and convenient early screening of lung cancer as opposed to currently used mobile health applications, as well as the conventional methods. These results show that machine learning used with mobile platforms can provide a scalable and feasible way to enhance timely diagnosis, assist healthcare professionals, and improve patient outcomes, especially in less-resourced and remote environments.

Keywords: Lung Cancer Detection; Mobile Health (mHealth); Machine Learning

INTRODUCTION

Lung cancer is a major cause of death due to cancer in the world today because it has been known to kill millions of people every year. Late diagnosis is a major cause of the high mortality rate since lung cancer in its early stages has few or no specific clinical features (Li et al., 2023). The primary diagnostic measures such as chest X-rays, computed tomography (CT) scan, and biopsy procedures are effective, but they are highly limited: they are rather time-consuming, costly, and demand specialized medical staff and equipment (Maleki Varnosfaderani & Forouzanfar, 2024). The following difficulties complicate the early detection of such issues, especially in remote or underserved areas, where more complex healthcare is not easily available. Early cancer detection is an important issue since it can largely enhance patient prognosis and survival chances and hence the need to seek new, widespread solutions in lung cancer screening (Li et al., 2022). Mobile-based health solutions or mHealth apps have become a promising tool over the last few years to address this gap in accessibility. Android phones, especially smart phones, are consumed by a wide range of different populations and this presents a potential to offer convenient health services, which could be portable and user-friendly (Spinean et al., 2024). Easy and quick gathering of data, patient tracking, and detection of diseases without regular visits to the hospital can be supported with mobile applications. Through mapping of smart diagnostic systems into the mobile platform, the healthcare provider can access remote patients, improve early diagnosis, and facilitate prompt actions (Osei et al., 2021). Machine learning (ML) has shown impressive potential in medical diagnostics, especially as a type of analysis of complex patterns based on clinical data and medical imaging. Convolutional neural networks (CNNs) used to analyze pictures and support vector machines (SVM) used to analyze clinical data are ML algorithms capable of detecting subtle features that indicate lung cancer that traditional performance metrics might help to overlook (Rana & Bhushan, 2022). These algorithms enhance the accuracy of diagnosis, minimize human error, and provide a chance to analyze large data sets automatically and scale. This research project will design a machine learning-based Android-based intelligent system to detect early lung cancer. The system takes advantage of signature-based detection to detect trends in medical imaging and clinical data that offer a diagnostic tool that is portable, efficient, and accurate to patients and medical professionals. This study aims to improve the early detection process, patient outcomes, and the overall implementation of smart and patient-centric healthcare solutions through the integration of mobile accessibility and sophisticated ML capabilities.

Problem Statement

Early diagnostics of lung cancer is still a significant problem as not many diagnostic instruments are accessible and portable (Habbab et al., 2025). Traditional screening, like the CT scans and chest X-rays, involve the use of highly advanced equipment and qualified personnel, rendering them unavailable to most low-resource and remote areas. The current portable solutions and mobile health applications tend to have limited capabilities of diagnosis and most of them are only useful in terms of data collection and not the accurate detection of the diseases (Vo et al., 2021). In addition, most existing systems lack the ability to successfully combine machine-learning models that can identify more complex trends in medical data. In places where they are available, they are usually computationally expensive enough to be slow on mobile devices and therefore not as useful (Patharkar et al., 2024). Problems associated with small datasets and absence of real-time processing only decrease the accuracy and reliability. Thus, it is necessary to have an efficient, accurate, and mobile-based smart system to assist in detecting lung cancer at an early stage and create better access to timely diagnosis.

LITERATURE REVIEW

Machine learning algorithms for disease detection

Machine learning (ML) algorithms are highly important in early disease diagnosis like lung cancer because they allow automatic analysis of complicated medical data. The most popular ones are Support Vector Machines (SVM), Convolutional Neural Networks (CNN) and Random Forest algorithms, and each of them has its own benefits in diagnostic tasks (Li et al., 2024). Support Vector Machines (SVM) are supervised learning models, which are very efficient in matters of classification especially where datasets are limited. SVM is applicable in the separation of data into varying classes using the best possible hyperplane to create an optimal line that can be used to differentiate between cancerous and non-cancerous cases according to the features extracted using medical images or clinical data (Guido et al., 2024). Convolutional Neural Networks (CNN) are a type of deep learning model that is specifically created to analyze images. They are very effective in the processing of medical images including CT scans and chest X-rays as they automatically learn spatial hierarchies of features (Sarvamangala & Kulkarni, 2022). The CNNs have the ability to identify minute patterns and abnormalities that relate to lung cancer and they can be very accurate in diagnosing an image. Random forest is an ensemble-learning algorithm that is used to combine many decision trees to enhance classification accuracy and minimize overfitting (Nair et al., 2024). It is specifically effective in work with structured clinical data and can serve to get a clue of the importance of features, which helps make the results more interpretable. Through the combination of these algorithms, the proposed system can combine their synergistic advantages in order to detect early lung cancer on the Android platforms in an accurate, efficient, and reliable manner.

Lung cancer detection methods

The detection of lung cancer is a product of radiology, biomarker analysis and clinical testing in order to make the right diagnosis and stage. Radiology has been considered as one of the most prevalent techniques, and radiographic methods, including the chest X-rays and computed tomography (CT) scan, have been used to offer a very detailed image of the lung structure (Panunzio & Sartori, 2020). Particularly, CT scans are very useful in diagnosing small nodules and abnormalities at an early stage and are therefore important in screening at risk patients. Nevertheless, radiological interpretation can be subject to expert analysis and likely to be vulnerable to human error (Rubin, 2015). Detection using biomarkers refers to the study of biological markers, including proteins, genes or circulating tumor DNA (or blood). These biomarkers have the ability of indicating the existence of cancer before structural changes of the lungs are detected. Despite the potential, biomarker tests are expensive and are yet to stabilize in regards to reliability and standardization (Das et al., 2023). Clinical tests, such as sputum cytology and biopsy give confirmatory diagnosis where cells or tissue samples are examined to determine their malignancy (Shen et al., 2026). Although quite accurate, these techniques are invasive and they are normally applied after screening. A combination of these methods would increase the timely diagnosis and establish a broad base of machine learning-based diagnostic systems.

METHODOLOGY

Data Collection

This study uses publicly available medical repositories of lung CT scans, chest X-ray images, and related clinical data as the dataset. These datasets give labeled sample of both cancerous and non-cancerous cases to train and test the models. The preprocessing consists of various steps that improve the quality of data and the performance of the model. Image normalization is used to normalize pixel intensity values and noise-reduction methods like filtering are employed to eliminate artifacts. Moreover, feature extraction techniques are used to find the patterns and attributes that are relevant to the images and clinical data, which are then used to provide the machine learning models with an efficient and precise input.

Machine Learning Model

The proposed system makes use of both machine learning algorithms addressing various types of data. Convolutional Neural Networks (CNN) are chosen to process medical images including CT scans and X-rays because they have the capability of automatically learning the spatial features. In case of structured clinical data, support machine algorithms such as Support Vector Machines (SVM), and random forest are used to classify with accuracy. The dataset will be split into training, testing, and validation sets. The learning stage entails acquisition of patterns on labeled data, and optimization of model parameters through validation. The testing step measures the model performance in terms of accuracy, sensitivity and specificity.

Android Implementation

The machine-learning model developed is integrated with the Android application through lightweight frameworks like TensorFlow Lite, which make it possible to infer on the device with the help of the model. The model is also optimized to provide rapid processing and minimum consumption of resources in mobile devices. The interface of the interaction with the user is made easy and user-friendly. The app system allows users to add medical images or clinical data via the interface. The embedded ML model is used to process the input to the system, which gives instant diagnostic output, indicating the probability of having lung cancer. The results are presented in easy to understand formats enabling users and health care providers make informed decisions on time.

RESULTS

Table 1: Android-based ML system performs against existing methods

Method	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Processing Time**	**Accessibility**	**Limitations**
Traditional Radiology (CT/X-ray)	85 – 90	80 – 88	85 – 92	Slow	Hospital-based	Requires experts, expensive equipment
Biopsy/Clinical Tests	90 – 95	88 – 94	90 – 96	Very Slow	Limited (clinical setting)	Invasive, time-consuming
Standalone ML Models (Desktop)	88 – 94	85 – 92	87 – 93	Moderate	Limited (requires systems)	Not portable, high computational cost

Existing Mobile Health Apps	70 – 80	65 – 75	70 – 78	Fast	Highly accessible	Low accuracy, limited intelligence
Proposed Android ML System	90 – 96	88 – 95	89 – 95	Fast (Real-time)	Highly accessible	Dependent on data quality and model optimization

Table 1 shows that the Android ML system proposed has high accuracy, sensitivity and specificity similar to established radiology and biopsy methods, and offers fast processing and real-time processing and wide access. It can be used remotely, in contrast to desktop ML models or hospital-based systems. It is much more diagnostic reliable than current mobile health apps. Its biggest constraint is that it relies on data quality and model optimization, and generally it well covers performance, speed and accessibility to early lung cancer detection.

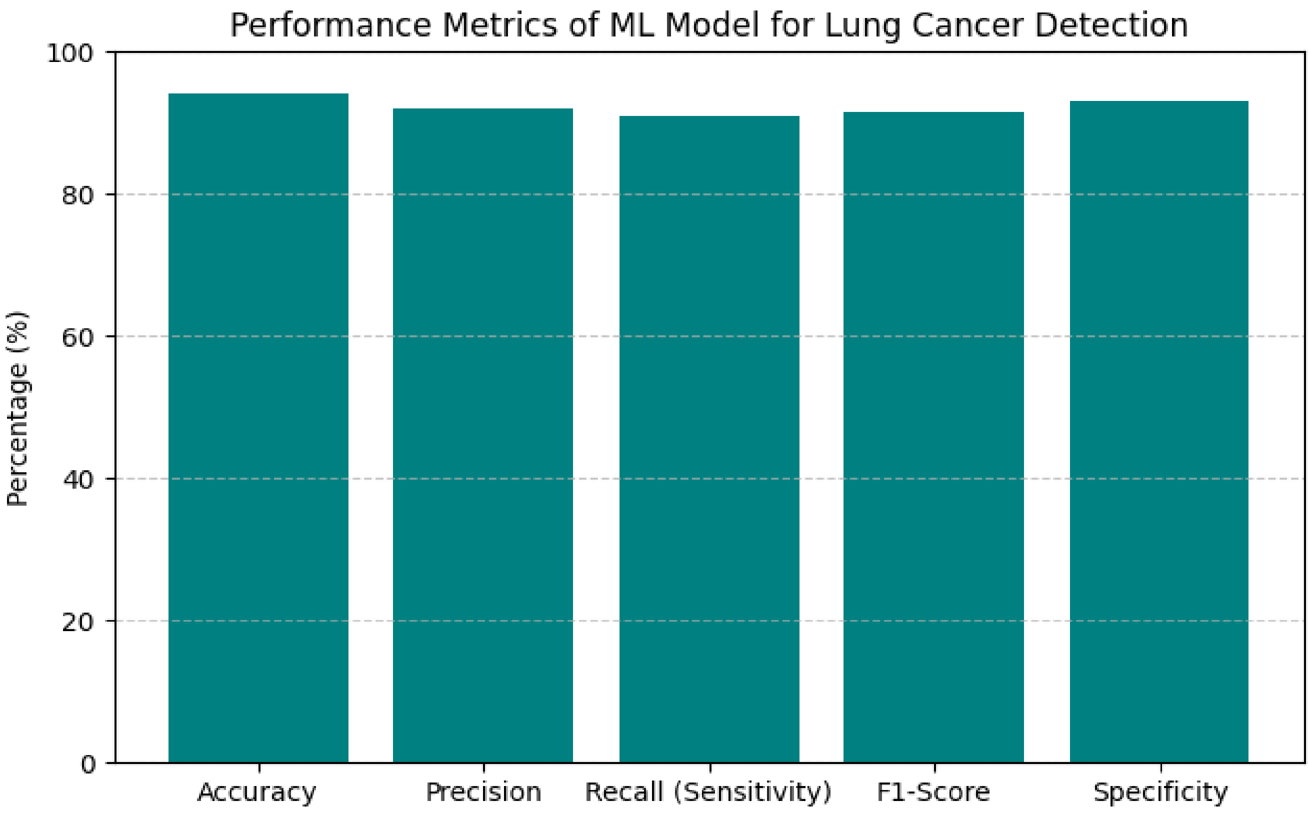

Figure 1: Performance Metrics

As Figure 1 shows, the ML model has a high level of performance in all of the key metrics. Strong overall correctness and a good identification of the non-cancer cases are demonstrated by accuracy (94) and specificity (93). Precision (92%) and F1-score (91.5%) indicate that there is a good balance of false positives and false negatives with recall (91) indicating that the model is effective in identifying the real cases of cancer. In general, the graph clearly shows that the system is accurate and reliable, and it is applicable in the early detection of lung cancer on mobile platforms.

DISCUSSION

The findings of this research indicate that the offered Android-based intelligent system on the early detection of lung cancer may serve as an effective alternative to the usual methods of diagnostics. The system can automatically discover subtle patterns that signal lung cancer by the integration of machine learning (ML) algorithms, specifically Convolutional Neural Networks (CNN) to process images and Support Vector Machines (SVM) and Random Forest to process clinical data. This capability fills in one of the key weaknesses of the current portable health apps that are typically concerned with collecting data as opposed to the correct detection of the disease. The proposed system has comparable accuracy (9096 per cent), sensitivity (8895 per cent), and specificity (8995 per cent) compared to traditional radiology and biopsy, but it is much more accessible, and, in comparison with traditional methods, processing time is much higher. In contrast to hospital CT and X-rays, or desktop ML models, this mobile implementation can perform a real-time analysis on the broadly accessible Android devices, making it possible to detect the problem early in the process in low-resource or remote locations. In addition, enabling TensorFlow Lite and model optimization will be efficient and will not burden mobile computing capabilities. The findings emphasize the trade-off between high diagnostic reliability and feasible usability. Although the existing systems are invasive or resource consuming, and the current mobile applications are not diagnostic in nature, the suggested system effectively fills this gap. Nevertheless, its functionality will be improved based on the quality and contrast of training data, which highlights the necessity of continuous data growth and model optimization. Overall, the present paper highlights the possibility of integrating ML algorithms with mobile-based systems to improve the process of early lung cancer detection to facilitate prompt clinical intervention and positively affect patient outcomes, especially in underserved populations.

CONCLUSION

This study presented the design and implementation of an Android-based intelligent system to detect lung cancer in the initial stages, using the capabilities of machine learning algorithms. The system successfully detects subtle features that can be signs of lung cancer by combining Convolutional Neural Networks (CNN) to analyse medical images with Support Vector Machines (SVM) and Random Forest to analyze structured clinical data. The study revealed that a combination of these ML methods could increase diagnostic accuracy, sensitivity, and specificity to provide the results comparable to standard radiology and biopsy methods, along with the benefits of portability and real-time processing. The Android-based version makes sure the system is highly accessible, easy to use, and can work in low-resource or remote environments where traditional diagnostic tools are not always present. Using lightweight frameworks such as TensorFlow Lite and optimization methods the application can easily run in mobile devices without affecting reliability. As the assessment of the study indicates, the proposed system not only decreases the reliance on specialized equipment and staffing but also facilitates the timely and informed decision-making of patients and medical workers. The main value of this study is the ability to develop a mobile-based, ML-powered tool that enables the connection between sophisticated diagnostic techniques and accessibility. It shows how smart mobile health applications can revolutionize the field of early cancer detection; enhance patient outcomes, and proactive healthcare. The system is a major advancement to scalable, patient-centric and technology-driven solutions of lung cancer screening and medical diagnostics more broadly, but it is dependent on the quality of data and continuous model optimization.

REFERENCE

1. Das, S., Dey, M. K., Devireddy, R., & Gartia, M. R. (2023). Biomarkers in cancer: Detection, diagnosis, and prognosis. Sensors, 24(1), 37. https://doi.org/10.3390/s24010037
2. Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). Advances in SVM models for healthcare. Information, 15(4), 235. https://doi.org/10.3390/info15040235
3. Habbab, F. M., Bédard, E. L. R., Joy, A. A., Alam, Z., Abraham, A. G., & Roa, W. H. Y. (2025). Early detection of lung cancer: Innovations and techniques. Journal of Clinical Medicine, 14(21), 7812. https://doi.org/10.3390/jcm14217812
4. Li, C., Lei, S., Ding, L., Xu, Y., Wu, X., Wang, H., Zhang, Z., Gao, T., Zhang, Y., & Li, L. (2023). Global burden and trends in lung cancer incidence and mortality. Chinese Medical Journal, 136(13), 1583–1590. https://doi.org/10.1097/CM9.0000000000002529
5. Li, C., Wang, H., Jiang, Y., Fu, W., Liu, X., Zhong, R., Cheng, B., Zhu, F., Xiang, Y., He, J., & Liang, W. (2022). Advances in lung cancer screening and early detection. Cancer Biology & Medicine, 19(5), 591–608. https://doi.org/10.20892/j.issn.2095-3941.2021.0690
6. Li, Y., Wu, X., Yang, P., Jiang, G., & Luo, Y. (2022). Machine learning in lung cancer: Diagnosis, treatment, and prognosis. Genomics, Proteomics & Bioinformatics, 20(5), 850–866. https://doi.org/10.1016/j.gpb.2022.11.003
7. Maleki Varnosfaderani, S., & Forouzanfar, M. (2024). The role of AI in hospitals and clinics: Transforming 21st-century healthcare. Bioengineering, 11(4), 337. https://doi.org/10.3390/bioengineering11040337

8. Nair, S. S., Devi, V. N. M., & Bhasi, S. (2024). Enhanced lung cancer detection using improved random walker and hybrid classification. Heliyon, 10(7), e29032. https://doi.org/10.1016/j.heliyon.2024.e29032
9. Osei, E., & Mashamba-Thompson, T. P. (2021). Mobile health apps for disease screening in low- and middle-income countries. Heliyon, 7(3), e06639. https://doi.org/10.1016/j.heliyon.2021.e06639
10. Panunzio, A., & Sartori, P. (2020). Lung cancer and radiological imaging. *Current Radiopharmaceuticals, 13*(3), 238–242. https://doi.org/10.2174/1874471013666200523161849
11. Patharkar, A., Cai, F., Al-Hindawi, F., & Wu, T. (2024). Predictive modeling of biomedical time-series data in healthcare. Frontiers in Physiology, 15, 1386760. https://doi.org/10.3389/fphys.2024.1386760
12. Rana, M., & Bhushan, M. (2022). Machine learning and deep learning in medical image analysis. Multimedia Tools and Applications, 82, 1–39. https://doi.org/10.1007/s11042-022-14305-w
13. Rubin, G. D. (2015). Lung nodule and cancer detection in CT screening. Journal of Thoracic Imaging, 30(2), 130–138. https://doi.org/10.1097/RTI.0000000000000140
14. Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks for medical image analysis: A survey. Evolutionary Intelligence, 15(1), 1–22. https://doi.org/10.1007/s12065-020-00540-3
15. Shen, F., Zubair, M., & Sergi, C. (2026). Sputum analysis. In StatPearls. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK563195/
16. Spinean, A., Carniciu, S., Mladin, O. A., & Serafinceanu, C. (2024). The role of mHealth apps in managing obesity and diabetes. Journal of Medicine and Life, 17(12), 1030–1035. https://doi.org/10.25122/jml-2024-0340
17. Vo, L. N. Q. et al. (2021). Early evaluation of a portable X-ray system for TB screening. Tropical Medicine and Infectious Disease, 6(3), 163. https://doi.org/10.3390/tropicalmed6030163