Research Article

Development of a Lung Cancer Detection System Using Residual Network 50 on Computed Tomography Scans

Angel Omoikhefe Egwaoje[1], A. Abiodun Oguntimilehin[2]

1. Department of Computing, Afe Babalola University, Nigeria; 2. Department of Computer Science, Federal University of Technology and Environmental Sciences, Nigeria

Lung cancer has continued to be the primary cause of deaths from cancer worldwide, and this is due to the late diagnosis of this disease. This study shows how deep learning can be used to detect lung cancer at an early stage, using lung CT scan images. A pre-trained Residual Network with 50 layers (ResNet50) model was used with transfer learning to classify lung CT images as either cancerous or non-cancerous. The CT images used in this study were sourced from three publicly available datasets: the IQ-OTH/NCCD, the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), and the Bernard Institute of Radiology (BIR) Lung dataset. Since the original dataset had a majority of one class, it was balanced to a 50:50 ratio, resulting in a total of 1,720 images. The images were then split into training, validation, and testing sets at a 70:15:15 ratio. The model showed strong performance in testing, achieving an accuracy of 98.46%, a recall of 99.21%, a precision of 97.69%, and an F1-score of 98.45%. The trained model was then built into a web application, giving healthcare professionals a useful tool for classifying lung CT scans. In conclusion, this system has shown the role that deep learning can play in detecting lung cancer at an earlier stage and helping doctors make better and faster decisions.

Corresponding author: A. Angel Egwaoje, egwaojeangel@gmail.com

 doi.org/10.32388/4QGQJX

Introduction

Background

Cancer is a disease caused when the body's cells divide chaotically, which is considered abnormal; these cells spread through the body by invading nearby tissues or organs. It can begin anywhere within the body's trillions of cells. Cancer develops due to changes in the DNA, but it occurs primarily in the genes [1]. Approximately 20 million people were newly diagnosed with cancer in 2022, and 9.7 million of these cases proved fatal. Statistics further show that cancer affects one in five people during their lifetime, with mortality rates of about one in nine for men and one in twelve for women [2]. Although the origin of cancer development still remains unclear, several influences that cause cancer growth have been identified. Some factors that cause cancer growth are tobacco intake and obesity; these factors are possibly changeable, while others, such as genetic changes which are inherited from parents, are not. These risk factors may act alongside or in an arrangement to stimulate cancer development [1]. The signs and symptoms of cancer are not the same for all types of cancer; they vary based on the location in which they are formed, the size of the growth and how intensely they affect nearby organs or tissues. The most common symptoms of cancer are tiredness, fever and weight loss. These symptoms occur because cancer consumes a large portion of the body's energy supply and cancer cells also release substances that negatively affect the way the body produces energy, which leads to changes in how the body functions, bringing about weakness in the immune system and triggering these symptoms [3]. According to the World Health Organisation, the effect of the various symptoms of cancer on patients' lives calls for the need for early identification of cancer [4]. Diagnosing cancer at an early stage can lead to improvements and positive results in the lives of cancer patients by preventing delays in treatment. The most common types of cancer experienced by people are lung cancer, breast cancer, colon cancer, rectal cancer and prostate cancer, and these cancer types have a high chance of being cured when managed properly with best practices [5].

Each year, a large number of individuals are affected by lung cancer, and this accounts for a significant number of deaths caused by cancer worldwide, with around 2 million new cases and 1.76 million fatalities among both men and women [6]. Approximately 80–90% of lung cancer deaths are caused by tobacco smoking, while the leading cause of lung cancer in individuals that do not smoke tobacco is exposure to radon gas. However, the incidence and death rates of lung cancer vary across the world because of different human lungs, reflecting terms of tobacco use, genetic influences and exposure to environmental risk factors [7]. A cough is the most frequently observed, though non-specific, symptom of lung cancer.

doi.org/10.32388/4QGQJX

Symptoms such as a bloody cough or shortness of breath, along with general symptoms such as loss of appetite and weight loss, increase an individual's chances of being diagnosed with lung cancer [8]. There are two categories of lung cancer, and they are of two primary types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). SCLC originates in the bronchi, accounts for approximately 20% of lung cancer cases, and tends to spread quickly to other parts of the body, including the lymph nodes. It is primarily caused by tobacco smoking and is the leading cause of deaths caused by cancer. NSCLC is responsible for about 90%, which is the majority of lung cancer cases, and tends to advance more gradually compared to SCLC. Lung cancer detection at an early stage is key to better patient survival. Despite improvements in initial diagnosis, lung cancer care continues to face some challenges, particularly in data analysis, and this is due to high clinical workloads [9]. Identifying if an individual has cancer at an early stage is essential because it allows for timely interventions, especially in high-risk individuals, which can help to prevent this disease from progressing, thereby improving patient outcomes. The five-year survival rate for lung cancer is much lower compared to many other cancer types, and this explains why lung cancer is the main cause of deaths caused by cancer worldwide. Although various methods for diagnosing and treating lung cancer are in use, survival outcomes for lung cancer patients still remain a concerning problem; this points out the need for better and more advanced tools in clinical practice. To address these issues, advancements in computer technology and statistics have brought about artificial intelligence (AI) as a much more promising solution for lung cancer diagnosis, particularly in tumour imaging [3].

Diagnosing lung cancer requires imaging of the chest area, as that is the location of the lungs in individuals. It is usually done with low-dose computed tomography (CT) to reduce radioactivity exposure in individuals, and this is the primary method for monitoring individuals at higher risk of lung cancer. [10] Early computer-aided diagnosis (CAD) systems relied on collecting features from the CT scan images that are considered important and common; measurements focusing on nodule size, count, and shape are the features most health professionals look out for. However, these measurements were often subjective and open to interpretation, leading to inconsistent and incomplete outputs [11]. These limitations have shown that there is still a need for better tools, creating a way for more machine learning advancements. Machine learning helps to prevent cancer cases from getting worse by detecting or predicting this disease at an early stage, allowing for early intervention and more positive patient outcomes [12].

Machine learning (ML), a subset of AI, focuses on enabling systems to learn from data [13], and numerous studies have demonstrated that it is an effective tool for the early detection of cancer [14]. Patient data are analysed by ML algorithms to identify nodules or early-stage signs of cancer that may be overlooked by healthcare professionals [15]. This helps to improve the accuracy and reduce the time consumed in diagnosing lung cancer, allowing for intervention at an early stage, which is crucial for improving survival rates in lung cancer patients [16].

The field of deep learning (DL) has been proven by several studies to improve lung cancer detection [17]. Convolutional Neural Networks (CNNs) have proven to be highly effective and high-performing in analysing medical images, including CT scans for lung cancer detection [18]. These models can automatically extract important features from imaging data, which leads to a much more accurate and efficient identification of potentially cancerous nodules present in the lungs [15]. However, DL uses artificial neural networks to learn patterns in raw data and offers a much better solution in this case, as it enables systems to handle larger datasets and improves the ability to generalise across diverse datasets [13].

Problem statement

Lung cancer continues to be a major health problem worldwide; an approximate 2.5 million people are diagnosed with this terminal disease each year, and nearly 1.8 million people die from it annually [19]. These figures are associated with late-stage diagnosis and limited screening coverage in many regions [20]. In 2022, lung cancer was estimated to cause about 18.7% of all cancer deaths, which shows the urgency of improved prevention and early intervention to increase patients' survival rates [19].

Recent studies have shown that although AI tools can improve the accuracy of lung cancer diagnostics and decrease the time radiologists spend interpreting CT scans, challenges remain. AI models show variability in performance depending on different patient data, demographics, and types of scans. This showcases the necessity for ongoing improvement to achieve consistent and reliable results in early-stage lung cancer detection [21].

Review of related works

In 2023, Ansari et al. [22] carried out a study in which the researchers made use of six CNN models: ResNet-50, VGG-16, VGG-19, ResNet-101, DenseNet-201, and EfficientNet-B4 to help detect lung cancer in CT scan images. The study made use of the LIDC-IDRI dataset. Their objective was to classify the presence of lung cancer into three types—squamous cell carcinoma (SCC), large cell carcinoma (LCC), and adenocarcinoma

doi.org/10.32388/4QGQJX

(ADC)—and normal benign (which means there is little or no cancer in the lung). The lung CT images were resized to 460 x 460 x 3 pixels as part of the preprocessing process; the models were also assessed in both transfer learning and in their original form for better generalisation. Transfer learning was useful in this case with DenseNet-201 and EfficientNet-B4, whereby these tools achieved validation accuracies of 96.88% and 96.53%, respectively. The study also showcased that hyperparameter tuning and the strength of transfer learning in deep learning models, especially DenseNet-201, helped the model perform more consistently. Although the study shows good results, as of today, the researchers still have not deployed the model as a usable application, and even though the model had a strong performance, the study noted that there was still overfitting during validation.

In 2021, a study was carried out by Abid et al. [23] and the aim of this research was to improve lung cancer detection by making use of a multi-view convolutional recurrent neural network (MV-CRecNet). The objective of this was to improve the identification of lung cancer nodules in CT scan images at an early stage by analysing the different parts of a lung that can be viewed: the axial, coronal, and sagittal planes. This method helps the model catch cancerous lung nodules early enough. This research made use of the LIDC-IDRI and ELCAP datasets. The model performed very well compared to the capabilities of all other models it was compared to. Despite the good results, the model still faced some challenges. The model's generalisability was limited because it was trained and tested only on the LIDC-IDRI and ELCAP datasets, and as of the publication of this study, the model has still not been deployed as a usable application.

In 2023, a study was carried out by Dadgar and Neshat [24]; they made use of a hybrid deep learning model based on convolutional networks to compare performance with other methods, and transfer learning was also used for lung cancer diagnosis. Several pre-trained CNN models were used to classify lung cancer into three types: Normal, Cancerous, and Benign. The models used in this study are ResNet152V2, VGG16, MobileNetV3 (both small and large versions), InceptionResNetV2, and EfficientNetV2. The researchers made use of 1,000 lung CT scans from a public dataset, which were preprocessed before being used. The InceptionResNetV2, with the help of transfer learning, achieved the highest performance of 91.1%, a precision of 84.9%, and an F1-score of 81.5% in classifying the three cancer types compared to the other models used. The results show that the model's performance is strong and reliable; even so, the study still faced limitations, such as a limited dataset. The authors stated that acquiring a dataset can be challenging and expensive, and as of the publication of this study, it had not explored the deployment of the model as a usable application.

doi.org/10.32388/4QGQJX

In 2023, a study was carried out by Mohamed et al. [12] in which the researchers made use of a hybrid Convolutional Neural Network (CNN) using the Ebola Optimisation Search Algorithm (EOSA) to classify lung CT scans into normal, benign and malignant tissues. The training process was performed on the IQ-OTH/NCCD dataset with a total of 1,097 images. The EOSA-CNN model achieved an overall accuracy of 93.21%, which strongly indicates that the model performed well and is reliable. The results showed that the EOSA-CNN model achieved specificities of 0.7941, 0.97951, and 0.9328; for the sensitivities of the classification tasks, the results were 0.9038 for normal cases, 0.1333 for benign cases, and 0.9071 for malignant cases. These findings show that the hybrid EOSA-CNN algorithm performs well in differentiating between lung tissue types, even though it has low sensitivity for benign cases, which suggests difficulty in correctly identifying that category. Although these results are encouraging, the study had some limitations. The dataset was imbalanced, which affected the model's ability to generalise by introducing bias towards the majority class. The training process required a large amount of computing power because of the optimisation algorithm used. The study, as of the publication date, has not explored the deployment of the model as a usable application.

In 2023, Raza et al. [18] carried out a study and applied a pretrained EfficientNetB1 convolutional neural network (CNN) to classify lung cancer CT scan images from the IQ-OTH/NCCD dataset. The model achieved very high classification accuracy of 99.10% in differentiating between benign, malignant and normal lung tissue using CT scan images, showing that the model performed very well. Although the results are promising, the authors stated that there were some limitations, which are a small dataset size and potential overfitting; they also did not provide detailed sensitivity and specificity metrics for different disease stages. These factors show that the model performs well but has not gone beyond research because of its limited generalisability and cannot be used in clinical environments. As of the date of publication, the model had not been deployed as a usable clinical application.

In 2023, Hendrix et al. [25] carried out a study which involved developing an AI system that analyses lung CT scans to identify whether lung nodules are harmless or cancerous in patients undergoing routine medical imaging. Their model performed very well; it achieved a sensitivity of 94.3% for benign nodules and 96.9% for lung cancers, which compares to or even surpasses the performance of several experienced radiologists. The study used CT scans collected from Dutch hospitals, rather than public databases, to ensure real-world clinical relevance. The aim of the model was to assist in the early detection of lung cancer by improving accuracy in diagnosis and reducing human error. However, despite the promising results that the model achieved, the researchers noted that the system is still in the research stage and has not yet been deployed as a usable application for clinical use.

In 2024, Gautam et al. [15] carried out a study with the aim of improving lung cancer detection using thoracic CT scans; they made use of an ensemble approach. The researchers utilised three convolutional neural networks: ResNet-152, DenseNet-169, and EfficientNet-B7. The ensemble model achieved strong results with an accuracy of 97.23% and a sensitivity of 98.6%. This approach proved to be better than traditional methods by reducing false positives and negatives. Although this study achieved very strong results, some limitations were still identified. The model still remains in the research phase and, as of the publication date, has still not been deployed as a usable application. The study also reported that the data were complex and the models required high computing power for both training and real-time application.

In 2025, Abdulqader et al. [11] carried out a study in which the researchers made use of a multi-objective deep learning model for lung cancer detection using CT images. The aim of this study was to improve tumour classification, localisation, and overall diagnostic accuracy for lung CT scans. They made use of three components: a transformer-based attention mechanism, adaptive anchor-free detection, and an improved feature pyramid network. This system's architecture allowed the model to learn important features and perform tasks simultaneously, such as identifying the presence of cancer and pinpointing the locations of the tumour within the lung. The model was trained on a dataset of 1,608 CT images (623 cancerous, 985 non-cancerous). The model achieved a mean Average Precision (mAP) of 96.26%, Intersection over Union (IoU) of 95.76%, precision of 98.11%, and recall of 98.83%. The model improved reliability in its detection ability, especially for early-stage lung cancer. Even though the model achieved these promising results, several limitations were still acknowledged. The dataset was small and specific to a single institution, which raised concerns about the model's ability to generalise to larger and more diverse patient populations. As of the date of publication, the model has not yet been deployed as a usable application.

In 2023, Shah et al. [16] carried out a study that suggested an ensemble model, combining multiple 2D Convolutional Neural Networks (CNNs) to improve the detection of lung cancer in CT scan images. The study made use of combining multiple CNN architectures. This ensemble approach achieved an overall accuracy of 95%; this method showed better accuracy compared to traditional methods that radiologists use, which can be time-consuming and sometimes inaccurate. The model also achieved a sensitivity of 95.9% and specificity of 94.5%, which shows that the model has a strong ability to differentiate between cancerous and non-cancerous nodules in the lungs. Although this study achieved these promising results, the system remains in the research phase and, as of the date of publication, it has still not been deployed as a usable application.

In 2024, Zhang et al. [21] carried out a study and built a deep learning approach to improve lung cancer diagnosis by combining Convolutional Neural Networks (CNNs) and DenseNet architectures, which were supported by mobile edge computing and data fusion techniques. The CNN was responsible for extracting important features from CT images, while DenseNet captured more complex patterns across many layers, which allowed it to understand the data better. This system analysed lung CT scans and classified the tissues into normal, benign, or malignant categories, achieving an impressive accuracy of 99% in the testing set. The researchers made use of data from multiple sources, such as imaging information and clinical data, through a data fusion process to further improve reliability in diagnosis. Despite the model's high performance, its effectiveness still depends on the availability of high-quality imaging data and controlled conditions. As of the publication date of this study, the system still remains in the research phase and has not yet been deployed as a usable diagnostic application that radiologists can use in real time.

The aim of this study is to develop an accurate and reliable deep learning system for early lung cancer detection using computed tomography (CT) scan images. The main contributions of this work include the design of a ResNet50-based binary classification model using transfer learning, which means the model has been trained on multiple images; the use of a balanced dataset that was obtained from three different sources to improve the model's generalisation; and the deployment of the trained model as a secure web-application diagnostic support system to assist clinical decision-making and reduce the time consumed.

Materials and Methods

Theoretical framework

Deep learning has become so powerful and effective in medical image analysis because of its ability to automatically learn important features directly from raw image data. In lung cancer detection, computed tomography (CT) scans provide high-resolution anatomical information that allows for the identification of abnormalities in the lungs, such as pulmonary nodules. However, the traditional method of using manual interpretation of CT images by radiologists or oncologists is time-consuming and may be affected by differences between observers, especially when nodules are still small or subtle. Automated deep

doi.org/10.32388/4QGQJX

learning systems can therefore serve as effective tools for supporting decisions for healthcare professionals.

Convolutional Neural Networks (CNNs) are suitable for image classification tasks due to their high capability for learning features effectively. A CNN consists of convolutional layers that extract important features from input images, followed by pooling layers that simplify the data and fully connected layers that perform classification. Early convolutional layers capture low-level features that are too small to be noticed by the average human, such as edges and textures, while deeper layers learn high-level important features that are useful for differentiating between cancerous and non-cancerous lung tissues.

Residual Networks (ResNets) help to address the drop in performance and vanishing gradient problems that occur when very deep networks are trained. ResNet50 also introduces residual learning through skip connections, which allows the network to learn a residual mapping rather than a direct mapping. This can be expressed as:

$$y = F(x, W) + x \tag{1}$$

As shown in Eq. (1), x represents the input to a residual block, F(x, $\{W_i\}$) is the residual function learned by the convolutional layers with weights $\{W_i\}$, and y is the output of the block. This structure allows for a smooth learning process and improves learning efficiency in deep architectures.

In this study, transfer learning was used by starting the ResNet50 model with weights pre-trained on a large image dataset. The pre-trained network provides generalised feature representations, which were fine-tuned for the lung cancer detection task. The classification problem was formulated as a binary classification task, with the output layer producing probabilities for cancerous and non-cancerous classes using the softmax function:

$$\mathrm{Softmax}(z_i) = e^{(z_i)} / \Sigma e^{(z_j)} \tag{2}$$

An objective function is required, as shown in Eq. (2), to measure how close the predicted probabilities are to the true labels. The binary cross-entropy loss function was used for optimisation. Model training was optimised using the binary cross-entropy loss function, defined as:

$$L = -(1/N)\Sigma[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \tag{3}$$

In Eq. (3), N represents the number of training samples, y_i is the true class label, and $\hat{y}_i$ is the predicted probability produced by the model. Performance was evaluated using accuracy, precision, recall, and F1-score, with recall being particularly important in medical diagnosis to minimise false negatives. The

doi.org/10.32388/4QGQJX

theoretical foundation presented here supports ResNet50 being a suitable deep learning tool for reliable lung cancer detection from CT scan images.

Dataset collection

CT scan images were obtained from three publicly available datasets: the IQ-OTH/NCCD, LIDC-IDRI [26], and the BIR Lung dataset [27]. The combined dataset consisted of 1,782 CT images categorised as cancerous and non-cancerous. Of these combined datasets, 922 images were cancerous and 860 were non-cancerous.

To address class imbalance and reduce bias towards the majority class, the dataset was balanced to a 50:50 ratio by randomly selecting 860 cancerous images to match the number of non-cancerous samples. This resulted in a final dataset of 1,720 CT scan images.

Data preprocessing

The balanced dataset was then divided into training, validation, and testing sets using a 70:15:15 ratio, corresponding to 1,200 training images and 260 images each for validation and testing. Preprocessing was then applied independently to each subset to prevent data from being leaked.

All images were resized to 224 × 224 pixels using the LANCZOS resampling method and were converted to Red-Green-Blue (RGB) format because ResNet50 is trained on coloured images. Pixel values were normalised using a mean of 0.485 and a standard deviation of 0.229 to improve model convergence.

Data augmentation

To improve generalisation and reduce overfitting, data augmentation techniques were applied only to the training set. These included horizontal and vertical flipping, 10° rotation, random cropping, and colour jittering to simulate variations in imaging conditions, so that the model can learn effectively.

Model architecture

ResNet50, which is a pre-trained convolutional neural network that was trained on ImageNet, was adapted for binary classification. All layers were frozen except for the final convolutional block (Layer 4) and the fully connected classification layer. A dropout layer with a rate of 0.5 was introduced before the output layer to reduce overfitting so that the model can generalise better. The final layer was then changed to produce only two output classes: cancerous and non-cancerous.

Training and optimisation

The model was built using PyTorch deep learning in Python. The training process was performed using the Adaptive Moment Estimation (Adam) optimiser with a learning rate scheduler to adjust the learning rate to help the model learn steadily. Rectified Linear Unit (ReLU) activation functions were used after each convolutional layer to make the model learn more complex patterns. Model performance was monitored using the validation dataset during training, and early stopping was used to terminate training when no improvement in validation loss was observed over a predefined number of epochs.

System deployment

After evaluation, the trained model was deployed as a web application and named LUNNY. The system provided a user interface that allows users to upload lung CT scan images and receive classification results almost immediately, along with confidence scores to describe how sure the model is about its results. The deployment process was documented to support anyone interested in reproducibility, maintenance, and future system improvements. The overall development workflow is illustrated in Fig. 1.

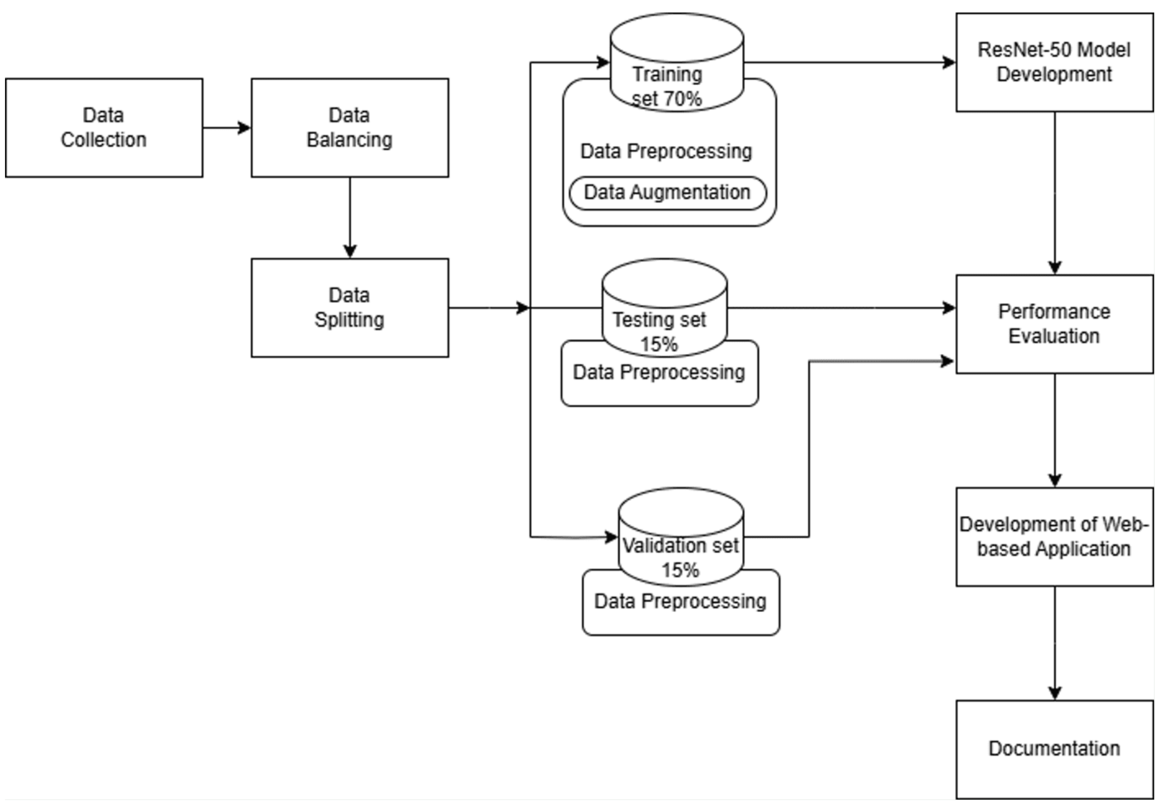

Figure 1. *Architecture of the Proposed System*

doi.org/10.32388/4QGQJX

Results

System implementation and model training

The lung cancer detection system was built using ResNet50 with transfer learning. The model was trained on a balanced dataset of 1,720 CT scan images, which was obtained from three publicly available datasets: the IQ-OTH/NCCD, LIDC-IDRI, and BIR Lung datasets. All the images were resized to 224 × 224 pixels and normalised before starting the training process. Flipping, rotation, cropping, and colour jittering, which are data augmentation techniques, were applied only to the training set to improve the model's ability to generalise even with diverse datasets.

During training, the early layers of the ResNet50 network were frozen to preserve learned low-level image features that were learned from ImageNet, while the final convolutional block and fully connected layers were fine-tuned for the lung cancer classification task. Model performance was monitored using the validation dataset to prevent overfitting through early stopping. The overall development workflow of the system is shown in Fig. 1.

Model performance evaluation

The trained model was evaluated using an unseen test dataset of lung CT scan images to test the performance of the model properly. The performance was assessed using accuracy, precision, recall, and F1-score, which are the most commonly used metrics in medical image classification tasks. The results achieved show how strong and consistent the performance is across all metrics used for evaluation.

Metric	**Value (%)**
Accuracy	98.46
Precision	97.69
Recall	99.21
F1-score	98.45

Table 1. Performance metrics of the proposed ResNet50 model.

doi.org/10.32388/4QGQJX

Web-based system deployment and user interface

The ResNet50 model was trained and was then deployed into a web application named LUNNY to support usability in the real world to help healthcare professionals. The landing page of the application is shown in Fig. 2, which requires authenticated access to prevent unauthorised access. The terms and conditions interface is presented in Fig. 3 to ensure responsible use of the system.

After authentication, users are redirected to the dashboard shown in Fig. 4, which provides access to scan upload, patient record management, and system logout functionalities. The scan upload interface is illustrated in Fig. 5, where users can submit CT scan images for analysis.

Once a scan is uploaded, the system processes the image and displays the classification result, confidence score, and Lung Reporting and Data System (Lung-RADS) category, as shown in Fig. 6. Patient data management interfaces are shown in Figs. 7-9, enabling healthcare professionals to store, edit, and review patient records efficiently.

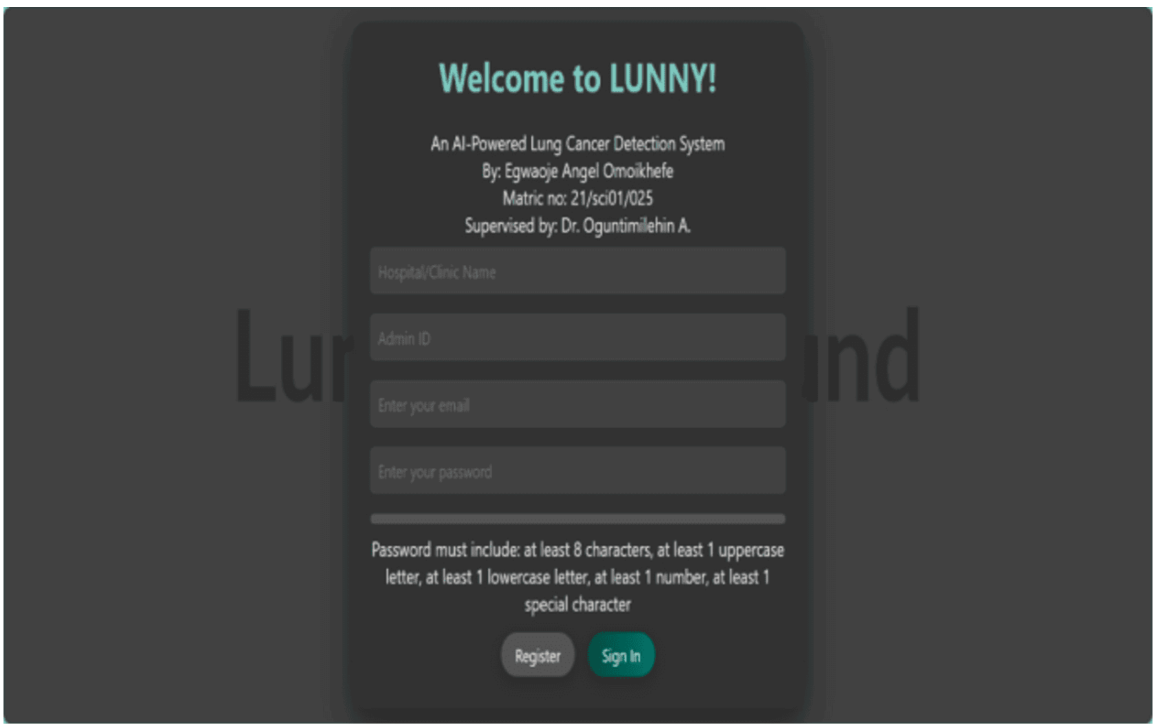

Figure 2. *LUNNY: Screenshot of the Landing Page Interface.*

doi.org/10.32388/4QGQJX

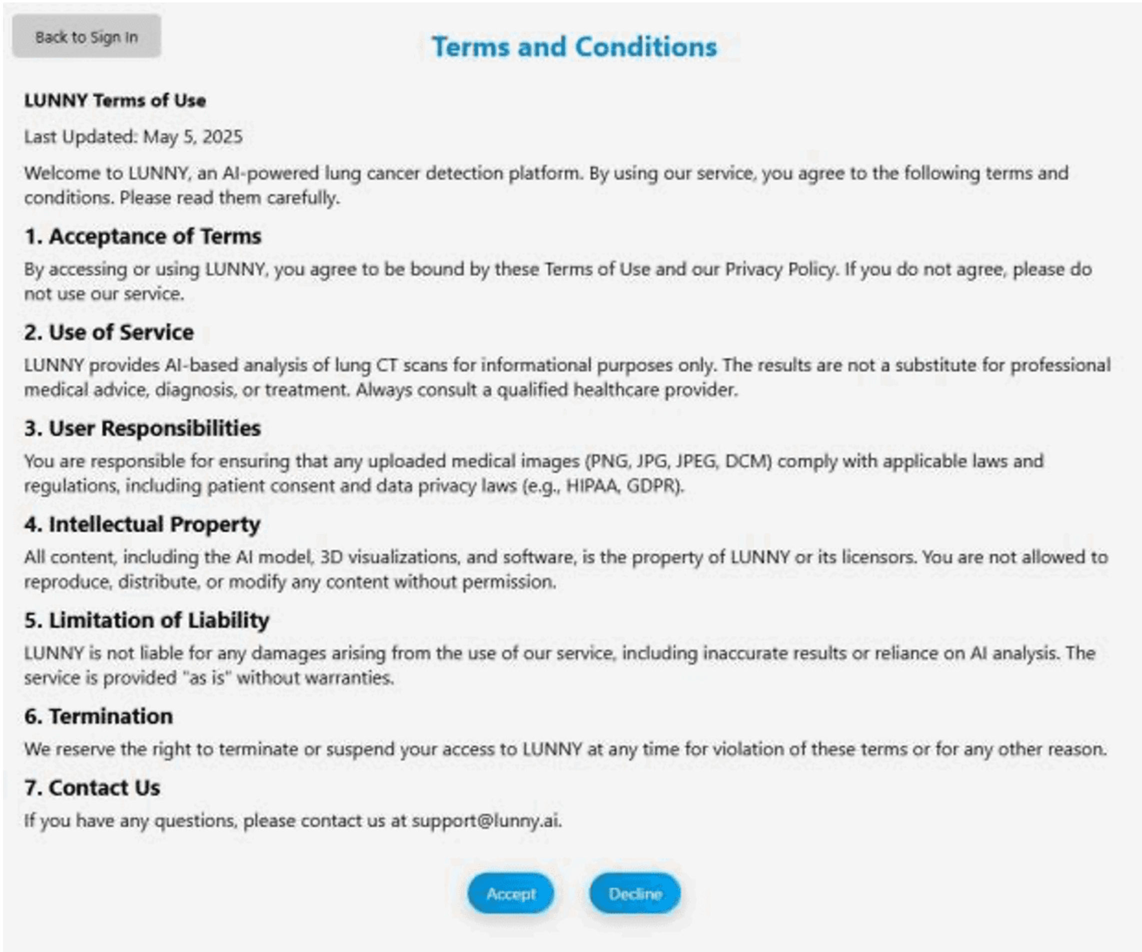

Figure 3. *LUNNY: Screenshot of the Terms and Conditions Interface.*

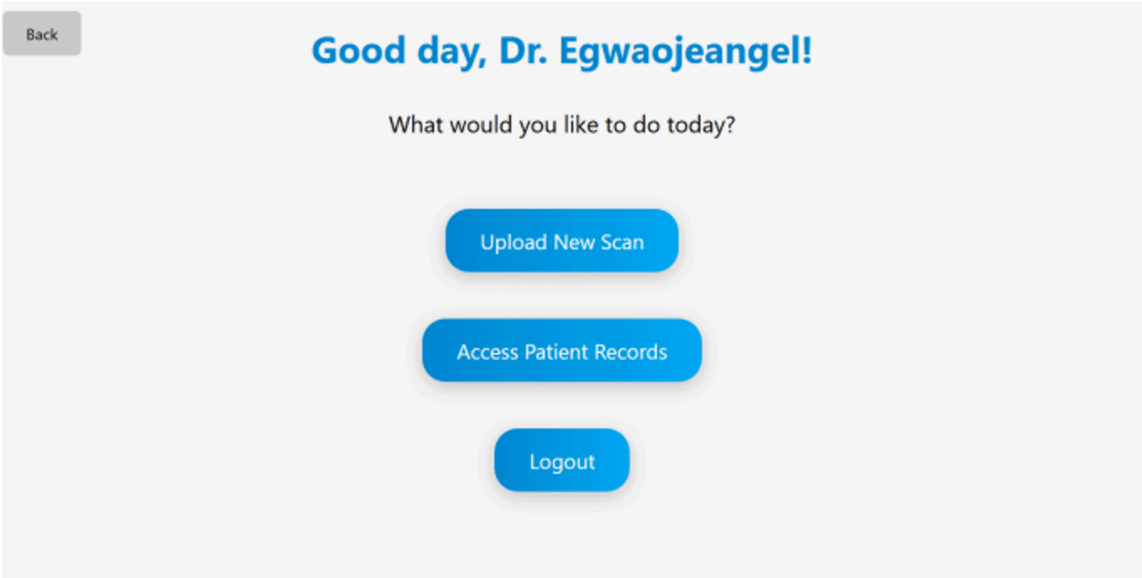

Figure 4. *LUNNY: Screenshot of the Post Login Interface.*

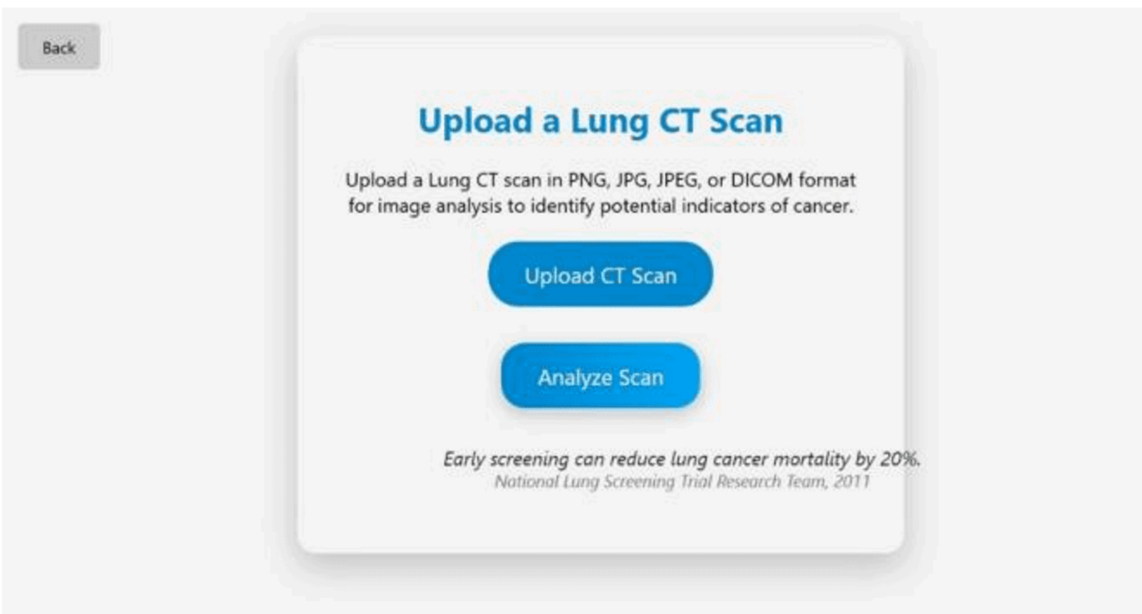

Figure 5. *LUNNY: Screenshot of the Upload Scan Page Interface.*

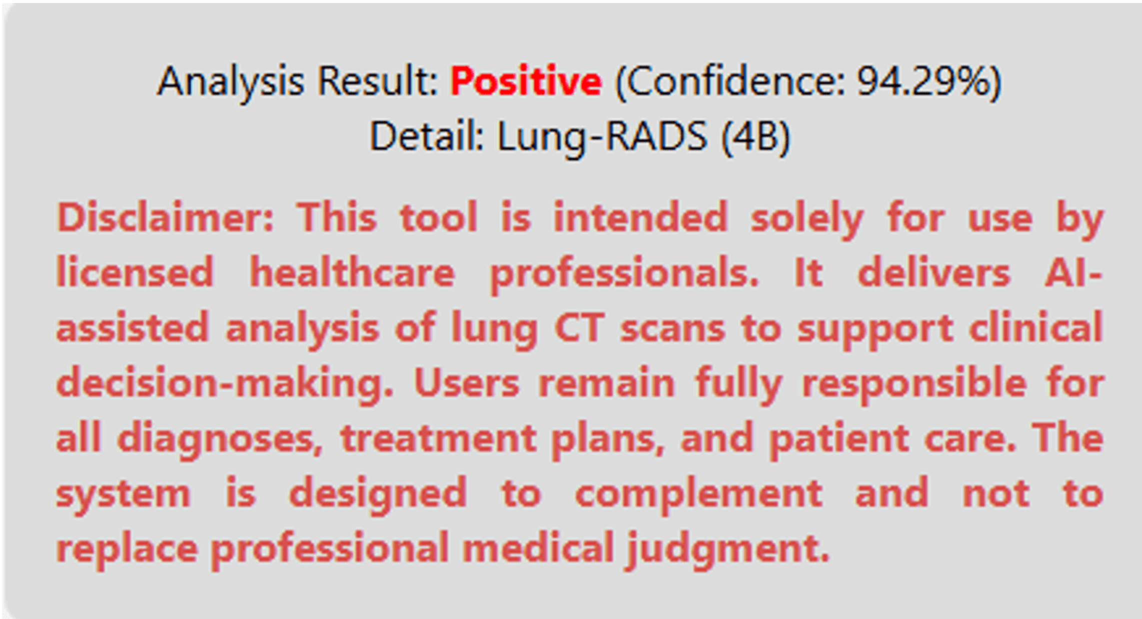

Figure 6. *LUNNY: Screenshot of the Analysis Result Page Interface.*

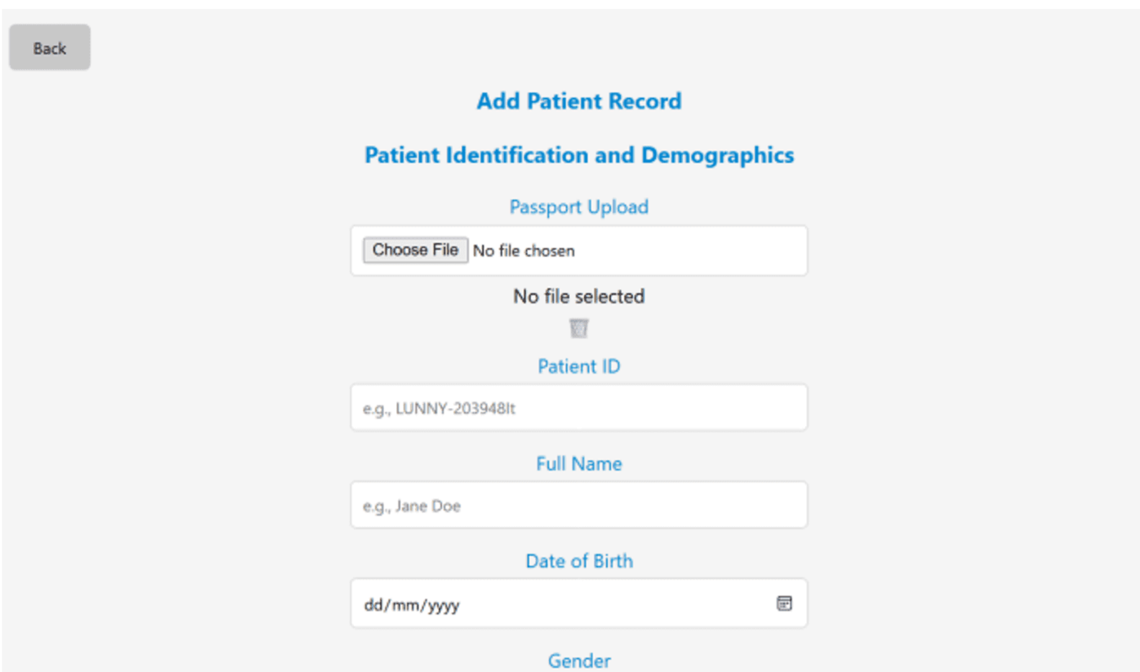

Figure 7. *LUNNY: Screenshot of the Input Records Page Interface.*

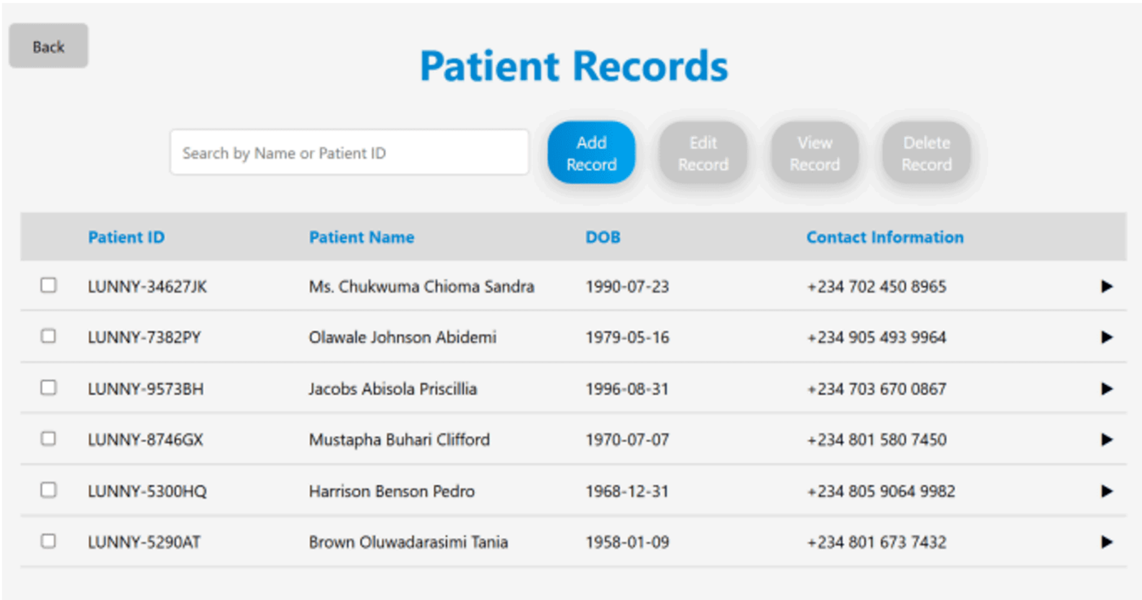

Figure 8. *LUNNY: Screenshot of the Patient Records Interface.*

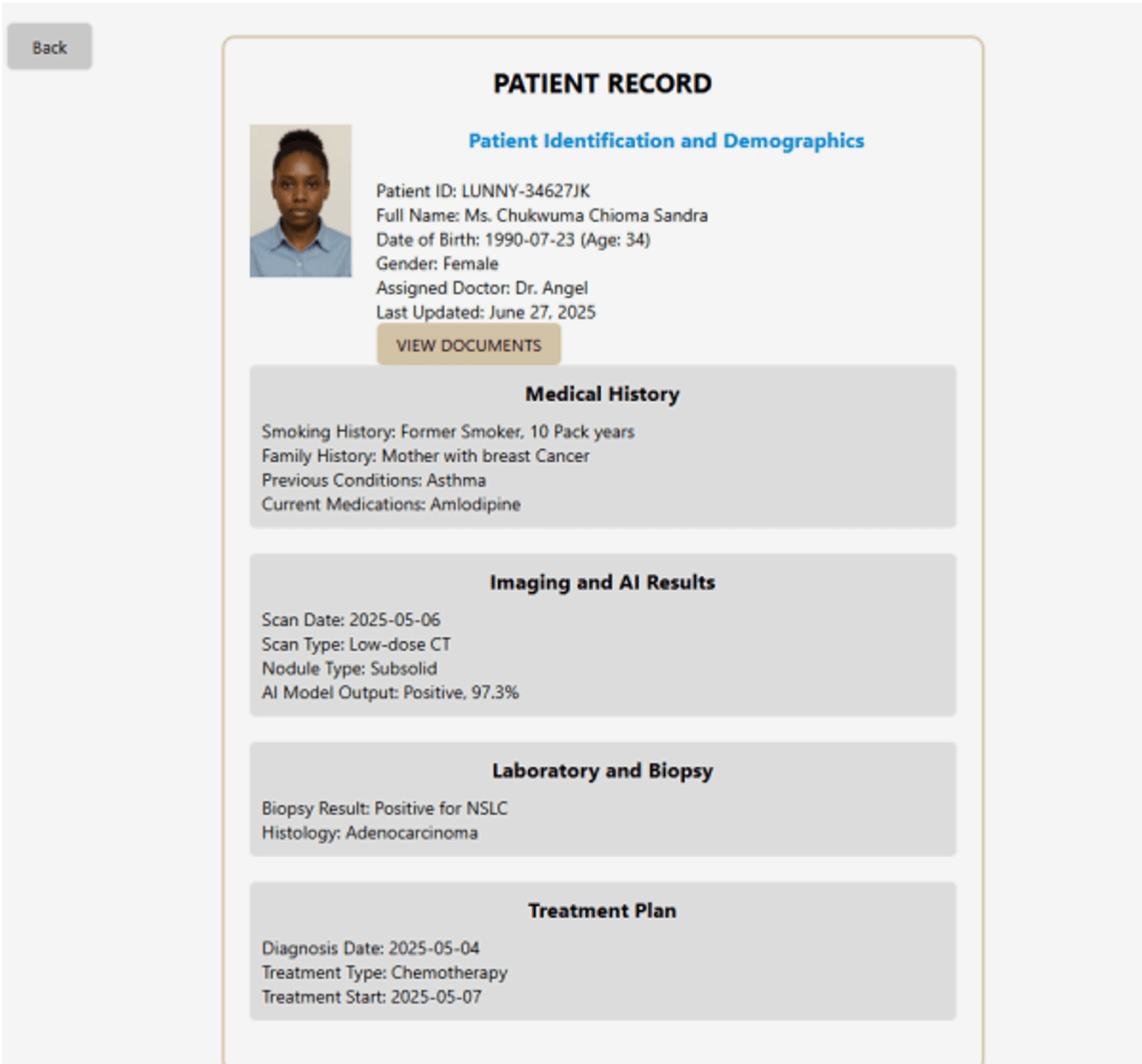

Figure 9. *LUNNY: Screenshot of the View Patient Record Interface.*

Discussion

The system achieved an accuracy of 98.46%, which shows the model has a strong ability to differentiate between cancerous and non-cancerous lung CT images. The recall value of 99.21% is important because it shows the model's effectiveness in correctly identifying the lungs that are cancerous, helping to reduce false negatives. This is important because, in clinical settings, missed diagnoses can have severe consequences.

The precision score of 97.69% shows a low false-positive rate by reducing the likelihood of unnecessary anxiety or follow-up procedures for patients. The high F1-score confirms the balance between precision

 doi.org/10.32388/4QGQJX

and recall, showing the strong effectiveness of the model in classifying between cancerous and non-cancerous cases.

The model performed well mainly because of the use of transfer learning with ResNet50, dataset balancing to reduce bias towards the majority class, and effective data augmentation strategies that made the model able to generalise better. Compared to several existing studies reported in the literature, the system used in this study shows competitive or superior performance while also offering practical deployment through a web application.

Conclusion

As reported in the literature, lung cancer has continually remained the leading cause of deaths caused by cancer, and this is largely due to delays in diagnosing this disease, which has therefore made a major contribution to poor survival rates. This work presents a deep learning system for the early detection of lung cancer using lung CT scan images. The ResNet50 model showed strong performance with high accuracy, precision, and recall when classifying lungs as either cancerous or non-cancerous and was then integrated into a web application. The system offers a user-friendly interface that can support healthcare professionals like radiologists and oncologists in clinical decision-making by providing reliable classification of lung images and helping to reduce the time consumed when using traditional methods. The system's ability to reduce delays in diagnosis and reduce false positive outcomes makes it a tool that can be useful in the medical field. Future work would focus on expanding the dataset, improving the interpretability of the system through explainable AI, and testing the system in hospitals for diagnosing in real-time and improving patient care support.

Notes

Short title: Lung cancer detection using ResNet50 on CT scans

Statements and Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Potential Competing Interests

The authors declare no conflicts of interest related to this study.

Ethics

This study did not involve any human or animal subjects and, therefore, did not require ethical approval. All datasets used in this study were publicly available.

Acknowledgements

The authors wish to acknowledge the providers of the publicly available datasets used in this study: the IQ-OTH/NCCD, LIDC-IDRI, and BIR Lung datasets.

References

1. ^a, b^ *National Cancer Institute (2025). "Risk Factors for Cancer." National Cancer Institute. https://www.cancer.gov/about-cancer/causes-prevention/risk.*
2. ^ *Kubryń N et al. (2025). "PROTAC Technology as a New Tool for Modern Pharmacotherapy." Molecules.* ***30****(10):2123. doi:10.3390/molecules30102123.*
3. ^a, b^ *Huang D, Li Z, Jiang T, Yang C, Li N (2024). "Artificial Intelligence in Lung Cancer: Current Applications, Future Perspectives, and Challenges." Front Oncol.* ***14****:1486310.*
4. ^ *World Health Organization (2024). "Cancer: Lung Cancer Overview." World Health Organization. https://www.who.int.*
5. ^ *National Cancer Institute (2025). "Non-Small Cell Lung Cancer Treatment (PDQ®) – Patient Version." National Cancer Institute. https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq.*
6. ^ *Sharma R, Khubchandani J (2024). "Global, Regional, and National Burden of Tracheal, Bronchus, and Lung Cancer in 2022: Evidence from the GLOBOCAN Study." Epidemiologia.* ***5****(4):785–795.*
7. ^ *Leiter A, Veluswamy RR, Wisnivesky JP (2023). "The Global Burden of Lung Cancer: Current Status and Future Trends." Nat Rev Clin Oncol.* ***20****(9):624–639.*
8. ^ *Thanoon MA, Zulkifley MA, Mohd Zainuri MA, Abdani SR (2022). "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images." Diagnostics.* ***13****(16):2617.*
9. ^ *Esteva A et al. (2021). "Deep Learning-Enabled Medical Computer Vision." NPJ Digit Med.* ***4****(1):5.*

 doi.org/10.32388/4QGQJX

10. ^Lancaster HL, Heuvelmans MA, Oudkerk M (2022). "Low-Dose Computed Tomography Lung Cancer Screening: Clinical Evidence and Implementation Research." J Intern Med. **292**(1):68–80.

11. a, bAbdulqader AF et al. (2025). "Multi-Objective Deep Learning for Lung Cancer Detection in CT Images: Enhancements in Tumor Classification, Localization, and Diagnostic Efficiency." Discover Oncology. **16**(1):529.

12. a, bMohamed ATI, Hassanien AE, Ghoneim A, Amer M, El-Fishawy N (2023). "Automatic Detection and Classification of Lung Cancer CT Scans Based on Deep Learning and Ebola Optimization Search Algorithm." PLoS One. **18**(8):e0285796.

13. a, bOguntimilehin A et al. (2024). "A Multilayer Perceptron-Based Mobile Diagnosis System for Malaria Fever." In: Proc 2024 Int Conf Sci, Eng Bus Driving Sustain Dev Goals (SEB4SDG). IEEE. pp. 1–7.

14. ^Mao L et al. (2024). "Knowledge-Informed Machine Learning for Cancer Diagnosis and Prognosis: A Review." arXiv preprint arXiv:2401.06406.

15. a, b, cGautam N, Basu A, Sarkar R (2024). "Lung Cancer Detection from Thoracic CT Scans Using an Ensemble of Deep Learning Models." Neural Comput Appl. **36**(5):2459–2477.

16. a, bShah AA, Ali F, Khan S, Rahman M, Jan Z, Kim D (2023). "Deep Learning Ensemble 2D CNN Approach Towards the Detection of Lung Cancer." Sci Rep. **13**:29856.

17. ^Jiang X, Hu Z, Wang S, Zhang Y (2022). "Deep Learning for Medical Image-Based Cancer Diagnosis." Cancers. **15**(14):3608.

18. a, bRaza R et al. (2023). "Lung-EffNet: Lung Cancer Classification Using EfficientNet from CT-Scan Images." Eng Appl Artif Intell. **126**:106902.

19. a, bBray F et al. (2024). "Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." CA Cancer J Clin. doi:10.3322/caac.21834.

20. ^Belachew SA, Bizuayehu HM, Diaz A (2025). "Facilitators and Barriers of Lung Cancer Screening Participation: Umbrella and Systematic Review of the Global Evidence." BMC Public Health. **25**:2993. doi:10.1186/s12889-025-23808-8.

21. a, bZhang C et al. (2024). "Enhancing Lung Cancer Diagnosis with Data Fusion and Mobile Edge Computing Using DenseNet and CNN." J Cloud Comput. **13**(1):1–10.

22. ^Ansari S, Akbar N, Javed H, Alam T, Nawaz F (2024). "Multi-Model Analysis for Lung CT Scan Classification Using Deep Learning." Comput Biol Med. **157**:106835.

23. ^Abid MMN, Zia T, Ghafoor M, Windridge D (2021). "Multi-View Convolutional Recurrent Neural Networks for Lung Cancer Nodule Identification." Neurocomputing. **453**:299–311.

24. ^Dadgar S, Neshat M (2023). "Comparative Hybrid Deep Convolutional Learning Framework with Transfer Learning for Diagnosis of Lung Cancer." In: Proc 14th Int Conf Soft Comput Pattern Recognit (SoCPaR 2022). Vol. 648. Springer. pp. 296–305.
25. ^Hendrix W et al. (2023). "Deep Learning for the Detection of Benign and Malignant Pulmonary Nodules in Non-Screening Chest CT Scans." Commun Med. **3**(1):1–12.
26. ^Armato SG III et al. (2011). "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans." Med Phys. **38**(2):915–931. doi:10.1118/1.3528204.
27. ^Prashanthi B, Claret SPA (2024). "Lung Nodule Detection for CT-Guided Biopsy Images Using Deep Learning." J Appl Eng Technol Sci. **5**(2):909–924. doi:10.37385/jaets.v5i2.3716.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.