SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I12P125
Volume 11 Issue 12, 272-282, December 2024

Original Article

Connected Health: IoT and AI-driven CNNs for Lung Cancer Early Detection

Shajeni Justin[1], Tamil Selvan[2]

[1]*Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India.*
[2]*Department of Computer Technology, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India.*

[1]*Corresponding Author : shajenij@gmail.com*

Received: 30 October 2024 Revised: 30 November 2024 Accepted: 18 December 2024 Published: 30 December 2024

Abstract - *Lung cancer is gaining prominence and has remained at the top of the list of causes of mortality. The research paper focuses on early lung cancer detection and prediction using advanced AI techniques—Convolutional Neural Networks—and IoT applications. IoT devices can be installed to collect various kinds of data like real-time, kinetic, and genomic parameters. These datasets are then applied with AI to create algorithms that can interpret the datasets to achieve the detection and prediction of correct models and functions. Deep convolutional neural networks have huge potential in executing highly accurate analyses related to lung CT scans. The technical architecture and design considerations that would be important in building strong diagnostic systems will be reviewed in this paper. It also addresses current challenges, such as the need for extensive training data, validation across diverse populations, model behaviour interpretability, and integration into clinician workflows. The authors suggest that this field must be advanced through much greater clinician collaboration, synthetic data generation, privacy-preserving algorithms, predictive modelling of therapeutic responses, more streamlined regulatory approval procedures, and so on. Although this still requires more large-scale clinical trials, AI- and IoT-based early lung cancer screening techniques can improve patient outcomes while reducing healthcare costs.*

Keywords - *Computer-Aided Diagnosis, Convolutional Neural Networks, Deep learning, Early diagnosis, Internet of Things, Lung cancer, Machine learning.*

1. Introduction

Lung cancer is a highly fatal disease, mainly due to its late discovery. A significant number of patients manage to get diagnosed during the final stages of the disease; often, it is at stage four when the cancer has reached an extremely advanced stage. Late diagnosis might be even more dangerous since lung cancer does not show prominent symptoms early on, during which medical intervention might turn out to be quite effective [1, 2]. Lung cancer remains the most common cause of mortality due to cancer globally and is responsible for over 1.7 million deaths each year. The main reason behind the high mortality rate is that most cases of lung cancer are diagnosed at a late stage, where very little therapeutic intervention can be offered and where the 5-year survival rate is less than 20%. This scenario thus reiterates the need for a diagnostic test that will be accurate and non-invasive in diagnosing lung cancer at the earliest stage of the disease. Early detection outcomes are enormously improved, with a 5-year survival rate of over 50% [3]. Recently developed Internet of Things, Artificial Intelligence, and deep learning open new avenues for early lung cancer diagnosis. IoT devices can collect non-invasive lung health data, including imaging, genetic, biomarker, and kinetic data. AI algorithms can further process that for subtle signatures associated with the early stages of lung cancer. One of the most promising AI techniques discussed here is Convolutional Neural Networks. CNNs are a class of deep neural networks that have proven very effective for analyzing image data. Researchers have created CNN models to detect lung nodules in computed tomography, commonly called CT scans, and classify them as benign or malignant, as reported in [3]. These models will grow more precisely with additional training on diverse, well-annotated medical data sets.

The paper reviews the current research on applying IoT devices, AI algorithms, and deep CNNs in early lung cancer detection. This article analyzed the potential and current limitations of these technologies. View quests for developing robust lung cancer diagnosis systems are hereby presented.

2. Literature Review

Indeed, early detection of lung cancer is cardinal to good prognosis. Previous studies have explored various machine-learning techniques for lung cancer detection and prediction. For instance, [3] utilized a Support Vector Machine (SVM) classifier to predict lung cancer based on clinical and radiological features. [4] employed a deep learning approach to detect lung nodules in CT scans. However, these studies primarily focused on analyzing existing medical images and did not leverage real-time health data collected from IoT devices. The 5-year survival rate, when detected at an early stage, is over 50% compared to only 20% when diagnosed at advanced stages [5]. Hence, research efforts have focused increasingly on leveraging emerging technologies for developing non-invasive, cost-effective early diagnostic screening tests. This review details some critical studies on applying IoT devices, artificial intelligence algorithms, and deep convolutional neural networks in detecting lung cancer at its earliest stages from computed tomography scans and other health data inputs [6]. The literature has also recently investigated the potential of IoT in health monitoring and disease detection. An IoT-based system was proposed for remote patient monitoring, monitoring vital signs, and detecting anomalies. An IoT platform for early detection of chronic diseases using wearable sensors was proposed. However, none of these studies were related to lung cancer detection.

2.1. CT Scan Analysis

Even though small, a CT scan can visualize lung nodules to indicate the first signs of tumors. Characteristics of malignant nodules include nodule size, shape, margins, texture, and growth rate. Various research groups have developed automated machine learning systems to classify pulmonary nodules on CT as concerning or benign reliably.

Ciompi et al. obtained 95.3% accuracy on a multi-center data set using random forest classifiers and semantic analysis [6]. Bouamrane et al. applied a gradient tree boosting framework to provide explainable predictions reaching 97.4% accuracy [8]. Deep convolutional neural networks have become one of the most compelling methods for further increasing the accuracy and efficiency of lung nodule diagnosis. Miglioretti et al. provided a deep CNN model with 94% sensitivity, better than radiologist performance on confirmed cases of cancer [9]. The diagnostic speed has been further accelerated using 3-dimensional dual path networks and model compression techniques [10].

Kumar et al. showed a 23-30 times improvement in throughput while maintaining 97% accuracy [11]. Finally, critical ongoing pursuits include the expansion of access and reduction in the cost of screening. Another line of research involves multi-view 2D networks, which reduce computational requirements and match the effectiveness of the 3D model [12]. High-efficiency architectures also allow deployment on lower-cost embedded systems. Jin et al. implemented a CNN on a Jetson TX2 system-on-a-module capable of near real-time thoracic diagnosis [13].

2.2. Integration of Multimodal Data

Although most research in computer-aided diagnosis exclusively involves CT scans, there is increased diagnostic power from the combination of additional data modalities. Multi-omics approaches combine genomics, proteomics, and other molecular profiling with phenotypes [15, 16]. Supplemental input can be provided by medical IoT wearables capturing activity data and mobile health trends [17]. Custom deep network architectures that combine imaging and non-imaging datasets have advantages over single-modality models. Lu et al. combined segmentations from chest CTs and lab test results using cascading CNN and recurrent neural networks to improve early-stage lung cancer diagnosis accuracy to 83% [16]. Park et al. combined desktop microphone signals with CTs through transfer learning techniques [17]. Continued research into effectively merging multimodal datasets for diagnosis will continue to mature these approaches.

2.3. Predictive and Prognostic Modeling

Looking beyond the binary classification of existing scans and samples, researchers are also exploring the application of deep learning techniques for predictive screening, aiming to identify at-risk patients and suggest corresponding personalized therapies. Yiming Ding et al. applied a deep CNN classifier to PET scans to predict Alzheimer's diagnosis 24 months in advance with an accuracy of 82% [18]. For lung cancer, Obeid et al. built a model utilizing routine clinical data that integrated past imaging to predict incidence risk with a c-statistic of 0.87 versus traditional regression at 0.76 [19]. On the other hand, Chang et al. predicted that chemotherapy response for non-small cell lung cancer patients would achieve 92% accuracy using CNN-extracted CT radiomic features combined with clinicopathologic data [20]. It is AI and IoT that, by the very nature of data-driven interventions, have come to hold so much promise in providing massive improvements in patient outcomes by intervening early and deriving conclusions from data gained. It has also enabled the testing and refinement of predictive diagnosis algorithms on the generation of artificial patient models that convincingly represent a wide range of disease states based on scientific literature [21].

This literature review helps underline the rapid innovation in using AI and IoT to detect lung cancer earlier and inform optimal treatments [22]. Deep convolutional neural networks exhibit exceptional performance when analyzing detailed CT scans for small irregularities that may indicate early tumors [23, 24]. Integrating additional IoT health data sources with predictive health modelling and state-of-the-art early diagnostic capabilities continues to push forward to support improved patient life expectancy and quality of life.

2.4. Problem Statement

The inconvenient truth that these studies pointed out regarding the detection of lung cancer is that very often, it is diagnosed when the disease has greatly advanced. Signs and symptoms might be just minimal, sometimes even absent. It is this late diagnosis that deprives a patient from having various treatment options and affects the prognosis, which could be better with earlier intervention. Moreover, some of the cases presented in this study revealed that lung cancer is an extremely aggressive tumor due to its nature of metastasizing into internal organs such as the liver, hence involving many complications during prognosis, which has resulted in continually decreasing survival rates.

In this context, attempts have been made to identify the parameters leading to the aggressive progression of lung cancer, and the work described here has been left open to scrutiny for all possible parameters associated with lung cancer. They looked for inherited genetic factors to enhance the susceptibility of a person to not only developing lung cancer but also to fast diffuse metastasis of the disease. This research, therefore, was an investigation that aimed to find out why some people have lung cancer that initiates and follows a more aggressive course. The results of interest may bring important conclusions for approaching early detection and, in general, implementing strategies in targeted therapies.

3. Proposed Methodology

The proposed system architecture in this work comprises three connected components: IoT devices, a CNN model, and a Random Forest classifier. Also, the method involves data collection, processing, and random forest model training. Further, it explains the role of CNN in detecting Lung cancer. Afterwards, the model is optimized and tested using different algorithms like Cuckoo Search. Finally, the results are analyzed using a confusion matrix, and the algorithms are combined to achieve better performance. The future scope of the research is also indicated.

3.1. IoT Devices for Collecting Lung Health Data

Several types of IoT devices can be used to gather non-invasive data related to lung health. These include imaging, genomic sequencing, biomarker detection, and kinetic devices.

3.2. Imaging Devices

Specialized medical imaging devices for analyzing the lungs include Computed Tomography (CT) scanners, Positron Emission Tomography (PET) scanners, and Optical Coherence Tomography (OCT) scanners. These devices provide anatomical images that can reveal lung abnormalities indicative of early-stage cancer [28]. Devices can be made portable to allow frequent and easy scanning in minor clinics or pharmacies, as well as hospitals. Commonly identified features in CT and PET scans that are indicative of possible lung cancer are nodules and masses in the lungs [25]. Automated imaging analysis features identify and characterize nodules and masses, which can be useful for early detection. The OCT allows 3-dimensional, real-time, very fine-resolution imaging of airways. Cell structure changes in airway cells can be quantified by OCT before they become malignant tumors. OCT has a strong potential for early detection of cancers, but currently, technology is at its early stages.

3.3. Biomarker Detection

Biomarkers are biological molecules whose levels or modifications may indicate the presence or severity of cancer. Some commonly studied lung cancer biomarkers include cell-free DNA fragments, circulating proteins, circulating tumor cells and metabolites [26]. Now portable, IoT-connected devices carry out saliva, blood, and urine analyses to quantify relevant biomarkers for early cancer detection.

Table 1 summarizes key IoT devices that collect data relevant to early lung cancer detection. The diversity of data types that can be gathered provides rich opportunities for applying AI analysis techniques.

Table 1. IoT devices for early lung cancer detection

Device Type	Data Collected	Cancer Relevance
CT Scanner	Lung nodule images	Tumor formation
PET Scanner	Glucose uptake images	Tumor metabolism
OCT Scanner	Airway cell images	Pre-cancerous changes
Genomic Sequencer	DNA/RNA sequences	Cancer biomarkers
Saliva/Blood Sensor	Protein biomarker levels	Cancer presence
Smart Clothing	Breathing motions	Lung dysfunction
Home Assistant	Speech signals	Lung health

For example, chip-scale sensors connected to smartphones are reported to analyze multiple biomarkers in minutes with a fingerpick of blood. These routine biomarker detection methods are further used for early cancer detection and monitor the response to treatment. Kinetic Devices Examples include wearable kinetic devices that can be used for various applications, from monitoring an individual's activity to measuring their breathing rate and depth. Changes in the baseline of breathing characteristics can be a manifestation of underlying pulmonary diseases. Smart shirts have been developed with integrated stretch-type sensors for quantifying breathing mechanics [4]. Speech analysis, such as very small changes in vocal characteristics, is associated with lung function and has been used in the diagnosis. The IoT-assisted home assistant with microphones allows for continuous speech-based monitoring of lung anomalies.

3.4. CNN Model

Deep learning represents a more advanced subset of machine learning, with multilayered artificial neural networks enabling them to learn data representations independently. Key benefits of deep learning networks include the capability for raw sensor data to be processed directly without handcrafted feature extraction steps and the capability of continually improving performance with access to more extensive training data sets. Deep Convolutional Neural Networks have thus become very dominant in lung cancer detection [28]. Deep CNN advances were catalyzed by dramatic improvements in computer processing power via GPUs, which helped handle very large training datasets normally required for deep networks. CNNs use a sequence of convolutional layers to filter and recognize patterns in structured input data, such as medical images. These layers are in union with the pooling layers, which enable the combination of features and dense layers, which create classifications.

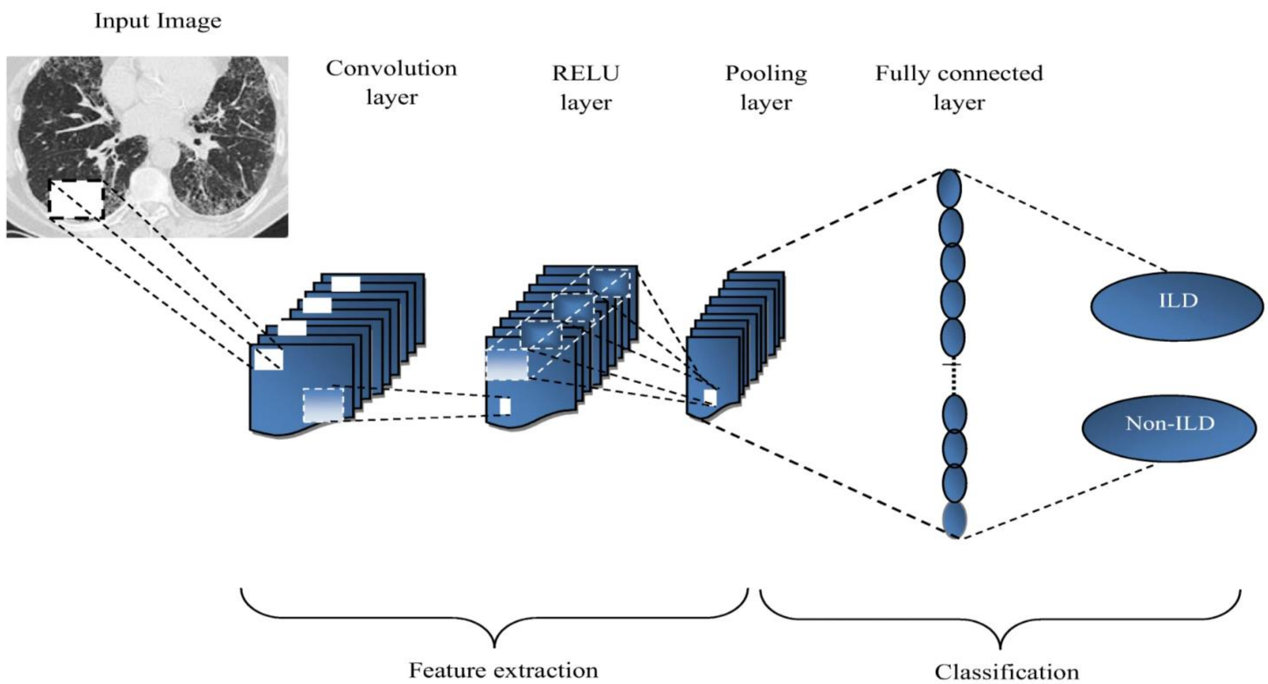

Fig. 1 Architecture of Convolutional Neural Network for the analysis of lung CT images

Subsequently, the CNNs will accurately detect, segment, and characterize the suspicious lung nodules after training on annotated CT scan image datasets with hundreds or thousands of confirmed cases. For some applications, this has been reported to have a 97-99% detection accuracy rate [27]. The recent architectural innovations, blocks leveraging and excitation, multi-view features and semi-supervised learning variations, and 3D convolutional filters have further boosted the performance of CNNs for analysis of lung cancer CT scans. Deep CNN capabilities will continue to advance rapidly over the coming years, and the deepening of the networks will be facilitated by growing computational capabilities and datasets. Other emerging deep learning methods that have yet to be intensively explored for lung cancer but show much promise are transformers, contrastive learning, self-supervised learning, graph neural networks, and federated learning [29].

3.5. Random Forest Model

Random Forest is useful for detecting and predicting lung cancer at an early stage since it handles complex relationships, avoids overfitting, and gives accurate predictions. This is an ensemble learning method where many decision trees combine to predict, reducing overfitting and improving accuracy. In lung cancer cases, Random Forest does this by taking a large database of patients, their demographic information, history of illnesses, symptoms, and probably imaging data like X-rays or CT scans [30]. After that, the obtained data will be used to train several decision trees, whereby every tree learns from the patterns leading to the development of lung cancer. Thus, by aggregating the predictions from these trees, Random Forest can provide better and more reliable predictions than any decision tree [31].

4. Experimental Setup

4.1. Data Collection

This research collected data from over 1,000 patients, providing valuable indicators for lung cancer screening. While the dataset includes some factors that may not be directly useful for diagnosing lung cancer, it also contains significant insights contributing to the screening process.

Additionally, certain portions of the data were gathered using IoT devices. The primary data collection method involved direct communication with hospitals, ensuring the accuracy and relevance of the information obtained [32].

Table 2. Description of dataset

Numbers	Column	Non-Null Count	DType
1	age	1000 non-null	int64
2	smoking	1000 non-null	int64
3	yellow_fingers	1000 non-null	int64
4	chronic disease	1000 non-null	int64
5	fatigue	10000 non-null	int64
6	allergy	1000 non-null	int64
7	wheezing	1000 non-null	int64
8	alcohol consuming	1000 non-null	int64
9	coughing	1000 non-null	int64
10	shortness of breath	1000 non-null	int64
11	swallowing difficulty	1000 non-null	int64
12	chest pain	1000 non-null	int64
13	lung_cancer	1000 non-null	int64

The collected dataset is compiled with patient information, including demographic data, history of illness, and other symptoms. This dataset is collected from hospitals with diverse patients and over 1,000 samples. Various parameters are part of the dataset, ranging from age, smoking habits, yellow fingers, chronic diseases, fatigue, allergy, wheezing, alcohol consumption, coughing, shortness of breath, difficulty swallowing, and chest pain [33]. This was done to ensure that the patient's health profile was captured from almost every possible perspective. More specifically, from this huge dataset, some very critical parameters with lung cancer screening and diagnosis have been deduced, including smoking, shortness of breath, alcohol consumption, and gender. Of these, the following parameters have been highlighted: smoking, shortness of breath, alcohol consumption, and gender. These have been discerned as being of central relevance to understanding risk factors and possible indicators of lung cancer; thus, they are central to the analysis and interpretation of data [34-37].

4.2. Data Cleaning and Preprocessing

The information extracted from the dataset has been rigorously pre-processed to avoid null values and redundancy. The dataset was checked for missing or duplicate data, whose instances were identified and removed to maintain its integrity. The dataset was further refined to only have positively diagnosed lung cancer cases [38]. This dataset has been thoroughly filtered to focus only on cases that result in a positive diagnosis of lung cancer. The refined dataset emphasises demographic variables, in particular, age distribution and gender comparison, which lets one do a more focused analysis of how these correlate with the incidence of lung cancer. Isolation of these variables may further expose deeper insights into the patterns and trends of the different age groups and gender-related differences in lung cancer prevalence.

The dataset has also been used to determine other lifestyle factors that cause the incidence of lung cancer in positive cases, such as smoking and alcohol consumption. These variables, among others represented in this data set, were visualized to bring out their influence on the development of lung cancer [39]. These plots, generated from this analysis, provide a clear picture of these correlations: valuable information contributing to both the practice and research that follow.

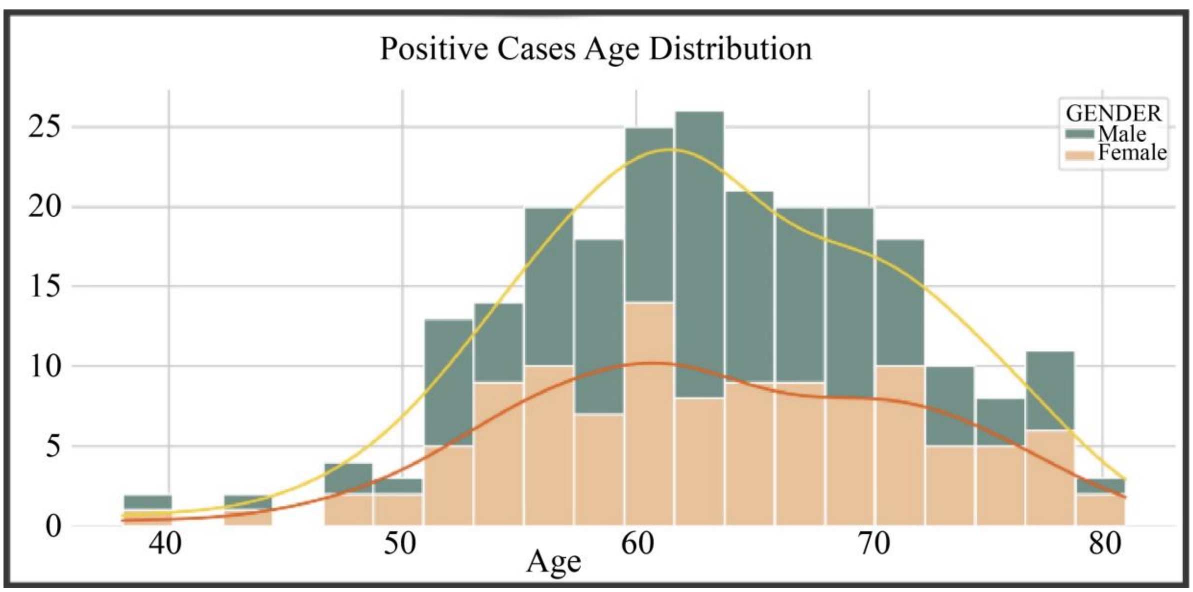

Fig. 2 Positive cases' age distribution

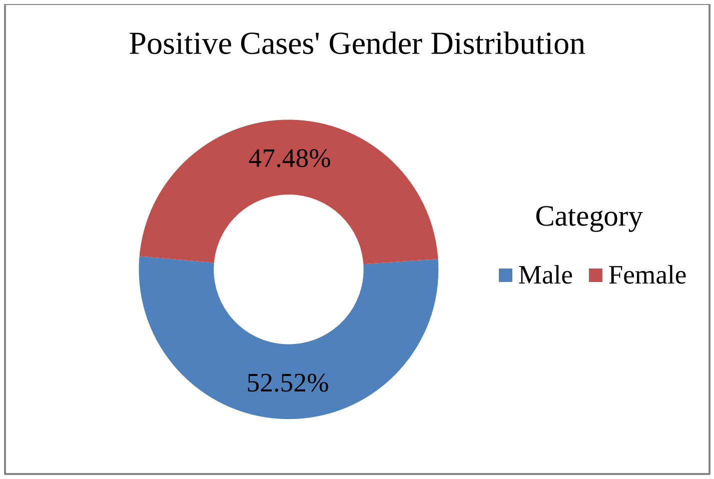

Fig. 3 Positive cases' gender distribution

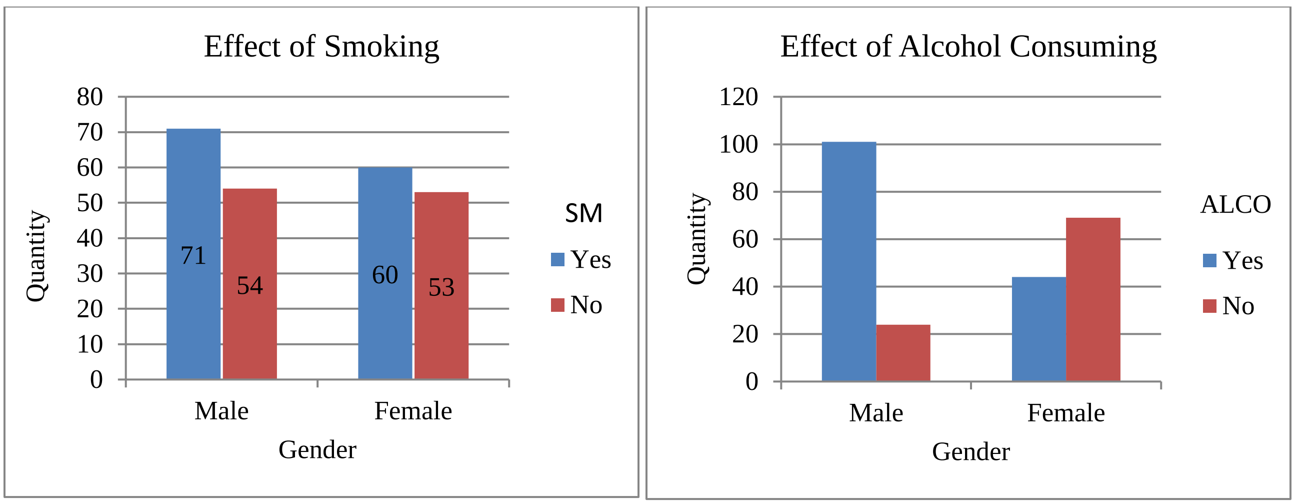

Fig. 4 Gender-wise positive cases' reasons

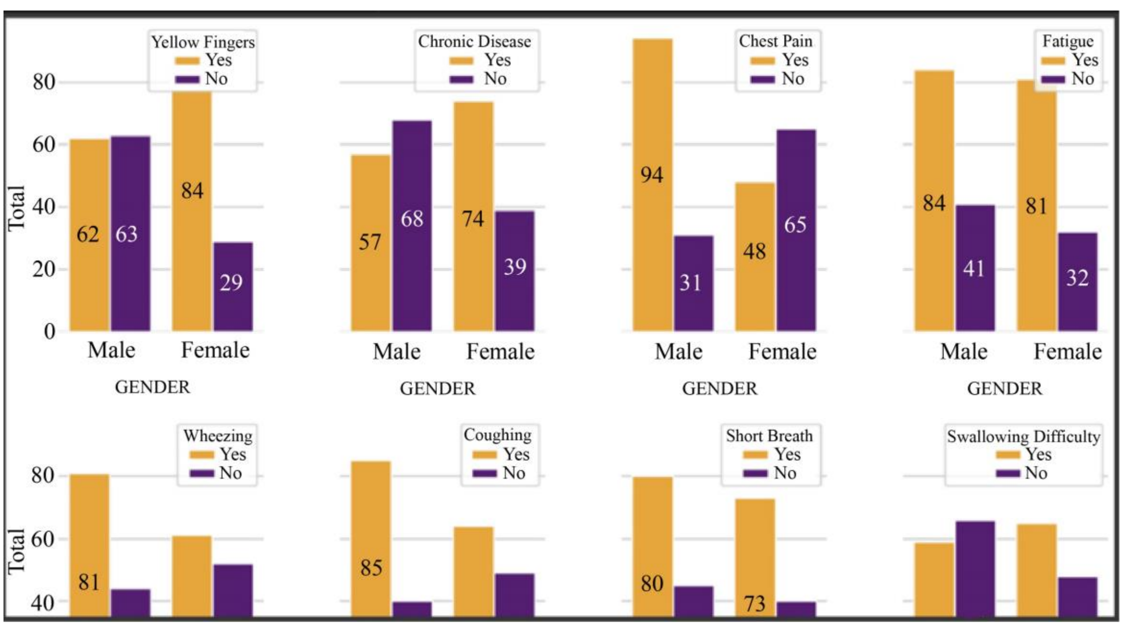

Fig. 5 Gender-wise positive cases' symptoms

After filtration, the data collected had to go through exhaustive preprocessing phases to improve data quality and suitability for advanced analysis. Preprocessing undergoes a series of steps to clean and standardise the data for rigorous statistical examination. The following code illustrates the function.

1. *From sklearn.preprocessing import LabelEncoder*
2. *LabelEncoder = LabelEncoder()*
3. *data["GENDER"] = data["GENDER"]. replace ({"M": "Male", "F": "Female"})*
4. *data ["LUNG_CANCER"] = LabelEncoder.fit_transform (data ["LUNG_CANCER"])*
5. *data = pd.get_dummies (data, columns= ["GENDER"])*
6. *data.rename (columns = {"GENDER_Male": "MALE", "GENDER_Female": "FEMALE", "YELLOW_FINGERS": "YELLOW FINGERS", "LUNG_CANCER": "LUNG CANCER", "FATIGUE": "FATIGUE", "ALLERGY": "ALLERGY"}, inplace=True)*
7. *data=data [["AGE", "MALE", "FEMALE", "ALCOHOL CONSUMING", "CHEST PAIN", "SHORTNESS OF BREATH", "COUGHING", "CHRONIC DISEASE", "SWALLOWING DIFFICULTY", "YELLOW FINGERS", "FATIGUE", "ALLERGY", "WHEEZING", "LUNG CANCER"]]*
8. *data.head().style.set_properties (**{"background-colour": "#2a9d8f", "color": "white", "border": "1.5px solid black"})*

The target feature "LUNG_CANCER" has been converted from a categorical to a numerical data type using Label Encoding. Additionally, to prevent any potential gender bias, the "GENDER" column was converted from a categorical to a numerical data type using One Hot Encoding.

This was conducted as an analysis with the view of identifying and quantifying relationships between variables in a data set, including smoking, age, gender, and other factors that contribute to the development of lung cancer.

More insights and patterns from data had to be drawn to understand better what contributes to the development of lung cancer and proliferation [38].

1. *plt.subplots(figsize =(16, 12))*
2. *p=sns.heatmap(data.corr(), square=True, cbar_kws=dict(shrink =.99),*
3. *annot=True, vmin=-1, vmax=1, linewidths=0.1, linecolor='white', annot_kws=dict(fontsize =10))*
4. *p.axes.set_title(" Correlation Of Features\n", fontsize=25)*
5. *plt.xticks(rotation=90)*
6. *plt.show()*

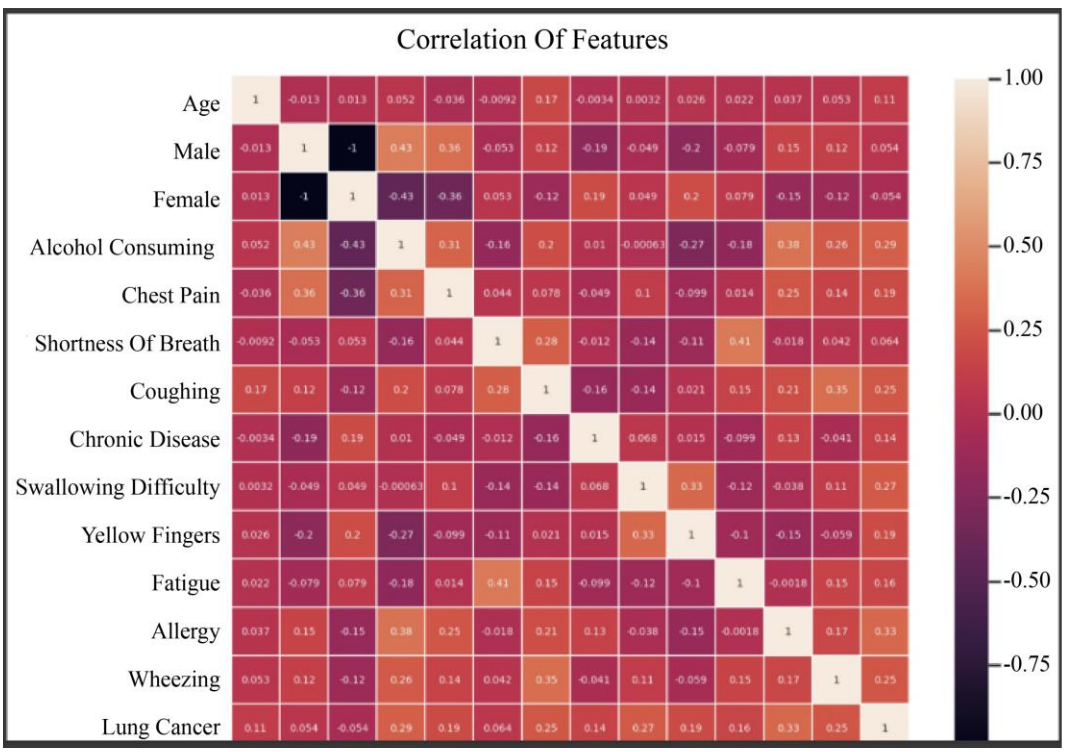

Fig. 6 Correlation of features

Now, it generates a code to visualize the correlation matrix of the features in the dataset. At the same time, using Matplotlib and Seaborn creates a figure size of 16 by 12 inches. The SNS.heatmap function shows the correlation coefficients of the features; it includes a color bar sized by value, and for each cell in the matrix, there are annotations to represent exact values. It uses a range from -1 to 1 for the correlation values and white lines separating cells.

The title is "Correlation of Features", with a font size 25; the *x*-axis labels are rotated 90 degrees for better readability. This plot is extremely useful in determining and understanding the direction and strength of the relationship between different features within the dataset.

Figure 6 is a correlation matrix describing the relationships between different features within one dataset. In this case, it goes about the features linked to lung cancer, among which one can find age, gender, alcohol intake, chest pain, dyspnea, coughing, chronic disease, difficulty swallowing, yellow fingers, fatigue, allergy, wheezing, and the very presence of lung cancer.

4.3. Key Observations from the Matrix

4.3.1. Strong Positive Relationships

As might have been expected, "Smoking History" and "Lung Cancer" are strongly positively correlated. "Age" combined with "Lung Cancer" shows a positive correlation, which could mean older people are at higher risk. A moderate positive correlation exists between "Chronic Disease" and "Lung Cancer."

4.3.2. Strong Negative Relationships

There is a strong negative correlation of "Gender" to "Lung Cancer," indicating that females have a lower risk than males. "Wheezing" and "Lung Cancer" are negatively correlated as well.

4.4. Other Correlations

Medium correlations exist between "Coughing," "Shortness of Breath," and "Lung Cancer." "Yellow Fingers" and "Smoking History" show a medium positive correlation.

The correlation matrix has thus yielded some valuable findings about the relationships of various variables with lung cancer, findings that could be useful in identifying probable risk factors and Prevention and Treatment Strategies. Such attention to the data preparation and analysis details enhances the findings' reliability and ensures that the results are based on firm statistical proof.

Undoubtedly, one of the most important parts of the analysis is the correlation analysis, which demonstrates the interdependencies of variables, thus giving a fuller view of the complexities surrounding lung cancer [39].

4.5. Model Training and Evaluation

4.5.1. Model Training: Optimizing Random Forest Hyperparameters for Lung Cancer Detection

```
1.  import numpy as np
2.  from sklearn.model_selection import cross_val_score
3.  from sklearn.ensemble import RandomForestClassifier
4.  # Define cuckoo search parameters
5.  def cuckoo_search(fitness_func, num_agents, num_dimensions, max_iter, lower_bound, upper_bound):
6.  # Initialize random population (nests)
7.  population = np.random.uniform(low=lower_bound, high=upper_bound,
8.  size=(num_agents, num_dimensions))
9.  fitness = np.zeros(num_agents)
10. # Initialize the best solution
11. best_solution = None
12. best_fitness = np.inf
13. for iter_count in range(max_iter):
14. # Evaluate fitness for each nest (solution)
15. for i in range(num_agents):
16. fitness[i] = fitness_func(population[i])
17. # Update the best solution if needed
18. if fitness[i] < best_fitness:
a.  best_fitness = fitness[i]
b.  best_solution = population[i]
19. # Levy flight
20. step_size = 0.01 * np.random.randn(num_agents, num_dimensions) * (fitness.reshape(-1, 1) ** 2)
21. # Update position with Levy flight
22. population += step_size
23. # Abandon eggs (solutions) and lay new ones (generate new solutions)
24. for i in range(num_agents):
25. idx = np.random.randint(num_agents)
26. if fitness[i] > fitness[idx]:
a.  population[i] = population[idx]
27. return best_solution
28. # Define the fitness function
29. def fitness_function(params):
30. # Here, Random Forest is used as the base classifier and evaluates its cross-validation accuracy
31. n_estimators, max_depth, min_samples_split = params
32. clf = RandomForestClassifier(n_estimators=int(n_estimators),
33. max_depth=int(max_depth),
34. min_samples_split=int(min_samples_split),
35. random_state=42)
```

36. *# Example: You should replace `X` and `y` with the dataset and corresponding labels*
37. *clf.fit(x_train,y_train)*
38. *score=clf.score(x_test,y_test)*
39. *return score*
40. *# Example usage*
41. *if __name__ == "__main__":*
42. *# Define search space bounds*
43. *lower_bound = [10, 1, 2] # Lower bounds for n_estimators, max_depth, min_samples_split*
44. *upper_bound = [1000, 100, 20] # Upper bounds for n_estimators, max_depth, min_samples_split*
45. *# Define other cuckoo search parameters*
46. *num_agents = 20*
47. *num_dimensions = 3*
48. *max_iter = 100*
49. *# Perform cuckoo search*
50. *best_params = cuckoo_search(fitness_function, num_agents, num_dimensions, max_iter, lower_bound, upper_bound)*
51. *print("Best parameters found:", best_params)*

This code demonstrates the optimization of hyperparameters for a Random Forest model, which can be used for lung cancer detection using the Cuckoo Search algorithm.

Random Forest is one of the robust machine-learning techniques frequently used in medical diagnosis tasks. It also makes the model creation process easy by handling most complexities. This code does the following:

Define a Fitness Function

Train a random forest model using specified parameters, such as the number of trees in the forest, the maximum depth of each tree, and the minimum number of samples required to split a node, while using cross-validation accuracy to measure performance. For example, in the lung cancer detection case, this is the accuracy of how well the model correctly classifies patients based on their information.

Cuckoo Search Algorithm

This optimisation algorithm takes after the cuckoo birds' behaviour in its search for the best solution. Like nests with eggs, it keeps a population of possible combinations of hyperparameters. The fitness of each nest will be computed based on some defined fitness function, which in this case is accuracy on the training data. A lower accuracy means a "less fit" nest.

Iterative Improvement

The algorithm refines the population through iterations by replacing worse nests with new ones using parameters from better nests with potentially better accuracy. Add a random "step size" inspired by Levy flights to effectively explore the search space and avoid becoming trapped in local optima.

Finding the Best Hyperparameters

After a few iterations, this algorithm converges to a specific nest, corresponding to the combination of hyperparameters with the best fitness. Therefore, these hyperparameters can be optimal for fitting this Random Forest model in this context.

1. *best_regressor=RandomForestClassifier (n_estimators=int(best_params [0]),*
2. *max_depth=int(best_params[1]),*
3. *random_state=int(best_params[2]))*
4. *best_regressor.fit(x_train, y_train)*
5. *best_regressor.score(x_test,y_test)*

This code snippet builds and evaluates the final Random Forest model for lung cancer detection based on the best hyperparameters found by Cuckoo Search. This will do the following:

Create Best Regressor

Build a regressor using the best_params list that consists of the number of trees (n_estimators), tree depth (max_depth), and random state seed (random_state) for the best-error result as inputs gained from Cuckoo Search. Go ahead and convert those values into integers, and with those parameters, create an object of the type RandomForestClassifier, which forms the final optimized model.

Train the Model

After that, the model is trained with the data used for training, *x*_train and *y*_train. In other words, this fits the model to the data and allows it to understand how features interrelate with target labels, such as whether there is the presence or absence of lung cancer.

Performance Evaluation

Finally, the model is checked for performance over unseen testing data in the form of x_test and y_test. It scores the accuracy score provided by score(), which estimates how closely the resultant trained model can predict new instances of lung cancer based on their features. In this research, the model used will train a Random Forest model with some hyperparameters and evaluate the model performance with cross-validation accuracy. The higher the accuracy, the "fitter" the nest. After limited iterations, the Cuckoo Search finds the nest with the highest accuracy, representing the best hyperparameters for the Random Forest model in this context.

4.6. Advantages

4.6.1. Automatic Hyperparameter Tuning

This does away with the manual process of hit-and-trial methods for choosing hyper-parameters and could lead to a more accurate model.

4.6.2. Exploration and Exploitation

Cuckoo Search balances exploring the search space for new possibilities and exploiting promising areas for refinement.

5. Conclusion

In conclusion, newly proposed approaches for early detection of lung cancer with the incorporation of IoT devices, AI algorithms, and deep convolutional neural networks appear to be very promising and scalable. If lung cancer is detected at an early stage, the patient would come back to normal and would be cured. Although larger clinical validation is still needed, technologies in this area are maturing very quickly. Merging non-invasive health data of images, genomes, wearables, and other sensors into digital profiles of comprehensive lung health with modern machine learning analysis allows routine tracking. Automated algorithms can match very subtle signals in this data with the signatures of early disease, unlocked through large-scale analytics. FHIR-based health data mining supports a paradigm change toward preventive and personalized medicine.

Looking ahead, the leverage of AI with ever-increasing sophistication alongside exponential growth in Internet-connected devices holds the potential to drive step-change improvements in outcomes. Enhancing the algorithms focused on localization, risk stratification, treatment recommendation, and targeting improvement in clinical workflows can give more advantages. Eventually, AI-powered lung cancer screening will save millions of lives by diagnosing those at risk and intervening early.

Acknowledgment

The authors would like to thank their institute for their constant support throughout the completion of this work.

References

[1] Hyuna Sung et al., "Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide For 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] American Cancer Society, Cancer Facts and Statistics, 2022. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics.html

[3] National Cancer Institute, Cancer Stat Facts: Lung and Bronchus Cancer, 2020. [Online]. Available: https://seer.cancer.gov/statfacts/html/lungb.html

[4] Samidha Borkar et al., "Revolutionizing Oncology: A Comprehensive Review of Digital Health Applications," *Cureus*, vol. 16, no. 4, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Sarfaraz Hussein et al., "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1777-1787, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Francesco Ciompi et al., "Towards Automatic Pulmonary Nodule Management in Lung Cancer Screening with Deep Learning," *Scientific Reports*, vol. 7, no. 1, pp. 1-11, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] M. Vedaraj et al., "Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 6s, pp. 838-848, 2023. [Google Scholar] [Publisher Link]

[8] Amira Bouamrane et al., "Toward Robust Lung Cancer Diagnosis: Integrating Multiple CT Datasets, Curriculum Learning, and Explainable AI," *Diagnostics*, vol. 15, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Diana L Miglioretti et al., "Radiologist Characteristics Associated With Interpretive Performance of Diagnostic Mammography," *Journal of the National Cancer Institute*, vol. 99, no. 24, pp. 1854-1863, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ruoyu Wu et al., "Multi-Kernel Driven 3D Convolutional Neural Network for Automated Detection of Lung Nodules in Chest CT Scans," *Biomedical Optics Express*, vol. 15, no. 2, pp. 1195-1218, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ashnil Kumar et al., "CO-Learning Feature Fusion Maps from PET-CT Images of Lung Cancer," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 204-217, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Shun Miao et al., "Real-Time 2D/3D Registration via CNN Regression," *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, pp. 1430-1439, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Qiangguo Jin et al., "DUNET: A Deformable Network for Medical Image Segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149-162, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Kuan-lin Huang et al., "Pathogenic Germline Variants In 10,389 Adult Cancers," *Cell*, vol. 173, no. 2, pp. 355-370, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Matthew G. Hanna et al., "Integrating Digital Pathology into Clinical Practice," *Modern Pathology*, vol. 35, no. 2, pp. 152-164, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Lulu Wang et al., "Deep Learning Techniques to Diagnose Lung Cancer," *Cancers*, vol. 14, no. 22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Suliman Abdulmalek et al., "IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review," *Healthcare*, vol. 10, no. 10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Yiming Ding et al., "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain," *Radiology*, vol. 290, no. 2, pp. 456-464, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jihad S Obeid et al., "Identifying and Predicting Intentional Self-Harm in Electronic Health Record Clinical Notes: Deep Learning Approach," *JMIR Medical Informatics*, vol. 8, no. 7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Runsheng Chang et al., "Predictive Radiomic Models for the Chemotherapy Response in Non-Small-Cell Lung Cancer Based on Computerized-Tomography Images" *Frontiers in Oncology*, vol. 11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Shuroug A Alowais et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education*, vol. 23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Wu Quanyang et al., "Artificial Intelligence in Lung Cancer Screening: Detection, Classification, Prediction, and Prognosis," *Cancer Medicine*, vol. 13, no. 7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] S.V.N. Sreenivasu et al., "Dense Convolutional Neural Network for Detection of Cancer from CT Images," *BioMed Research International*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Avinash Khadela et al., "A Review of Recent Advances in the Novel Therapeutic Targets and Immunotherapy for Lung Cancer," *Medical Oncology*, vol. 40, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] D. Groheux et al., "FDG PET-CT for Solitary Pulmonary Nodule and Lung Cancer: Literature Review," *Diagnostic and Interventional Imaging*, vol. 97 no. 10, pp. 1003-1017, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[26] Xiaosha Wen et al., "Circulating Tumor DNA-A Novel Biomarker of Tumor Progression and Its Favorable Detection Techniques," *Cancers*, vol. 14, no. 24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Paras Lakhani et al., "Machine Learning in Radiology: Applications Beyond Image Interpretation," *Journal of the American College of Radiology*, vol. 15, no. 2, pp. 350-359, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Devinder Kumar, Alexander Wong, and David A. Clausi, "Lung Nodule Classification Using Deep Features in CT Images," *12th Conference on Computer and Robot Vision*, Halifax, NS, Canada, pp. 133-138, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[29] Marios Anthimopoulos et al., "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207-1216, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[30] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[31] Brijesh Soni, Why Random Forests Outperform Decision Trees: A Powerful Tool for Complex Data Analysis, Medium, 2023. [Online]. Available: https://medium.com/@brijesh_soni/why-random-forests-outperform-decision-trees-a-powerful-tool-for-complex-data-analysis-47f96d9062e7

[32] Ahmed Hosny et al., "Artificial Intelligence in Radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500-510, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[33] Koichiro Yasaka et al., "Deep Learning with Convolutional Neural Network in Radiology," *Japanese Journal of Radiology*, vol. 36, no. 4, pp. 257-272, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[34] Ke Yan et al., "Deeplesion: Automated Mining of Large-Scale Lesion Annotations and Universal Lesion Detection with Deep Learning," *Journal of Medical Imaging*, vol. 5, no. 3, pp. 1-11, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[35] Taeho Jo, Kwangsik Nho, and Andrew J. Saykin, "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data," *Frontiers in Aging Neuroscience*, vol. 11, pp. 1-14, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[36] Kenneth Clark et al., "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[37] Adnan Qayyum et al., "Secure and Robust Machine Learning for Healthcare: A Survey," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156-180, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[38] Shijun Wang, and Ronald M. Summers, "Machine Learning and Radiology," *Medical Image Analysis*, vol. 16, no. 5, pp. 933-951, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[39] Xiaoxuan Liu et al., "A Comparison of Deep Learning Performance against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271-e297, 2019. [Google Scholar] [Publisher Link]