

Lung Cancer Disease Prediction using Artificial Intelligence: A Systematic Review

Sneha Khoria¹ and Raghuraj Singh²

Review Paper

¹Reserach Scholar, Harcourt Butler Technical University, Kanpur, INDIA

² Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, INDIA

Email: snehakhoria@gmail.com

Received: 1 July 2023, Revised: 26 August 2023, Accepted: 31 August 2023

Abstract:

Lung cancer remains a global health challenge with profound implications for patient outcomes and healthcare systems. In recent years, artificial intelligence (AI) has emerged as a promising tool for enhancing the accuracy and efficiency of lung cancer prediction. This systematic review aims to provide a comprehensive synthesis of the state-of-the-art AI-based approaches in lung cancer prediction, shedding light on the advancements, challenges, and future directions in the field. The findings of this systematic review highlight the remarkable progress made in the use of AI for lung cancer prediction. Image-based approaches, particularly those utilizing deep learning techniques, have shown exceptional promise in improving early diagnosis through the analysis of radiological and histological data. Additionally, the integration of diverse data sources, such as clinical records, genetic information, and patient demographics, has led to more holistic predictive models. In nutshell, this systematic review provides a comprehensive overview of the role of artificial intelligence in lung cancer prediction. It highlights the transformative potential of AI-driven approaches in improving early detection and prognosis. As the field continues to evolve, addressing the identified challenges and fostering interdisciplinary collaboration will be essential in harnessing the full potential of AI to mitigate the impact of lung cancer on public health.

Keywords: Lung Cancer, AI, Deep Learning

1. Introduction

Lung cancer is one of the most common cancers globally and a leading cause of cancer-related deaths. It can be broadly categorized into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is more common and accounts for about 85% of all lung cancer cases [1]. In terms of cases and deaths, lung cancer statistics can vary by region, time period, and various other factors. Lung cancer was one of the most commonly diagnosed cancers worldwide. The number of new cases each year varies by country and region. High rates of smoking, exposure to environmental pollutants, and other risk factors contribute to the incidence of lung cancer. Lung cancer is also a major cause of cancer-related deaths. The mortality rate associated with lung cancer is high, partly due to late-stage diagnoses and the aggressive nature of the disease. Treatment outcomes depend on factors such as the type and stage of lung cancer, the patient's overall health, and the availability of effective treatments. Efforts to reduce lung cancer cases and deaths include anti-smoking campaigns, early detection through screening programs, advances in treatment options, and personalized medicine approaches targeting specific genetic mutations. It's important to consult recent and reliable sources for the most up-to-date lung cancer statistics and trends. The spectrum of imaging features linked to lung cancer is extraordinarily varied, ranging from tiny single nodules to complex patterns [2]. It turns out to be a very difficult task to find small, uncomplicated lesions [3]. Lung cancer has a wide range of histopathological subtypes, including adenocarcinoma, squamous cell carcinoma, small cell carcinoma, and a number of unusual histological forms. The existence of numerous subtypes adds to the complexity of histology. As an example, the classification of lung cancers by the World Health Organization in 2015

identified at least six predominant subtypes and a total of eleven subtypes for adenocarcinoma alone [4], with more subtypes added in the 2021 [5]. A trio of criteria, including the lung cancer's genomic characteristics, histology, and clinical staging, play a significant role in determining the best course of treatment. Clinicians must carefully compile all relevant information in the era of precision medicine in order to make judgments regarding chemotherapy, targeted therapy, immunotherapy, and possibly combined methods involving surgery or radiotherapy. In routine clinical practice, the choice of whether to start therapy or take a wait-and-see attitude looms constantly. Clinicians work to identify the inherent relationship between their observations and interventions (inputs) and the resulting results (outputs) in this situation. In essence, this endeavor entails developing disease detection, categorization, or prediction models. Currently, clinical trial findings and the combined experience of medical professionals serve as the foundation for such understanding. However, this method places a heavy strain on medical professionals because it requires repeated examination of pathology slides and imaging data in order to make accurate diagnosis. The process of studying patient information to identify the best therapy options also takes a lot of time. A powerful prediction or classification model has the potential to simplify this complex operation. Here we have artificial intelligence (AI), a term that encompasses a broad range of technologies without having a strict meaning. AI uses an algorithmic framework that is powered by pre-existing data to anticipate or categorize different items [6]. Its fundamental components include the dataset used for training, preprocessing techniques used on the data, the algorithm used to build the predictive model, and pre-trained models that speed up model construction by drawing on prior knowledge. Machine Learning (ML), a discrete subset of the larger field of artificial intelligence, represents the scientific quest to create algorithms that can solve problems without explicit programming.

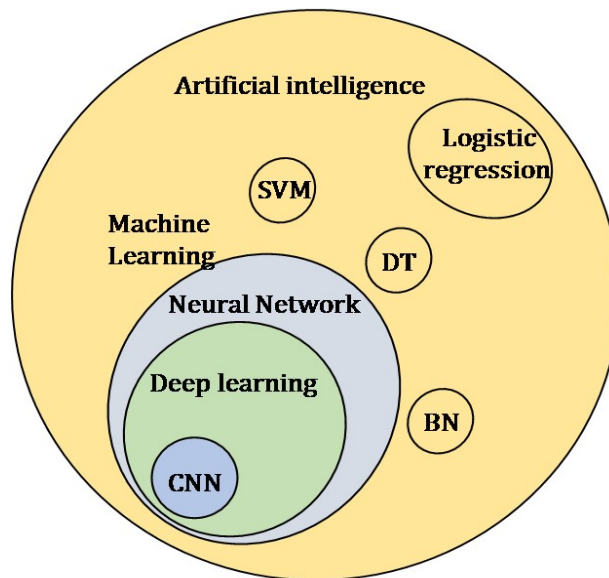


Figure 1: Classification of AI techniques

In-depth research reveals Deep Learning (DL) as a specific subcategory nested within ML. Deep learning accomplishes feature extraction and model fitting simultaneously because to its hierarchical design [7]. The algorithm may gradually pick up higher-level features thanks to the layered architecture, which eventually helps it interpret complicated patterns in the data more precisely. To elucidate the hierarchical relationship among these concepts, refer to Figure 1, which visually encapsulates the interconnections between AI, ML, and DL.

However, a significant number of computational resources are required for the creation of such complex models. Building a multidimensional algorithm for the interpretation of images used to be a laborious procedure that frequently required hours or even days of human brainpower. The development of complex predictive models is significantly hampered by this enormous computing power demand. Fortunately, the rapid development of chip technology and software optimization has increased the amount of computational power that is readily available, making it much easier to carry out complex calculations [8, 9]. The design of models that greatly exceed the constraints of logistic or linear regression has become possible thanks to the capacity to quickly compute enormous matrices.

2. Screening Methods

Around 7% of those with lung cancer have no detectable symptoms, which is a significant number [10]. More than 50% of people who have lung cancer resection surgeries also show no symptoms, which complicate this complex dynamic [11]. With the help of these revelations, numerous initiatives to put into practice efficient screening procedures have been made, utilizing a variety of tools as imaging. The ability to offer useful insights remains most obvious in the imaging domain among these techniques. Although techniques such as sputum cytology, blood testing, and breath analysis have virtues of their own, they frequently have drawbacks in terms of sensitivity, specificity, or usefulness.

Surprisingly, when it comes to imaging, the landscape changes toward two commonly used methods: low-dose computed tomography (LDCT) and chest X-rays (CXR). LDCT stands out as the technique that has shown to be able to detect lung cancer at an earlier stage and, as a result, lengthen the survival rates of people dealing with this condition, despite the fact that CXRs are a mainstay in clinical practice [12].

As long durations of image interpretation may cause human visual issues, the repeated nature of the imaging reading process provides a fertile environment for the incorporation of AI. Long periods of picture reading can cause weariness and pressure on the eyes, which progressively robs the images of their clarity and focus. AI is positioned to offer a novel remedy to this over time occurring reduction in interpretational accuracy. The inherent proneness to errors during the interpretation of CXR and LDCT images further complicates this issue. These mistakes can have serious consequences and are frequently a factor in numerous medical malpractice claims [13]. Given the possible repercussions for medical care, these legal challenges highlight the vulnerability that results when human error sneaks into picture interpretation.

It's important to remember that even specialists have limitations, despite research showing that expert radiologists tend to recognize more pulmonary nodules on CXRs than their less experienced peers [14]. In reality, a sizable portion of lung nodules roughly 20% go unnoticed, highlighting the inherent difficulty of this undertaking. This discrepancy in detection rates highlights the complex nature of radiological interpretation and the potential for AI to support human expertise.

A. CXR

The most widely used modality in the medical field, (CXR) serve as the foundation of medical imaging. CXRs provide a thorough assessment of the thoracic region of a patient with a radiation exposure of just 0.1 mSv, or about 10 days' worth of naturally occurring background radiation. With the development of computer-aided diagnostic (CAD) systems specifically created for interpreting CXRs dating back to the 1960s [15], this non-invasive imaging approach has been around from a time before the digital imaging era. According to the conventional paradigm, significant image characteristics including shape, size, intensity, and texture required manual annotation before being subjected to further analytical procedures. This environment has undergone a dramatic change as a result of the advent of the digital era. These days, computers are equipped with the ability to analyze photos on their own. The field of radiomics evolves from a painstaking pixel-by-pixel computation of the image content, redefining the idea of image features from a computational perspective. This method simplifies the computation of image texture and density by utilizing a wide variety of mathematical techniques. As a result, the area of interest within an image can be transformed into a multi-dimensional dataset that is represented by a large matrix.

Radiomics is based on mathematical principles, which makes it difficult for computers to handle the wide range of visual characteristics included in CXRs. The pursuit of accurate radiomics data emerges as a crucial activity when navigating this environment. In this regard, the year 2022 saw a significant advancement: image augmentation, a critical operation that arose as a necessary step before beginning nodule detection [16], became a requirement method. This complex procedure includes crucial steps including pre-processing, lung segmentation [17], and rib suppression [18]. Each of these components is essential for improving the CXR data's quality and interpretability, which supports the precision and dependability of radiomics analysis. Also, malignancy/benign classification were done with the help of a different algorithm. Prior to 2011, algorithms based on Decision Tree (DT) techniques were the most popular analytical tools for handling these complex aspects. But in 2011, the market witnessed a dramatic change as deep learning-based algorithms became strong competitors and shown their expertise in the field of picture analysis. The CheXNet algorithm is a prominent example of this paradigm shift. CheXNet, which was designed as a deep learning model, acquired a level of expertise comparable to that of a qualified radiologist. The world's largest CXR database of its kind, Chest-Xray14, served as the training ground for this remarkable algorithm. The results were

astounding, outperforming radiologists' abilities to identify a variety of 14 lung conditions. Area Under the Receiver Operating Curve (AUROC) values of 0.78 and 0.87, respectively, were used to capture conditions such lung nodules and lung masses [19]. Notably, AUROC levels are a measure of diagnostic effectiveness. Later models reached even higher levels of sensitivity by delving deeper into the world of deep learning. While retaining a false-positive rate of 0.2 per Chest X-ray, these models were able to increase sensitivity to an extraordinary 0.83 [20]. The algorithm's ability to reliably identify actual positive situations while minimizing false alarms is enhanced by this improvement in sensitivity. It's interesting that many software packages have received the Food and Drug Administration's (FDA) seal of approval in terms of regulatory approval. This acknowledgment highlights the reliability and clinical viability of these algorithms, attesting to their ability to make a significant contribution to the field of patient care and medical diagnostics [21].

B. Chest CT

A non-invasive way to examine the complex three-dimensional anatomy of the thoracic area has been made possible using CT technology. As this technology has developed, there has been a noteworthy decrease in radiation exposure, going from the once typical 7 mSv associated with traditional chest CT to a noticeably reduced 1.6 mSv with Low-Dose CT (LDCT) techniques. The National Lung Screening Trial (NLST) [22] and the Dutch-Belgian Randomized Lung Cancer Screening Trial (abbreviated as NELSON study) [23] are two important randomized control trials that demonstrate the key shift that the introduction of LDCT screening has sparked. In these studies, LDCT screening showed an impressive 20% reduction in fatality rates.

3. AI Based Lung Cancer Prediction Techniques

The use of AI into daily life has become a dominant trend in the twenty-first century, spreading its effect to a variety of industries, including the field of medicine. Particularly in this change, the medical industry has embraced the possibilities of AI. Lung cancer stands out among medical illnesses as a top prospect for the use of AI-driven solutions because to its intrinsic heterogeneity. Numerous studies have investigated the use of AI to address various aspects of lung cancer, including tasks like the identification of lung nodules, the diagnostic interpretation of histopathology samples, the stratification of disease risks, the facilitation of drug development, and even the prediction of prognosis. This article acts as a thorough narrative review, exploring the various ways that AI is being used to treat lung cancer. The strategy used entails a thorough investigation of reported applications based on the sequential steps of the clinical workflow, followed by a systematic analysis of AI model architectures. These phases involve prognosis prediction at the end of the diagnosis and screening phases.

AI Models

Currently, several AI models built using various techniques are developed. Generally speaking, there are four categories into which AI models can be split: reinforcement learning, semi-supervised learning, and supervised learning (Figure 2).

A. Supervised Learning

Within the realm of supervised learning, the foundation lies in the creation of meticulously curated datasets encompassing both inputs and their corresponding desired outputs or answers. This assembly of data serves as the training ground for algorithms, equipping them to undertake prediction-based challenges, including the realms of classification and regression. The very architecture of these algorithms is notably diverse, offering a spectrum of strategies for researchers to employ. A prominent approach involves the utilization of binary nodes, which when amalgamated, give rise to Decision Trees (DTs), effectively serving as classifiers. On the other hand, the methodology of Support Vector Machines (SVMs) involves the exploration of multidimensional spaces to discern planes that effectively demarcate different classes, thereby yielding a robust classifier. In a different vein, Bayesian classifiers leverage input data to computationally derive probabilities that inform the correct classification, establishing a probabilistic foundation. Interestingly, the probabilities extracted through the aforementioned algorithms can be harnessed to transform categorical answers into continuous variables, effectively providing a versatile tool for tackling regression problems, and conversely, offering a mechanism to convert continuous outcomes into categorical solutions. This adaptability lends itself to a host of scenarios, particularly within the domain of AI applications. Many of these AI

applications, which revolve around predicting outcomes such as survival rates [24], cancer risk [25, 26], the detection of nodules [27], and the characterization of these nodules [28], find their cornerstone in the realm of supervised learning. By employing these techniques, researchers are able to harness the power of labeled data to enhance the precision and efficacy of their predictive models, ultimately contributing to advancements in fields such as medical diagnosis and prognosis.

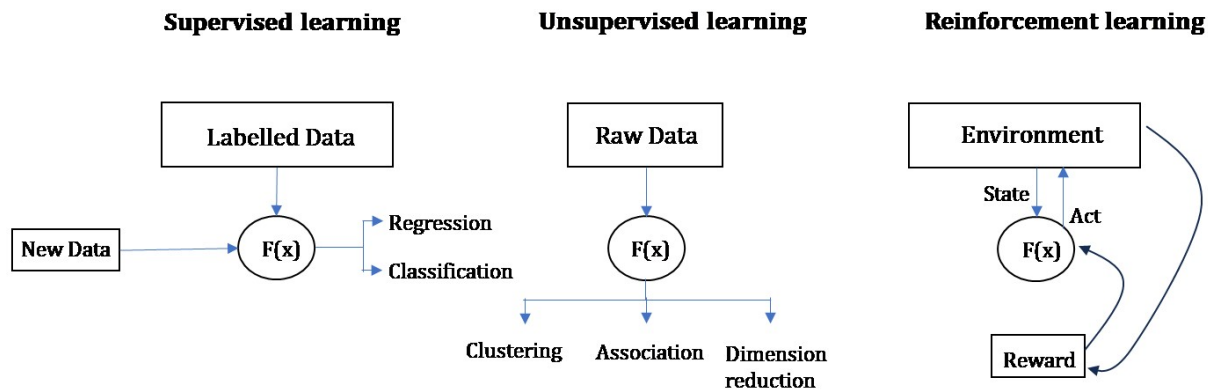


Figure 2: Types of various learning techniques

Sesen et al. [24] provides a profound exploration of how Bayesian networks can be harnessed to enhance clinical decision-making within the context of lung cancer care. This research is a testament to the ongoing pursuit of innovative solutions to the challenges posed by the complex landscape of lung cancer diagnosis and treatment. The primary objective of this study was to introduce a sophisticated framework that leverages Bayesian networks to assist medical professionals in making well-informed decisions related to lung cancer care. Recognizing the intricate interplay of various clinical variables and their impact on disease progression, the researchers aimed to construct a probabilistic model that could effectively capture these relationships, thus aiding in accurate diagnosis, prognosis, and treatment planning.

Huang et al., [25] presented a pioneering effort to harness cutting-edge technology in the realm of lung cancer detection. This study marks a significant advancement in leveraging chemical sensor arrays and machine learning techniques for accurate and non-invasive diagnosis, reflecting the constant quest for innovative approaches to address the challenges posed by lung cancer diagnosis.

Kort et al., [26] work stands as a significant stride toward enhancing the accuracy and effectiveness of lung cancer diagnosis through a multidimensional approach. This study represents a significant advancement in the realm of medical diagnostics, where the integration of exhaled-breath data and clinical parameters demonstrates the potential to revolutionize the landscape of lung cancer detection. The core objective of this study was to explore the synergistic potential of combining exhaled-breath data with conventional clinical parameters to enhance the diagnostic accuracy of lung cancer. Acknowledging the complexities and challenges associated with early lung cancer diagnosis, the researchers aimed to create a holistic diagnostic strategy that capitalizes on the complementary nature of these two distinct data types.

Kort et al. [27] work stands as a significant stride toward enhancing the accuracy and effectiveness of lung cancer diagnosis through a multidimensional approach. This study represents a significant advancement in the realm of medical diagnostics, where the integration of exhaled-breath data and clinical parameters demonstrates the potential to revolutionize the landscape of lung cancer detection. The core objective of this study was to explore the synergistic potential of combining exhaled-breath data with conventional clinical parameters to enhance the diagnostic accuracy of lung cancer. Acknowledging the complexities and challenges associated with early lung cancer diagnosis, the researchers aimed to create a holistic diagnostic strategy that capitalizes on the complementary nature of these two distinct data types.

B. Unsupervised Learning

Unsupervised learning constitutes a distinct paradigm where algorithms autonomously segregate samples based solely on their input characteristics, obviating the requirement for labeled data. This approach finds its niche in tasks such as clustering analysis, uncovering inter-sample associations, and facilitating dimensionality reduction. Clustering analysis stands as a prime example of unsupervised learning's prowess. By allowing algorithms to autonomously group samples with similar attributes, valuable insights can be gleaned from the data landscape. This technique has found meaningful application within the study of lung cancer, exemplified by instances where cluster analysis was harnessed to unveil key oncogenic signatures within this complex disease landscape [29,68]. This process not only refines our understanding of the intricate nature of lung cancer but also paves the way for more targeted and personalized therapeutic interventions.

Girard et al., [29] work signifies a crucial milestone in the realm of lung cancer genetics. This study represents a groundbreaking effort to dissect the genomic landscape of lung cancer through comprehensive allelotyping, unravelling distinctive patterns of genetic alterations that differentiate SCLC from (NSCLC, and highlighting potential oncogenic loci clustering.

Shen et al., [30] work presents a pioneering effort in the field of cancer genomics. This study introduces a sophisticated integrative approach that leverages multiple genomic data types to unravel hidden patterns in breast and lung cancer subtypes, thus exemplifying the power of computational techniques to extract meaningful insights from complex biological datasets.

Lynch et al.[31] constitutes a significant stride in the realm of lung cancer research and personalized medicine. This study exemplifies the innovative application of unsupervised analysis techniques to large-scale patient data, shedding light on the potential of data-driven approaches to uncover hidden insights, patterns, and subgroups within lung cancer patient populations. The primary objective of this study was to explore the utilization of unsupervised analysis techniques to dissect and comprehend the complexities of lung cancer patient data. Recognizing the diverse nature of lung cancer and the inherent variability within patient populations, the researchers aimed to unravel latent structures that could potentially contribute to improved understanding and stratification of the disease.

C. Semi-Supervised Learning

While supervised learning undoubtedly offers superior accuracy in algorithmic outcomes, a significant drawback looms in the form of the scarcity of labeled data, coupled with the labor-intensive nature of the labeling process. Conversely, unsupervised learning leverages the advantage of utilizing unlabeled data, but its predictive accuracy often falls short of its supervised counterpart. In this delicate balancing act between precision and data availability, a compelling solution emerges in the form of semi-supervised learning.

Semi-supervised learning stands as a strategic middle ground, effectively harnessing the strengths of both supervised and unsupervised learning paradigms. At its core, this approach capitalizes on the power of supervised learning to forge a labeling tool, thus circumventing the need for exhaustive manual labeling. This inventive strategy not only alleviates the laborious burden of labeling large datasets but also contributes to enhancing the overall efficiency of the learning process.

Eckardt et al., [32] work marks a significant advancement in the field of cancer diagnostics, specifically in the context of leveraging semi-supervised learning techniques. This study underscores the innovative use of machine learning to tackle the challenges inherent in cancer diagnostics, emphasizing the potential of harnessing both labeled and unlabeled data to enhance accuracy and efficacy. The primary objective of this study was to explore the application of semi-supervised learning methodologies in the domain of cancer diagnostics. While supervised learning relies on labeled data for training, semi-supervised learning capitalizes on the wealth of unlabeled data available in medical contexts. Given the complexities of cancer diagnostics and the limitations of obtaining large amounts of labeled data, the researchers aimed to investigate how semi-supervised learning could address this data scarcity challenge.

Nguyen et al., [33], represented a significant leap forward in the realm of lung cancer detection using advanced machine learning techniques. This study highlights the innovative utilization of active semi-

supervised learning to address the challenges of limited labeled training data, specifically in the context of lung cancer detection from Computerized Tomography (CT) images. The primary objective of this study was to tackle the scarcity of labeled training data, a common bottleneck in training accurate machine learning models. This is particularly crucial in medical imaging where obtaining large labeled datasets can be resource-intensive and time-consuming. The researchers aimed to address this challenge by introducing active semi-supervised expectation maximization learning, a powerful framework that capitalizes on both labeled and unlabeled data.

Shak et al., [34] presents a significant advancement in the domain of lung cancer prediction by introducing a novel semi-supervised self-training approach. This study underscores the innovative use of semi-supervised learning to leverage both labeled and unlabeled data for improved accuracy and efficiency in lung cancer prediction. The central objective of this study was to address the challenge of limited labeled training data in the field of lung cancer prediction. Recognizing the potential of unlabeled data to enhance model performance, the researchers aimed to devise a method that harnesses the insights from unlabeled data while maintaining the accuracy of a supervised learning model.

D. Reinforcement Learning

Reinforcement learning represents a distinctive paradigm grounded in a reward-based framework. The algorithm's growth is intricately interwoven with its interaction with the environment, often manifesting as the dataset it engages with. Within this dynamic landscape, a reward function takes center stage, orchestrating the fine-tuning of either the algorithm itself or the underlying neural network. This methodology transcends mere algorithmic evolution; it's a voyage of refinement driven by rewards. By navigating through interactions with the environment, the algorithm progressively learns to optimize its responses based on the feedback it receives. This iterative loop of action, feedback, and adjustment mirrors the learning process observed in living organisms, leading to a more sophisticated and contextually aware AI system. This adaptive nature of reinforcement learning carries profound implications. Researchers have harnessed this technique to fashion AI systems that not only self-evolve but also excel in intricate tasks, such as detecting nodules within CT images.

Zhao et al.,[35] presented a pioneering effort in the realm of clinical trial design within the context of (NSCLC. This study introduces the innovative application of reinforcement learning strategies to optimize the design of clinical trials, showcasing the potential of computational techniques to enhance the efficiency and effectiveness of trial protocols.

Tseng et al., [36] presented a groundbreaking advancement at the intersection of artificial intelligence, medical physics, and cancer treatment. This study showcases the pioneering utilization of deep reinforcement learning to automate the process of radiation adaptation in lung cancer treatment, highlighting the potential of cutting-edge technology to optimize therapeutic interventions. The primary objective of this study was to address the intricate challenge of radiation adaptation in lung cancer treatment. Radiation therapy is a cornerstone of cancer treatment, aiming to target tumors while minimizing damage to healthy tissues. However, adapting radiation plans to accommodate changes in patient anatomy and tumor response poses a significant challenge. The researchers aimed to harness deep reinforcement learning to create an automated system that dynamically adjusts radiation plans based on real-time patient data.

Wang et al., [37] did significant breakthrough in the realm of cancer treatment planning using cutting-edge technology. This study introduces an integrated approach that leverages deep reinforcement learning to automate the process of intensity-modulated radiation therapy (IMRT) treatment planning for NSCLC, demonstrating the potential of advanced AI techniques to revolutionize therapeutic strategies. The primary objective of this study was to address the complexities and challenges associated with IMRT treatment planning in NSCLC cases. IMRT is a sophisticated technique that delivers precise radiation doses to tumors while sparing nearby healthy tissues. However, crafting optimal IMRT treatment plans requires expert knowledge, intricate calculations, and time-consuming iterations. The researchers aimed to harness deep reinforcement learning to create an automated solution that streamlines the treatment planning process while maintaining high-quality plans.

E. Notable Techniques

Alsinglawi et al. [38] presented a novel predictive model in their research with the goal of calculating the length of stay (LOS) for people with lung cancer. This structure used a variety of supervised machine learning classifiers, such as Logistic Regression, Random Forest, and XGBoost. The authors used a K-fold 10 cross-validation procedure to make sure their results were reliable. The MIMIC-III dataset was used by the researchers to examine the patterns of length of stay among patients who had been admitted to the intensive care unit (ICU). It is important that the dataset displayed an inherent imbalance, indicating that the number of samples for various groups (lengths of stay) was not distributed evenly. To fix this problem, the authors used over-sampling techniques, particularly the Synthetic Minority Over-sampling Technique (SMOTE), during the validation process. 53,423 adult patients made up the study's large sample size, which was used to conduct the analysis. The authors' thorough investigation resulted in some important findings. They found that Random Forest, when combined with the SMOTE class balancing method, performed the best among the three classifiers (Random Forest, XGBoost, and Logistic Regression). The outcomes of their research were surprising. An outstanding area under the curve (AUC) score of 98% was attained by the Random Forest classifier using SMOTE, indicating an exceptionally high degree of accuracy in differentiating between various lengths of stay.

Venkatesh et al. [39] in order to forecast cases of lung cancer, unique methodologies using ensemble learning were applied in the study of The Surveillance, Epidemiology and End Results (SEER) dataset, which consisted of 1000 samples and 149 different attributes for each of them, was used to build the predictive models. The researchers started using a variety of machine learning classifiers, including K-Nearest Neighbors, Decision Trees, and Neural Networks, along with ensemble approaches, especially Bagging and Adaboost. These methods were applied to thoroughly evaluate the accuracy of lung cancer prediction. The researchers came to some important results after careful analysis. Adaboost performed best among the ensemble approaches used, demonstrating its greater prediction accuracy. This method impressively has a 98.2% accuracy rate, supporting its promise as a reliable tool for accurately finding lung cancer cases in the SEER dataset.

Vikas et al. [40] used two different machine learning techniques, Support Vector Machine and Random Forest. The algorithms were compared with and without the feature selection method known as Chi-square in order to refine the procedure. The researchers discovered something important as a result of their analysis. The Support Vector Machine showed superior results in both computational effectiveness and prediction performance. Incredibly impressive metrics were attained by this algorithm, including a 98% accuracy rate, a flawless precision score of 100%, an excellent recall rate of 100%, and an exceptional F1-Score of 100%. Additionally, this algorithm's execution time, which was very quick at 0.010 seconds, was impressive.

Puneet et al.'s [41] work focused on creating a lung cancer prediction model using machine learning techniques that made use of routine blood indices. A combination of methods, including XGBoost, GridSearchCV, Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, Decision tree, and K-Nearest Neighbor classifiers, were used for this task. A K-fold 10 cross-validation strategy was used to evaluate the predicting skills in order to assure robustness. The dataset used in this study came from Lanzhou University and included 277 patients in total. The results of this dataset were carefully predicted using the aforementioned classifiers. Surprising results came from this study. Among the classifiers being evaluated, XGBoost won out, displaying higher performance across important evaluation metrics. Notably, XGBoost demonstrated its accuracy in prediction by achieving a rate of 92.16%. Additionally, it had a recall rate of 96.97%, which is amazing and shows how well it can detect real positive cases. Another indicator of XGBoost's superiority was the Area Under the Curve (AUC), which measures the classifier's capacity to distinguish between various classes and reached a value of 95%.

Sim et al.'s [42] analysis focused on health-related quality of life (HRQOL) in the context of forecasting lung cancer 5-year survival rates. In this work, different machine learning models were developed, and their effectiveness in this ability for prediction was evaluated. Decision Tree, Logistic Regression, Bagging, Random Forest, and AdaBoost were all included in the group of models. The process of assessment was carried out across two distinct feature sets and a K-fold 5 cross-validation technique was used. By using these models to clinical data derived from 809 individuals who underwent surgical interventions for lung cancer, the

researchers targeted to gain insightful knowledge into their predictive capabilities. The results of this thorough investigation produced a few interesting findings. AdaBoost and Random Forest emerged out as the top models from the wide range of models evaluated, displaying excellent performance metrics. Notably, AdaBoost attained a noteworthy accuracy rate of 94.8%, demonstrating its skill in precise prediction. Additionally, it achieved a great Area Under the Curve (AUC) value of 94.9%, demonstrating its skill at differentiating between various outcomes.

Patra [43] carried out a thorough investigation using a variety of machine learning classifiers, each of which is unique in type. Radial Basis Function Network (RBF), K Nearest Neighbors (KNN), SVM, Artificial Neural Network (ANN) and Random Forest were among the classifiers used. WEKA tools were used to make it easier to develop these classifiers. A dataset of 32 instances was used for the analysis, and each instance's 57 properties were taken from the "UCI repository." The researchers sought to determine the relative efficacy of these classifiers in predicting lung cancer through a careful comparative study of the results produced by them. This study's main finding was especially significant. Radial Basis Function Network (RBF) stood up as the top performer among the variety of classifiers examined, outperforming the others in terms of predictive power. RBF impressively earned an accuracy rate of 81.25%, demonstrating the accuracy of its predictions. Additionally, it showed 81.3% precision, recall, and F1-score, indicating its skill in both detecting positive situations and accurately forecasting outcomes. Calculated at 74.9%, the Area Under the Curve (AUC) measures how well the classifier can distinguish between various classes.

P.R. et al [44] started a study that used four different Machine Learning algorithms and was intended to predict lung cancer. Naive Bayes, Support Vector Machine, Decision Tree, and Logistic Regression were some of these methods. The two main goals of this research project were to enable early lung cancer diagnosis and to undertake a thorough examination of the performance of the algorithms. The researchers gathered a dataset of 1000 examples, each identified by 25 distinct attributes, to aid their inquiry. The "Data World" website was used to obtain this dataset. The scientists set out to identify which algorithm showed greater prediction powers and might, thus, be successfully used for the early diagnosis of lung cancer. The Support Vector Machine algorithm was the best performer of the four studied algorithms. This algorithm's impressive accuracy rate of 99.2% highlights how precise it is at making reliable predictions. This result demonstrated the algorithm's advantage over the other classifiers included in the study.

Wu et al. [45] used Random Forest, to distinguish and identify lung cancer cases, using common cancer indices as the foundation for their predictions. A reliable K-fold 10 cross-validation procedure was used to verify the effectiveness of their suggested model. The data for 277 patients were included in the precisely assembled dataset for this study from Lanzhou University. The researchers wanted to test the prediction power of their Random Forest model by applying it to this dataset. The Random Forest model achieved a noteworthy rate of 95.7%, demonstrating extraordinary accuracy. This demonstrated how skilled it was at appropriately classifying situations. The model also showed a remarkable recall rate of 96.3%, demonstrating its efficiency in catching real positive cases. The model's ability to differentiate between various classes was evaluated using the Area Under the Curve (AUC), which was calculated at a high value of 99.01%.

Faisal et al. [46] undertook a thorough investigation involving numerous machine learning approaches and ensemble learning methods. The researchers used a variety of classifiers, including Naive Bayes, Neural Network, SVM, Gradient Boosted Tree, and Random Forest, to evaluate and compare the performance of these approaches. K-fold 10 cross-validation was used to carefully carry out this evaluation method. The authors focused on the Gradient Boosted Tree, a well-known Ensemble Learning method, in search of discoveries. The Gradient Boosted Tree outperformed all other individual classifiers in predictive power, which made their findings more striking. This approach successfully generated accurate forecasts with a noteworthy accuracy rate of 90%, which is impressive. The precision score of 87.82%, demonstrating its accuracy in making positive predictions, was one of the other performance indicators that was equally impressive. Furthermore, an F1-score of 85.71% and a recall rate of 83.71% demonstrated its effectiveness in identifying true positive cases. The dataset used in this research came from the UCI repository and consisted of 32 instances in total, each of which had 57 attributes. Faisal and colleagues [15] established the framework for reliable lung cancer prediction by utilizing this dataset and meticulously evaluating the ensemble and individual methodologies.

Safiyari et al. [47] introduced a wide range of methods, including Bagging, Dagging, AdaBoost, MultiBoosting, and Random SubSpace. In order to forecast lung cancer survival rates, a variety of other classification methods, including RIPPER, Decision Stump, Simple Cart, C4.5, SMO, Logistic Regression, Bayes Net, and Random Forest, were also used. The researchers used the under-sampling method to conduct a thorough review of these methods' effectiveness. They focused their investigation on the Surveillance, Epidemiology and End Results (SEER) dataset, a sizable collection that includes a whopping 643,924 samples, each of which is distinguished by 149 different attributes. The researchers made an important finding after carefully comparing and contrasting the results of these various methodologies. The AdaBoost algorithm stood out as the top performer, excelling in crucial parameters like accuracy and Area Under the Curve (AUC). AUC was calculated to be an excellent 94.9%, demonstrating the algorithm's ability to differentiate across various classes. Additionally, the accuracy rate rose to a remarkable 88.98%, underscoring the strength of AdaBoost's capacity to make precise predictions.

4. Limitation of the Lung Cancer Prediction Methods

Lung cancer detection methods, while continually improving, still have certain limitations. These limitations can impact the accuracy of diagnosis and the effectiveness of treatment. As of my last update in September 2021, here are some key limitations of lung cancer detection methods:

Early Stage Detection Challenges: Lung cancer's elusive nature during its early stages presents a significant challenge in its detection and management. This phenomenon is rooted in the biology of the disease and the fact that symptoms typically only manifest when the cancer has already reached more advanced stages. This delay in symptom appearance can have critical implications for patient outcomes, treatment effectiveness, and overall survival rates.

Silent Development: Lung cancer is notorious for its silent growth in its initial phases. The disease can progress within the lungs without causing noticeable symptoms that would prompt patients to seek medical attention. Unlike more symptomatic conditions, such as infections or inflammatory disorders, lung cancer's early growth does not usually trigger alarms in the body's systems. As a result, individuals may remain unaware of its presence until the disease has already reached a more advanced stage.

Delayed Presentation: By the time lung cancer begins to produce symptoms, it often signifies that the disease has reached an advanced stage. Symptoms can vary depending on the tumor's location, size, and type, but they might include persistent coughing, coughing up blood, chest pain, unexplained weight loss, fatigue, and difficulty breathing. Unfortunately, these symptoms are not specific to lung cancer and can be associated with other conditions as well, leading to potential misdiagnoses or delayed detection.

Impact on Treatment and Survival: Early detection of cancer is crucial because treatments are often more effective when initiated at an early stage. When lung cancer is diagnosed in its advanced stages, it is more likely to have spread beyond the lungs, limiting treatment options and complicating the management of the disease. At these stages, curative treatments like surgery might not be feasible, and the focus shifts to palliative care to manage symptoms and improve quality of life. Moreover, the limited treatment options for advanced-stage lung cancer can lead to reduced overall survival rates. Effective treatments, such as surgery, targeted therapies, and radiation, are most successful when applied early, aiming to eradicate the disease or control its growth. In contrast, advanced lung cancer might require more aggressive treatments with potentially harsher side effects, which can further impact a patient's well-being.

Importance of Early Detection: The challenge of detecting lung cancer in its early stages highlights the significance of regular screenings for individuals at high risk, such as heavy smokers or those with a family history of lung cancer. Screening methods, such as LDCT scans, can increase the chances of identifying lung cancer at an earlier, more treatable stage. Educating healthcare professionals and the general public about the risk factors, importance of early screening, and potential symptoms of lung cancer is essential. By raising awareness and promoting early intervention, there is a better chance of identifying lung cancer when treatment options are most effective, potentially improving patient outcomes and overall survival rates.

False Positives and False Negatives: A false-positive result occurs when a diagnostic test indicates the presence of a condition, such as lung cancer, when there is no actual disease. In the context of lung cancer detection, a false-positive result could lead to unnecessary anxiety and distress for the patient. Being informed that one has cancer, only to later find out that it was a false alarm, can cause emotional turmoil and stress for the individual and their loved ones. False positives may also trigger additional invasive procedures and follow-up tests, such as biopsies or surgeries, to confirm the initial finding. These procedures can carry risks, discomfort, and potential complications, all of which might have been avoidable if the initial test had been accurate. On the other hand, a false-negative result occurs when a diagnostic test fails to detect the presence of a condition that is actually present. In the case of lung cancer, a false-negative result could delay the diagnosis and subsequent treatment. This delay might allow the cancer to progress to more advanced stages, potentially reducing the effectiveness of treatment options and lowering the chances of successful outcomes. False negatives can lead to missed opportunities for early intervention and timely treatment. Patients might experience a sense of security due to the negative test result, even as the disease continues to develop within their bodies.

Size and Location of Tumors: The sensitivity of lung cancer detection methods is intricately tied to the characteristics of the tumors themselves, including their size, location within the lungs, and specific tumor type. These factors play a crucial role in determining the effectiveness of various imaging techniques, which in turn can impact the accuracy of diagnosis and the subsequent treatment strategy. The complexity of lung tumors makes it imperative to consider these aspects when designing diagnostic approaches.

Tumor Size and Detection: The size of a lung tumor is a critical factor influencing its detectability. Larger tumors are more likely to be detected during routine screenings or imaging tests due to their greater mass and potential to cause noticeable changes in lung structure. However, smaller tumors, especially those in their early stages of growth, can be much more challenging to detect. These tumors might be too small to cause symptoms or visible changes on imaging scans, leading to their potential escape from early detection efforts.

Tumor Location and Visibility: The location of a lung tumor within the lung tissue also has a significant impact on detection. Tumors located near the airways or central regions of the lungs are more likely to be visible on imaging scans and might cause symptoms like persistent coughing or wheezing. However, tumors located in peripheral regions, farther from the airways and closer to the lung's outer edges, can be more challenging to identify. These peripheral tumors might not cause noticeable symptoms until they have grown significantly or spread to nearby tissues.

Peripheral Tumors and Imaging Techniques: Peripheral tumors, due to their location away from the central airways, can pose specific challenges to certain imaging techniques. For example, conventional chest X-rays might not capture peripheral tumors effectively, as the tumors are not directly adjacent to the X-ray beam's path. Computed tomography (CT) scans are more sensitive and can detect smaller tumors, but even they might struggle to visualize very small peripheral tumors or those hidden behind other lung structures.

Innovations and Challenges: Advancements in imaging technologies and radiological techniques aim to address these challenges. High-resolution CT scans, positron emission tomography (PET) scans, and hybrid imaging approaches (PET-CT) have improved the ability to detect smaller or peripheral tumors. However, the complexity of lung anatomy and the variability of tumor locations mean that no single imaging method is foolproof in all cases. The development of more targeted and personalized approaches, such as liquid biopsies that analyze blood for tumor-related genetic material, holds promise for detecting even early-stage and hard-to-reach tumors. However, the variability of tumor characteristics among individuals adds an additional layer of complexity to the detection process.

Challenges in Small Cell Lung Cancer: SCLC is a subtype of lung cancer known for its aggressive nature, rapid growth, and tendency to spread swiftly to other parts of the body. These characteristics present unique challenges in both the early detection and treatment of the disease. Understanding the aggressive behavior of SCLC is crucial for devising effective diagnostic and therapeutic strategies.

Aggressive Growth and Early Detection: One of the defining features of SCLC is its aggressive growth pattern. SCLC cells multiply and divide rapidly, leading to the formation of tumors that can grow and spread at an accelerated rate. This rapid growth contributes to the challenge of early detection. By the time SCLC becomes clinically detectable or produces noticeable symptoms, it often indicates that the cancer has already reached an advanced stage. Early detection of SCLC is challenging due to its propensity to remain hidden within the lung tissue until it has reached a significant size. This characteristic is in contrast to NSCLC, where certain tumors might present symptoms or changes on imaging scans at an earlier stage.

5. Importance of Research and Early Intervention

In conclusion, the framework presented herein harnesses the potential of Wireless Sensor Networks to revolutionize underwater rescue operations. Through a synergy of advanced sensor technologies, communication infrastructure, data analytics, localization methods, and resource management strategies, this framework offers a comprehensive solution to the unique challenges of underwater disasters. While facing ongoing challenges in communication reliability and power efficiency, the promise of enhanced survivor detection, reduced response times, and optimized resource allocation is undeniable. As we navigate an era marked by rapid technological advancement and increasing marine safety needs, this framework serves as a pivotal milestone, offering hope for safer and more efficient underwater rescue missions, ultimately benefiting both human lives and marine ecosystems.

6. Future Development

Upcoming lung cancer AI applications might emphasize integration and applications. First, as AI is a data-driven technology, researchers can combine small datasets to produce enormous training data sets. Regulated data exchange is a significant barrier for researchers, though. A straightforward solution is federated learning, which shares the trained parameters instead of the data [48, 49]. In federated learning, the models were trained independently at various hospitals, and only the learned models were provided to the main server, shielding the raw data from the main server. After that, each hospital received a report based on the final model.

Second, the majority of earlier studies were carried out by independent experts and concentrated on distinct specialties like radiology, pathology, surgery, or clinical oncology. The reality, however, might be more accurately reflected by combining all factors, including imaging, pathology, demographics, clinical data, and both new and ancient technology. Researchers can create predictive models by combining several features [50].

A further difficulty is the deployment of AI algorithms, in addition to improving model accuracy by boosting the training sample size and integrating other disciplines. While research discussed above shown the potential benefits of using AI in lung cancer treatment and several products received FDA approval [51, 52], actual clinical workflow implementation is uncommon. Real-world applications are hindered by the user interface, data analysis speed, size, internet connection, and resource consumption of the AI program. Before moving in the AI-assisted world, more infrastructure has to be built.

References

1. Wei, Ming-Ming, and Guang-Biao Zhou. "Long non-coding RNAs and their roles in non-small-cell lung cancer." *Genomics, proteomics & bioinformatics* 14, no. 5 (2016): 280-288.
2. Rampinelli, Cristiano, Sonia Francesca Calloni, Marta Minotti, and Massimo Bellomi. "Spectrum of early lung cancer presentation in low-dose screening CT: a pictorial review." *Insights into imaging* 7, no. 3 (2016): 449-459.
3. Migliore, Marcello, Stefano Palmucci, Marco Nardini, and Antonio Basile. "Imaging patterns of early stage lung cancer for the thoracic surgeon." *Journal of Thoracic Disease* 12, no. 6 (2020): 3349.
4. Travis, William D., Elisabeth Brambilla, Andrew G. Nicholson, Yasushi Yatabe, John HM Austin, Mary Beth Beasley, Lucian R. Chirieac et al. "The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification." *Journal of thoracic oncology* 10, no. 9 (2015): 1243-1260.

5. Nicholson, Andrew G., Ming S. Tsao, Mary Beth Beasley, Alain C. Borczuk, Elisabeth Brambilla, Wendy A. Cooper, Sanja Dacic et al. "The 2021 WHO classification of lung tumors: impact of advances since 2015." *Journal of Thoracic Oncology* 17, no. 3 (2022): 362-387.
6. Suzuki, Kenji. "Overview of deep learning in medical imaging." *Radiological physics and technology* 10, no. 3 (2017): 257-273.
7. Lawson, Christopher E., Jose Manuel Martí, Tijana Radivojevic, Sai Vamshi R. Jonnalagadda, Reinhard Gentz, Nathan J. Hillson, Sean Peisert et al. "Machine learning for metabolic engineering: A review." *Metabolic Engineering* 63 (2021): 34-60.
8. Leiserson, Charles E., Neil C. Thompson, Joel S. Emer, Bradley C. Kuszmaul, Butler W. Lampson, Daniel Sanchez, and Tao B. Schardl. "There's plenty of room at the Top: What will drive computer performance after Moore's law?." *Science* 368, no. 6495 (2020): eaam9744.
9. Shalf, John. "The future of computing beyond Moore's Law." *Philosophical Transactions of the Royal Society A* 378, no. 2166 (2020): 20190061.
10. In, Kwang-Ho, Yong-Soo Kwon, In-Jae Oh, Kyu-Sik Kim, Maan-Hong Jung, Kwan-Ho Lee, Sun-Young Kim et al. "Lung cancer patients who are asymptomatic at diagnosis show favorable prognosis: a Korean Lung Cancer Registry Study." *Lung cancer* 64, no. 2 (2009): 232-237.
11. Quadrelli, Silvia, G. Lyons, H. Colt, D. Chimondeguy, and A. Buerro. "Clinical characteristics and prognosis of incidentally detected lung cancers." *International journal of surgical oncology* 2015 (2015).
12. de Koning, Harry J., Carlijn M. van der Aalst, Pim A. de Jong, Ernst T. Scholten, Kristiaan Nackaerts, Marjolein A. Heuvelmans, Jan-Willem J. Lammers et al. "Reduced lung-cancer mortality with volume CT screening in a randomized trial." *New England journal of medicine* 382, no. 6 (2020): 503-513.
13. Baker, Stephen R., Ronak H. Patel, Lily Yang, Valdis M. Lelkes, and Alejandro Castro III. "Malpractice suits in chest radiology: an evaluation of the histories of 8265 radiologists." *Journal of thoracic imaging* 28, no. 6 (2013): 388-391.
14. Baker, Stephen R., Ronak H. Patel, Lily Yang, Valdis M. Lelkes, and Alejandro Castro III. "Malpractice suits in chest radiology: an evaluation of the histories of 8265 radiologists." *Journal of thoracic imaging* 28, no. 6 (2013): 388-391.
15. Lodwick, Gwilym S., Theodore E. Keats, and John P. Dorst. "The coding of roentgen images for computer analysis as applied to lung cancer." *Radiology* 81, no. 2 (1963): 185-200.
16. Munir, Khushboo, Hassan Elahi, Afsheen Ayub, Fabrizio Frezza, and Antonello Rizzi. "Cancer diagnosis using deep learning: a bibliographic review." *Cancers* 11, no. 9 (2019): 1235.
17. Jaeger, Stefan, Sema Candemir, Sameer Antani, Yi-Xiang J. Wang, Pu-Xuan Lu, and George Thoma. "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases." *Quantitative imaging in medicine and surgery* 4, no. 6 (2014): 475.
18. Nguyen, Hoang C., Tung T. Le, Hieu H. Pham, and Ha Q. Nguyen. "VinDr-RibCXR: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest X-rays." *arXiv preprint arXiv:2107.01327* (2021).
19. Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).
20. Kim, Young-Gon, Yongwon Cho, Chen-Jiang Wu, Sejin Park, Kyu-Hwan Jung, Joon Beom Seo, Hyun Joo Lee, Hye Jeon Hwang, Sang Min Lee, and Namkug Kim. "Short-term reproducibility of pulmonary nodule and mass detection in chest radiographs: Comparison among radiologists and four different computer-aided detections with convolutional neural net." *Scientific Reports* 9, no. 1 (2019): 18738.
21. Kim, Ji Hoon, Sang Gil Han, Ara Cho, Hye Jung Shin, and Song-Ee Baek. "Effect of deep learning-based assistive technology use on chest radiograph interpretation by emergency department physicians: a prospective interventional simulation-based study." *BMC Medical Informatics and Decision Making* 21 (2021): 1-9.
22. National Lung Screening Trial Research Team. "Reduced lung-cancer mortality with low-dose computed tomographic screening." *New England Journal of Medicine* 365, no. 5 (2011): 395-409.
23. de Koning, Harry J., Carlijn M. van der Aalst, Pim A. de Jong, Ernst T. Scholten, Kristiaan Nackaerts, Marjolein A. Heuvelmans, Jan-Willem J. Lammers et al. "Reduced lung-cancer mortality with volume CT screening in a randomized trial." *New England journal of medicine* 382, no. 6 (2020): 503-513.

24. Sesen, M. Berkan, Ann E. Nicholson, Rene Banares-Alcantara, Timor Kadir, and Michael Brady. "Bayesian networks for clinical decision support in lung cancer care." *PLoS one* 8, no. 12 (2013): e82349.
25. Huang, Chi-Hsiang, Chian Zeng, Yi-Chia Wang, Hsin-Yi Peng, Chia-Sheng Lin, Che-Jui Chang, and Hsiao-Yu Yang. "A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer." *Sensors* 18, no. 9 (2018): 2845.
26. Kort, Sharina, Marjolein Brusse-Keizer, Jan Willem Gerritsen, Hugo Schouwink, Emanuel Citgez, Frans de Jongh, Jan van der Maten, Suzy Samii, Marco van den Bogart, and Job van der Palen. "Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters." *ERJ open research* 6, no. 1 (2020).
27. Kort, Sharina, Marjolein Brusse-Keizer, Jan Willem Gerritsen, Hugo Schouwink, Emanuel Citgez, Frans de Jongh, Jan van der Maten, Suzy Samii, Marco van den Bogart, and Job van der Palen. "Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters." *ERJ open research* 6, no. 1 (2020).
28. Sun, Yingli, Cheng Li, Liang Jin, Pan Gao, Wei Zhao, Weiling Ma, Mingyu Tan et al. "Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction." *European Radiology* 30 (2020): 3650-3659.
29. Girard, Luc, Sabine Zochbauer-Muller, Arvind K. Virmani, Adi F. Gazdar, and John D. Minna. "Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering." *Cancer research* 60, no. 17 (2000): 4894-4906.
30. Shen, Ronglai, Adam B. Olshen, and Marc Ladanyi. "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis." *Bioinformatics* 25, no. 22 (2009): 2906-2912.
31. Lynch, Chip M., Victor H. van Berkel, and Hermann B. Frieboes. "Application of unsupervised analysis techniques to lung cancer patient data." *PLoS One* 12, no. 9 (2017): e0184370.
32. Eckardt, Jan-Niklas, Martin Bornhäuser, Karsten Wendt, and Jan Moritz Middeke. "Semi-supervised learning in cancer diagnostics." *Frontiers in oncology* 12 (2022): 960984.
33. Nguyen, Phuong, David Chapman, Sumeet Menon, Michael Morris, and Yelena Yesha. "Active semi-supervised expectation maximization learning for lung cancer detection from Computerized Tomography (CT) images with minimally label training data." In *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314, pp. 553-564. SPIE, 2020.
34. Shak, Kelvin, Mundher Al-Shabi, Andrea Liew, Boon Leong Lan, Wai Yee Chan, Kwan Hoong Ng, and Maxine Tan. "A new semi-supervised self-training method for lung cancer prediction." *arXiv preprint arXiv:2012.09472* (2020).
35. Zhao, Yufan, Donglin Zeng, Mark A. Socinski, and Michael R. Kosorok. "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer." *Biometrics* 67, no. 4 (2011): 1422-1433.
36. Tseng, Huan-Hsin, Yi Luo, Sunan Cui, Jen-Tzung Chien, Randall K. Ten Haken, and Issam El Naqa. "Deep reinforcement learning for automated radiation adaptation in lung cancer." *Medical physics* 44, no. 12 (2017): 6690-6705.
37. Wang, Hanlin, Xue Bai, Yajuan Wang, Yanfei Lu, and Binbing Wang. "An integrated solution of deep reinforcement learning for automatic IMRT treatment planning in non-small-cell lung cancer." *Frontiers in Oncology* 13 (2023): 1124458.
38. Alsinglawi, Belal, Osama Alshari, Mohammed Alorjani, Omar Mubin, Fady Alnajjar, Mauricio Nova, and Omar Darwish. "An explainable machine learning framework for lung cancer hospital length of stay prediction." *Scientific reports* 12, no. 1 (2022): 607.
39. Venkatesh, Surraj Kumar Prabhu, and Lilly Raamesh. "Predicting lung cancer survivability: A machine learning ensemble method on seer data." (2022).
40. Vikas, Prabhpreet Kaur. "Lung cancer detection using chi-square feature selection and support vector machine algorithm." *International Journal of Advanced Trends in Computer Science and Engineering* (2021).
41. Chauhan, Anamika. "Detection of lung cancer using machine learning techniques based on routine blood indices." In *2020 IEEE international conference for innovation in technology (INOCON)*, pp. 1-6. IEEE, 2020.

42. Sim, Jin-ah, Young Ae Kim, Ju Han Kim, Jong Mog Lee, Moon Soo Kim, Young Mog Shim, Jae Ill Zo, and Young Ho Yun. "The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning." *Scientific reports* 10, no. 1 (2020): 10693.
43. Patra, Radhanath. "Prediction of lung cancer using machine learning classifier." In *Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1*, pp. 132-142. Springer Singapore, 2020.
44. Radhika, P. R., Rakhi AS Nair, and G. Veena. "A comparative study of lung cancer detection using machine learning algorithms." In *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*, pp. 1-4. IEEE, 2019.
45. Wu, Jiangpeng, Xiangyi Zan, Liping Gao, Jianhong Zhao, Jing Fan, Hengxue Shi, Yixin Wan, E. Yu, Shuyan Li, and Xiaodong Xie. "A machine learning method for identifying lung cancer based on routine blood indices: qualitative feasibility study." *JMIR medical informatics* 7, no. 3 (2019): e13476.
46. Faisal, Muhammad Imran, Saba Bashir, Zain Sikandar Khan, and Farhan Hassan Khan. "An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer." In *2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST)*, pp. 1-4. IEEE, 2018.
47. Safiyari, Ali, and Reza Javidan. "Predicting lung cancer survivability using ensemble learning methods." In *2017 intelligent systems conference (IntelliSys)*, pp. 684-688. IEEE, 2017.
48. Jochems, Arthur, Timo M. Deist, Johan Van Soest, Michael Eble, Paul Bulens, Philippe Coucke, Wim Dries, Philippe Lambin, and Andre Dekker. "Distributed learning: developing a predictive model based on data from multiple hospitals without data leaving the hospital—a real life proof of concept." *Radiotherapy and Oncology* 121, no. 3 (2016): 459-467.
49. Jochems, Arthur, Timo M. Deist, Issam El Naqa, Marc Kessler, Chuck Mayo, Jackson Reeves, Shruti Jolly et al. "Developing and validating a survival prediction model for NSCLC patients through distributed learning across 3 countries." *International Journal of Radiation Oncology* Biology* Physics* 99, no. 2 (2017): 344-352.
50. Wang, Debby D., Weiqiang Zhou, Hong Yan, Maria Wong, and Victor Lee. "Personalized prediction of EGFR mutation-induced drug resistance in lung cancer." *Scientific reports* 3, no. 1 (2013): 2855.
51. Abels, Esther, and Liron Pantanowitz. "Current state of the regulatory trajectory for whole slide imaging devices in the USA." *Journal of pathology informatics* 8, no. 1 (2017): 23.
52. Wichmann, Julian L., Martin J. Willeminck, and Carlo N. De Cecco. "Artificial intelligence and machine learning in radiology: current state and considerations for routine clinical implementation." *Investigative Radiology* 55, no. 9 (2020): 619-627.