HIGH SCHOOL EDITION
Journal of Student Research

Volume 12 Issue 2 (2023)

How Artificial Intelligence Can Aid Screening and Detecting Lung Cancer in Lung Cancer Patients

Ronak Ramesh[1], Simran Saluja[#], Rajagopal Appavu[#] and Jothsna Kethar[#]

[1]The Academy of Aerospace and Engineering
[#]Advisor

ABSTRACT

In modern society, lung cancer stands as one of the most dangerous and commonly diagnosed cancers in North America. In recent years, lung cancer has caused more people to die than any other cancer. Most diseases like cancer get worse as time goes on, so the best way to combat it is to identify the disease in its early stages. However, lung cancer shares symptoms with many other diseases like COVID-19, asthma, and pneumonia which leads to doctors incorrectly diagnosing their lung cancer patients. If left unchecked, this catastrophic crisis will only lead to more deaths. Previous research studies have investigated if using Artificial Intelligence in Lung Cancer screening and treatment diagnosis would be successful but have yet to make a conclusive verdict. This paper has put together knowledge about lung cancer detection and artificial intelligence algorithms to determine whether AI can be accurately implemented to detect lung cancer.

Introduction

Lung cancer is one of the most commonly diagnosed cancers in North America and by far the deadliest cancer as it is the leading cause of cancer death. In fact, each year more people die from lung cancer than from breast, prostate, and colon cancers combined (*What Is Lung Cancer? | Types of Lung Cancer*, n.d.). Lung cancer starts off in the lungs and can spread to other organs in a process called metastasis. Previous research studies have investigated if using Artificial Intelligence in Lung Cancer screening and treatment diagnosis would be successful since they seem to have an improvement in the detection of biomarkers (term used to describe abnormal cells of a disease). Although these exploratory investigations have shown potential value, rigorous validation and standardization are necessary before AI may be used in clinical decision-making. I will go over modern strategies of implementing AI in Lung Cancer screening and I'll also talk about the difficulties and restrictions that need to be overcome before these AI prediction models are used in healthcare settings.

What is Lung Cancer?

To put it simply, cancer is a disease where the cells of a human grow abnormally and uncontrollably while spreading across the body. Lung cancer is the growth of these abnormal cells starting in one or both of the lungs (*What Is Lung Cancer? | Types of Lung Cancer*, n.d.). We categorize these abnormal cells or growths that form in our lungs as lung nodules or pulmonary nodules. When you breathe in, air enters either your mouth or your nose and travels into your lungs through the trachea. The trachea is split into bronchi, which are tubes that divide and connect to other smaller bronchi in the lungs. These bronchi split to form smaller branches called bronchioles, which have tiny air sacs called alveoli at the ends of them. The alveoli's main function is to take in the oxygen from the inhaled air and supply it to the blood while also removing the carbon dioxide from your blood when you exhale. The cells lining the bronchi and other areas of the lung, such as the bronchioles or alveoli, are where lung cancers generally begin (*What Is Lung Cancer? | Types of Lung Cancer*, n.d.).

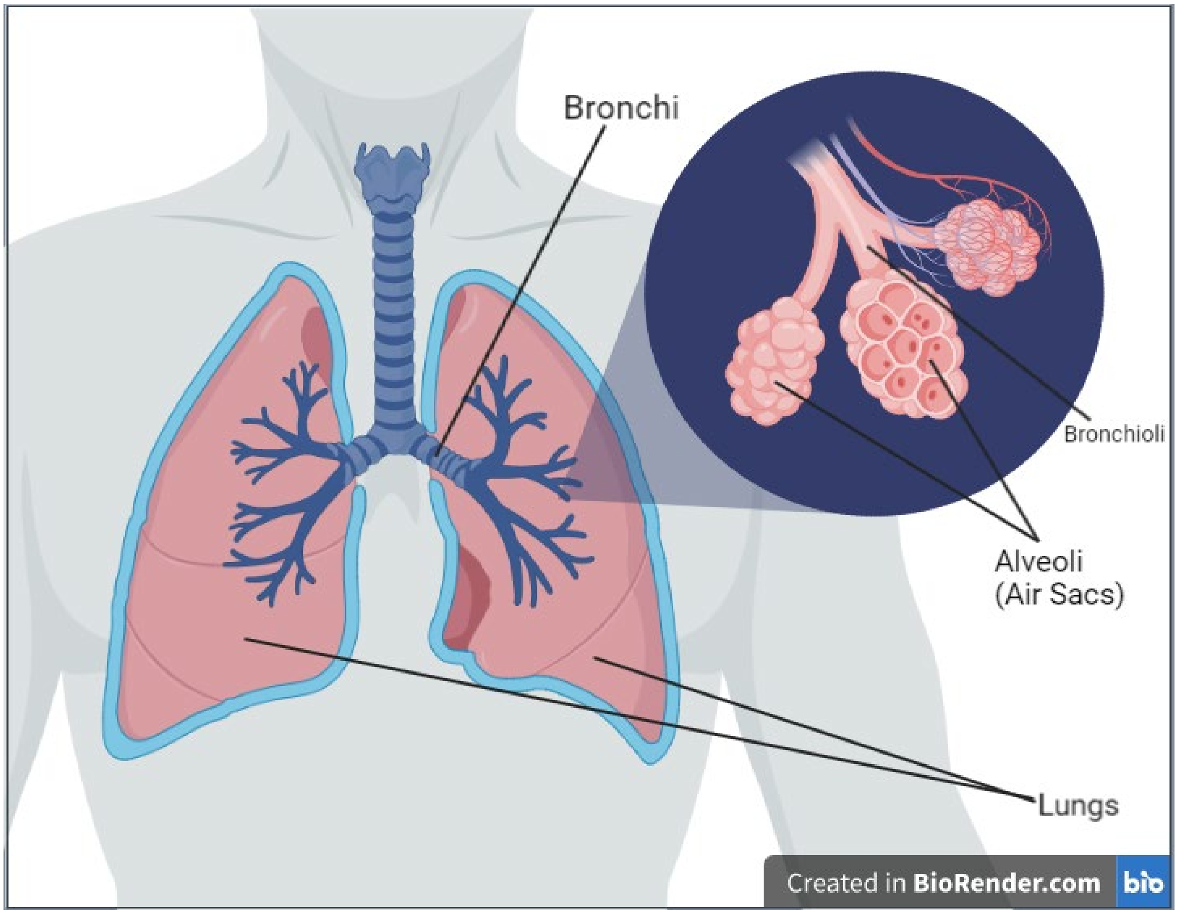

Figure 1. This illustration of the respiratory system shows the lungs, bronchi, bronchiole, and alveoli. Created and copyrighted by Ronak Ramesh

There are two main types of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC covers close to 85% of all lung cancers while the rest is SCLC. Compared to NSCLC, this kind of lung cancer tends to develop and spread more quickly. In fact, when diagnosed with SCLC, almost 70% of patients already have the disease disseminated throughout their bodies(*Lung Cancer Signs & Symptoms | Common Symptoms of Lung Cancer*, n.d.). One of the major causes of lung cancer is cigarette smoking which is linked to approximately 80-90% of lung cancer deaths. In fact, the risk of developing lung cancer or passing away from it is 15–30 times higher in smokers than in nonsmokers. Other tobacco products are also known to damage your lungs and lead to lung cancer. It is extremely rare for lung cancer to be hereditary (*What Are the Risk Factors for Lung Cancer? | CDC*, n.d.). Another way someone can get lung cancer is through prolonged exposure to a gas called Radon. Radon is a natural gas often formed in rocks, soil, and water. Radon is hard for an average human with no equipment to detect because it cannot be seen, tasted, nor smelled (*What Are the Risk Factors for Lung Cancer? | CDC*, n.d.). Radon can sneak into homes through cracks and once it is trapped inside, the increased exposure can cause lung cancer. According to the U.S. Environmental Protection Agency (EPA), radon causes about 21,000 lung cancer deaths each year and one in every 15 homes have high radon exposure (*What Are the Risk Factors for Lung Cancer? | CDC*, n.d.). Radon is so dangerous that it is the second leading cause of cancer, right after cigarette smoking. Symptoms of lung cancer include chest discomfort, shortness of breath, wheezing, coughing up blood, feeling extremely exhausted all the time, and weight loss with no apparent cause. One of the most dangerous aspects of lung cancer is that it is often diagnosed far too late. Most cases of lung cancer show symptoms after the cancer has spread throughout the body, making diagnosis and treatment even more difficult. Furthermore, the symptoms shown above are also shown in numerous other diseases such as asthma, allergies, pneumonia, COVID-19, heart disease, bronchitis, and bronchiolitis.

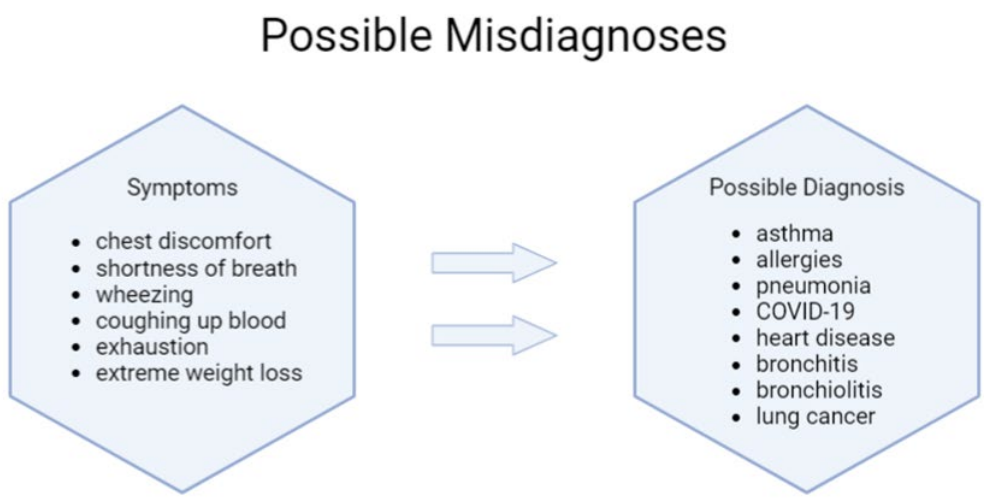

Figure 2. Possible misdiagnosis of patient symptoms. Created and copyrighted by Ronak Ramesh.

Since these symptoms are common in so many other diseases, doctors often misdiagnose patients and only find lung cancer in the later stages. The later a cancer is diagnosed, the more difficult it is to treat it.

AI Literature Review

Artificial intelligence is simply the usage of computer systems to perform tasks that would otherwise require human intelligence. It simulates human intelligence through computer programs and algorithms machines. Artificial intelligence is so important because it can reduce the number of human errors made in the real world. Especially in the medical field, when using computers and machines, we can be more precise when going through processes like diagnosis and screening. For the software to learn automatically from data feeds and data patterns or characteristics in the data, artificial intelligence (AI) combines massive volumes of data with quick, iterative processing and sophisticated algorithms in a process called deep learning, a subset of a process named machine learning (Pessach, D., & Shmueli, E., 2020). Deep-learning algorithms understand what a tumor is from real-world instances rather than searching for tumor traits that a programmer has predefined in advance. Researchers provide the systems with a sizable data set made up of the lung CT scans of thousands of patients, with some cancer-specific data and some with non-cancer data, some of whom had cancer and others who did not. This lets the machines understand what a lung cancer nodule looks like for themselves. The more data we feed to the machines, the more accurate the predictions it makes will become. Meaning, without large amounts of data provided to the AI machine, it will often make inaccurate predictions. This would be detrimental to any program that incorporated AI since the algorithm would be making incorrect choices in the real world. An artificial intelligence algorithm has to be educated with relevant data to establish a logical process for deciding before it can be utilized to make actual judgments (Pessach, D., & Shmueli, E., 2020). Issues that are observed as a result of an algorithm's choice are frequently the result of mistakes made during the algorithm's training (Pessach & Shmueli, 2020).

Although AI algorithms are present in many aspects of our life, most individuals are not immediately aware of their influence. For instance, AI has found considerable utility in the fields of marketing, product creation, and more (Pessach, D., & Shmueli, E., 2020). In terms of the field of biology, radiomics is a method of study that uses quantitative data derived from medical pictures along with machine learning for applications in staging, therapy selection, and response evaluation. In other words, radiomics is the method of the extracting of data from medical imaging. Segmentation, feature extraction, and machine learning modeling can all be used to accomplish this. Initial segmentation of the area of interest, typically a tumor, is followed by analysis to derive numerical information known as handcrafted characteristics. In order to build prediction models, these features (which are often made up of first-order, texture, size, and shape features) can be combined with data from genetics, histology, blood biomarkers, and patient demographics.

How Lung Cancer is Currently Screened

Modern day physicians may request diagnostic tests, such as the following, to get a precise diagnosis.

1. Imaging testing: To assist identify the source of your symptoms, imaging tests that take images of your lungs include chest X-rays, computed tomography (CT) scans, magnetic resonance imaging (MRI) scans, and/or PET scans.
2. Lab procedures/tests: Lung cells can be examined to assist make a precise diagnosis. The cells may be extracted from fluid that has accumulated around the lungs or from mucus that you cough up.
3. Biopsy: A tiny sample of the nodule will be removed with a needle or during surgery in order to conduct a biopsy and assess if it is malignant if imaging tests reveal a nodule (or mass) in your lungs. This tiny sample is then further tested in laboratories to draw information about the overall nodule.
4. Breathing test: When patients undergo breathing problems, doctors may recommend a breath test, also known as a pulmonary function test. These tests help show how well a patient's lungs can breathe in and out. Apart from lung cancer, these tests can be used to diagnose many other diseases.

CXR

CXRs or Chest X-Rays are one of the most often utilized imaging modalities in medicine. CXRs provide an x-ray of the anatomy inside a specific person's chest. CXRs have been used to give excellent assessments of the patients' thorax. In fact, even with 0.1 mSv radiation exposure, which is equivalent to 10 days of natural background radiation, a CXR was still able to provide an accurate representation of the patient's thorax.
Traditionally, radiologists would analyze the x-ray results and diagnose patients based on their interpretation. However, in the modern age, computers can immediately examine photographs by analyzing image attributes such as form, size, intensity, and texture (Luo et al., 2022).

In fact, radiomics broadens the definition of image attributes from a computer standpoint by calculating the picture pixel-by-pixel. In the picture, the region of interest area may be translated to higher dimension data and expressed as a massive matrix by calculating the picture texture and density using various mathematical approaches. Because radiomics is based on arithmetic, the diverse picture characteristics of CXR provide the computer with other work. Image augmentation, comprising pre-processing, lung segmentation, and rib suppression, is a critical method before nodule diagnosis to acquire correct radiomics data. To make the process of analyzing these pictures even more efficient, deep-learning-based algorithms were employed and the results were astounding. It was found that CheXNet, a radiologist-level deep learning system trained on Chest-X Ray 14, one of the world's biggest CXR databases, outperforms radiologist performance in the diagnosis of 14 pulmonary illnesses, including lung nodules and lung masses, with AUROC (Area Under the Receiver Operating Characteristics) and AUC (Area Under the Curve) values of 0.78 and 0.87, respectively.

Chest CT

CT technology allows for noninvasive exploration of the thorax's three-dimensional structure.

In two major, randomized control studies, screening with LDCT reduced mortality by around 20%: the National Lung Screening Trial (NLST) and the Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON study).

AI in LDCT reading was made with intentions to assist radiologists

1. reduce arduous work,
2. reduce reader variability, and
3. enhance screening efficiency.

The primary objective for AI in LDCT reading is the same as it is in CXR: nodule identification and classification/malignancy prediction. Many researchers have employed artificial intelligence systems to detect lung nodules in chest CT scans. It was difficult to evaluate the models scientifically since they employed various models on different datasets and evaluated the models using multiple criteria such as sensitivity, specificity, AUC, and accuracy. All these different subsets of information also made it more likely for humans to make silly interpretation errors. In modern times, not only has this AI implementation shown to assist radiologists in increasing the sensitivity of nodule identification, but it has also decreased interpretation time. Furthermore, AI has shown to be an excellent concurrent or second reader (Conant et al., 2019). It was concluded that AI wasn't yet ready to take over the primary reader spot, in case radiologists weren't able to find the missed nodules from the AI's diagnosis.

In the process of identifying critical nodules, lung nodule classification and malignancy prediction are important because they help radiologists and artificial intelligence prioritize nodules by harmfulness. Nodules are characterized based on their texture, which might be solid, part-solid, or non-solid, as well as their size. Research shows that an AI model trained on the MILD trial dataset and externally verified on the Danish Lung Cancer Screening Trial (DLCST) has proven that AI performed equally to human experts on differential six textures(sold, part-solid, non-solid, calcified, perifissural, and speculated) (Chiu et al., 2022). The artificial intelligence categorization was then utilized to forecast the likelihood of malignancy, as advised by the Lung CT Screening Reporting and Data System (LUNG-RADS) (Chiu et al., 2022).

Novel Screening Tests

Some novel screening tools used for lung cancer are genomics and proteomics. Although both of these screening approaches provide a high number of signals for each patient, a sophisticated algorithm that can build off of the high yield of information would improve diagnostic usefulness.

One of the most popular areas in oncology is genomics. Scientists can now study the full genome, exome, transcriptome, and epigenome of malignancies and create massive sets of information on patients and their tumors using the polymerase chain reaction (PCR) amplification technique and associated technology. Machine learning algorithms can effectively distinguish cancer-free healthy controls from cancer patients by evaluating whole-genome and whole-transcriptome sequencing data from treatment-naive patients in The Cancer Genome Atlas (TCGA).

Proteomics is known as the study of proteomes which are proteins produced in a specific biological context. Proteomics are so important because it can find biomarkers in the human body by analyzing the proteins in the body. Proteins and other metabolites obtained from plasma and urine samples are extremely simple to collect and have been explored for decades as a screening tool for lung cancer. To deal with the enormous number of variables generated by proteomics, researchers employed machine learning approaches to decrease dimensionality and select features. The most accurate forecasts, however, were produced using less interpretable models. Following on from this concept, urine proteome analysis combined with machine learning analysis effectively created models that can distinguish lung cancer samples not just from healthy ones, but also from samples from different malignancies (Zhang et al., 2018). These methods show how productive and useful algorithms can be to these screening methods when incorporated correctly.

Incorporating Artificial Intelligence in Lung Cancer Screening

In efforts of producing accurate diagnoses, several screening methods have been tried, including imaging, sputum cytology, blood test screening, and breath testing. However, only picture screening can give the necessary insights. Doctors have been using chest X-rays (CXRs) as a commonly utilized clinical practice to screen lung cancer for a very long time. The only technology that has been proved to identify lung cancer sooner and extend lung cancer patients' lives is computed tomography (LDCT). However, there are also instances of diagnosing errors in interpreting CXR or LDCT pictures. This may be due to human error as human eyes grow weary and visuals begin to blur after reading images for an extended period of time. Since these inaccurate interpretations harmed patient lives, they also resulted in a substantial number of malpractice lawsuits. Furthermore, research shows that even though specialists were found to detect more pulmonary nodules on CXRs, radiologists missed roughly 20% of lung nodules that are less than 3 cm. With computer-aided diagnostic methods and AI-based algorithms that have been designed in the twenty-first century, the prediction accuracy of pulmonary nodules on CXRs has been substantially increased. Digital imaging and communications in medicine (DICOM) is the industry standard for picture restoration and transfer to allow communication between servers, manufacturers, and hospitals. In a format governed by the Medical Imaging and Technology Alliance, a branch of the National Electrical Manufacturers Association, the DICOM not only stores image file pixel data but also a patient identification number, picture type, machine-related characteristics, and other information (Masood et al., 2018).

Humans cannot always be 100% accurate because we have many limitations. For example, after hours of reviewing lung cancer screening data, the human eye can become tired and images can seem blurry for scientists. This can lead to false interpretations of the screening causing catastrophic effects for the patient. By incorporating artificial intelligence into lung cancer screening, we can prevent most, if not all, human errors. Research has shown that approximately 7% of patients who were diagnosed with lung cancer are asymptomatic, and more than half of the patients who had lung cancer resection were asymptomatic (Chiu et al., 2022). This means that while 7% of patients had lung cancer, they were not affected by it at all. When incorporating AI into lung cancer screenings, scientists hope to not only make more accurate conclusions but also create helpful categories. Scientists hope that AI algorithms could sort out the results from Lung Cancer screening and categorize them by priority. For example: emergency level, abnormal level and normal level (Armitage, 2018).

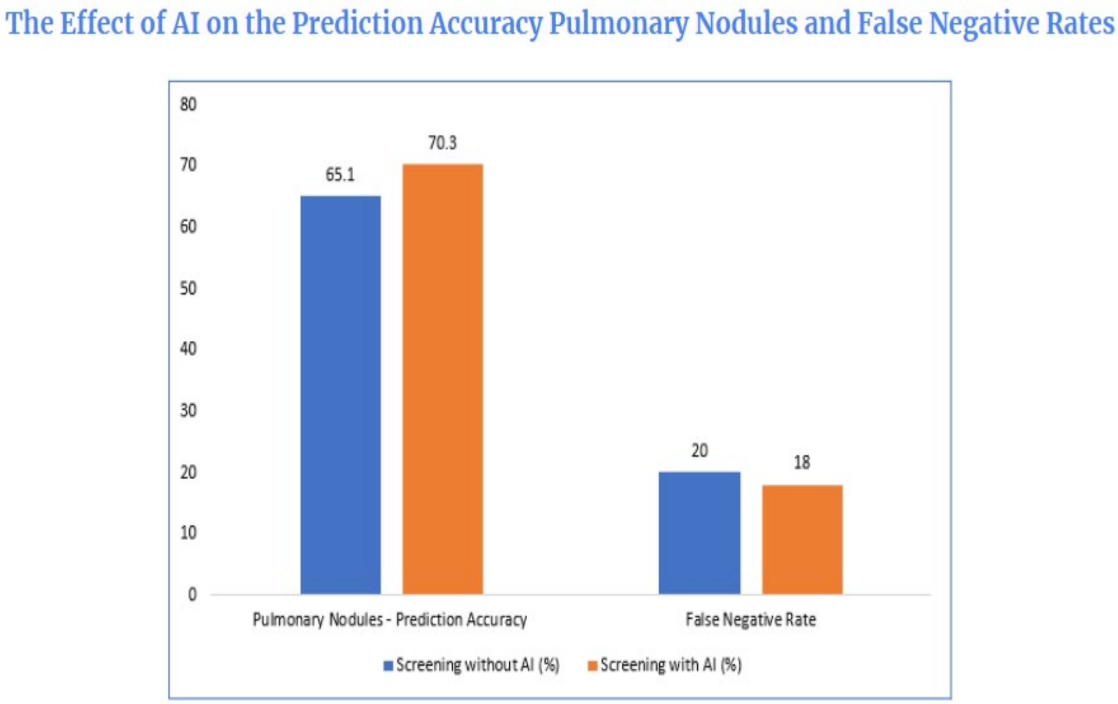

Figure 3. This graph clearly shows how using artificial intelligence can decrease false negative rates and increase the prediction and identification accuracy of pulmonary nodules. This is yet another piece of evidence that supports the fact that Artificial Intelligence can be of great use for Lung Cancer detection. Created and copyrighted by Ronak Ramesh.

Discussion

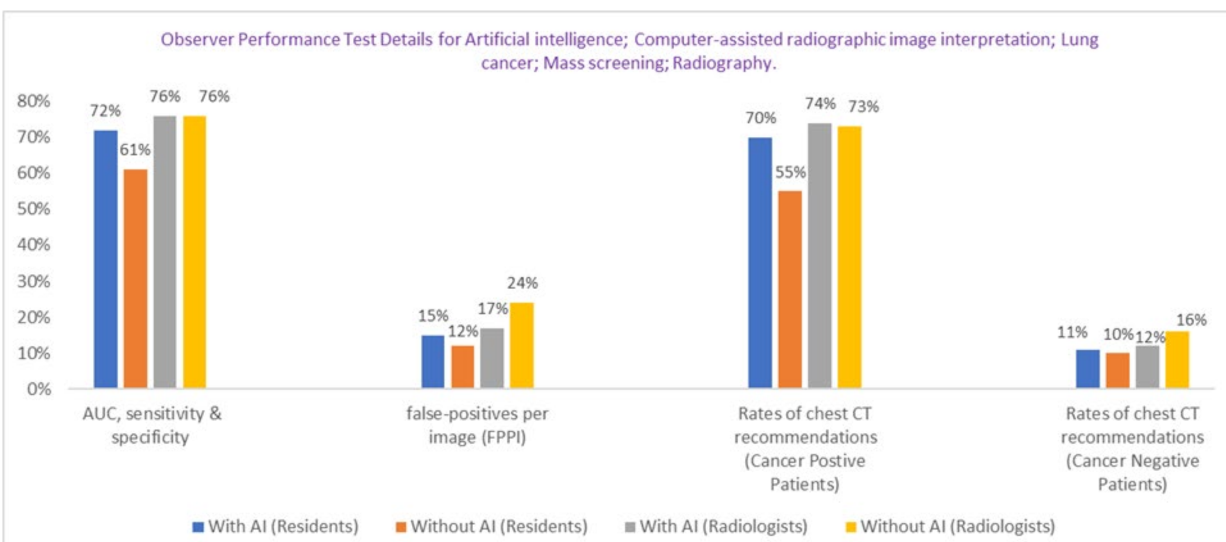

Figure 4: Performance results from a research study conducted on detection of Lung Cancer Nodules using Artificial Intelligence Model (Yoo et al., 2021).

The experiment whose results are displayed in Figure 4, involves 3 radiologist residents, 5 Board-Certified Radiologists, 173 images from cancer-positive patients and 346 images from cancer-negative patients (provided by the National Lung Screening Trial). The residents and radiologists were tested on their ability to detect visible lung cancer both with and without AI. In the bottom table, it compares the sensitivity rates, false-positive rates, rates of chest computed tomography recommendations for patients with cancer, and rates of chest computed tomography recommendations for patients with cancer from the residents and radiologists both with/without the aid of Artificial Intelligence (Yoo et al., 2021). Created and copyrighted by Ronak Ramesh.

As seen in Figure 4, while using Artificial Intelligence, the radiologist residents were able to increase their average sensitivity for the detection of visible lung cancer when compared to when they were not using AI (72% vs 61%). On the other hand, board certified radiologists' average sensitivity rates for detecting visible lung cancer with and without AI were roughly the same both with and without Artificial Intelligence (76%). This might be due to the heavy experience board-certified radiologists have. Nonetheless, it shows how accurate using Artificial intelligence is, since residents are clearly much more accurate in identifying visible lung cancer with Artificial Intelligence aiding them. Another conclusion we can draw from Figure 4 is that the number of false-positive results was lower for radiologists when using Artificial Intelligence (17% with AI vs 24% without AI). This proves that the aid of Artificial Intelligence is useful because with the help of AI, even the most experienced radiologists are able to reduce the number of false-positive results they identify. The average rate of Computed Tomography recommendation in patients positive for cancer significantly rose for residents using AI vs without AI, (70% vs 55%). In cancer-free patients, the average rate of Computed Tomography recommendation was lower for radiologists with the aid of AI (12% with the aid of AI vs 16% without AI). Once again, all this evidence proves that the aid of Artificial Intelligence would be invaluable in the process of detecting lung cancer because medical professionals are able to be so much more precise with the use of AI. To summarize, we can conclude that when used as a primary approach, AI algorithms can improve readers' performance in detecting lung cancers on chest radiographs.

Conclusion

In modern society, scientists are striving to improve the accuracy and precision of lung cancer diagnoses. This is of utmost importance, because lung cancer (like all other cancers) is harder to treat the later it is found. One idea is to use computer algorithms and artificial intelligence to assist doctors in detecting lung cancer symptoms. This would theoretically reduce the number of errors made during lung cancer screenings. However, it is not widely used in our society because although initial data and studies have shown that AI has potential, rigorous validation and standardization are necessary before AI may be used in clinical decision-making. This paper delved deeper into this problem by taking a dataset from the National Lung Screening Trial (NLST) and analyzing the results to see whether or not AI can provide a significant difference in lung cancer diagnoses. The analysis and discussion on the NLST dataset found that the AI algorithm improved sensitivity for residents and specificity for radiologists in the diagnosis of lung cancer. Furthermore, with AI, radiology residents were able to prescribe more chest CT tests for patients with visible lung cancer, which means that they were able to identify more lung cancer symptoms/signs. In cancer-negative patients, radiologists advised considerably fewer unnecessary chest CT exams with AI. With all this evidence, we can confidently say that AI assistance can indeed improve the accuracy and quality of lung cancer detection earlier.

Acknowledgments

I would like to thank Coach Jo, Mr. Rajgopal, and Simran Saluja for guiding me while writing this paper and for helping me polish it to the best it could be.

References

Chiu, H. Y., Chao, H. S., & Chen, Y. M. (2022). Application of Artificial Intelligence in Lung Cancer. *Cancers*, *14*(6), 1370. https://doi.org/10.3390/cancers14061370

Lung Cancer Signs & Symptoms | Common Symptoms of Lung Cancer. (n.d.). Retrieved October 25, 2022, from https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/signs-sy mptom/signs-symptoms.html

What Are the Risk Factors for Lung Cancer? | CDC. (n.d.). Retrieved October 25, 2022, from https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm

What Is Lung Cancer? | Types of Lung Cancer. (n.d.). Retrieved October 25, 2022, from https://www.cancer.org/cancer/lung-cancer/about/what-is.html

Pessach, D., & Shmueli, E. (2020). Algorithmic fairness. arXiv preprint arXiv:2001.09784. https://doi.org/10.48550/arXiv.2001.09784.

Yoo, H., Lee, S. H., Arru, C. D., Doda Khera, R., Singh, R., Siebert, S., Kim, D., Lee, Y., Park, J. H., Eom, H. J., Digumarthy, S. R., & Kalra, M. K. (2021). AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset. *European Radiology*, *31*(12), 9664–9674. https://doi.org/10.1007/s00330-021-08074-7

Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D. (2018). Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. *Journal of Biomedical Informatics*, *79*, 117–128. https://doi.org/10.1016/j.jbi.2018.01.005

Armitage, H. (2018, November 20). Artificial intelligence rivals radiologists in screening X-rays for certain diseases. Retrieved October 25, 2022, from https://med.stanford.edu/news/all-news/2018/11/ai-outperformed-radiologists-in-screening-x-rays-for-certain-diseases.html

Conant, E. F., Toledano, A. Y., Periaswamy, S., Fotin, S. V., Go, J., Boatsman, J. E., & Hoffmeister, J. W. (2019). Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis. *Radiology. Artificial intelligence*, *1*(4), e180096. https://doi.org/10.1148/ryai.2019180096

Luo, L., Chen, H., Xiao, Y., Zhou, Y., Wang, X., Vardhanabhuti, V., Wu, M., Han, C., Liu, Z., Fang, X. H. B., Tsougenis, E., Lin, H., & Heng, P. A. (2022). Rethinking Annotation Granularity for Overcoming Shortcuts in Deep Learning-based Radiograph Diagnosis: A Multicenter Study. Radiology. Artificial intelligence, 4(5), e210299. https://doi.org/10.1148/ryai.210299

Zhang, C., Leng, W., Sun, C., Lu, T., Chen, Z., Men, X., Wang, Y., Wang, G., Zhen, B., & Qin, J. (2018). Urine Proteome Profiling Predicts Lung Cancer from Control Cases and Other Tumors. *EBioMedicine*, *30*, 120–128. https://doi.org/10.1016/j.ebiom.2018.03.009