Investigating the Effects of Algorithmic Bias in Lung Cancer Diagnosis and Treatment

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Preethi Chidambaram

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

MC Forelle, Department of Engineering and Society

Introduction

Lung cancer is the leading cause of cancer deaths worldwide in both men and women, with an overall 5-year survival rate that is less than 20% (Christie et al., 2021). Mortality rates become increasingly high in later stages of lung cancer, which makes screening and early detection an integral part of patient survival. In fact, the survival rate of patients is so low in most countries because patients are only diagnosed in later stages, which limits the treatment methods that can be used (Pei et al., 2022). Machine learning based prediction models have become increasingly prevalent in this field of medicine to improve cancer detection and treatment. However, current research on this topic focuses largely on the benefits these systems bring to cancer care and fails to discuss major design flaws that may have serious implications for the overall healthcare system.

Artificial intelligence, or AI, is an algorithm that is used to recognize patterns between various inputs and outputs and make decisions based on these patterns when given unseen inputs (Z.-H. Chen et al., 2021). AI is being used in cancer care to interpret medical images and provide doctors with patient specific diagnostic information which can help create personalized treatment plans. Machine learning (ML) is a subfield of AI that focuses on training algorithms to learn information based on past behavior. Furthermore, deep learning, which is a subset of ML, allows for unsupervised learning amongst these algorithms (Z.-H. Chen et al., 2021). Medical professionals are utilizing AI and its subfields to combat lung cancer by increasing the accuracy of detecting pulmonary nodules, which are abnormal areas of growth within lungs (Pei et al., 2022). One of the main concerns regarding the use of AI in any field is the possibility of algorithmic bias that may reproduce existing social disparities because of the people developing the algorithms or the model's training data. With medical AI systems specifically, it is imperative that these algorithms are adequately trained and tested on datasets representative of global populations before being deployed in clinical settings.

This paper will provide an overview of the applications of AI in lung cancer treatment, with emphasis on the implications of algorithmic bias in this specific sector of healthcare. Research indicates that the use of AI systems in lung cancer diagnosis and treatment can potentially exacerbate existing socioeconomic disparities due to the lack of representative training datasets, minimal insurance coverage, and inaccessibility amongst lower income groups. The literature review will cover the development of AI systems in lung cancer detection, existing disparities in lung cancer treatment, and technical design factors of an AI system that can contribute to implicit bias. I will then analyze the socioeconomic background of patients and the role of insurance companies to determine how accessible advanced treatment plans are for patients of all backgrounds. Additionally, the type of datasets used to develop AI systems in the context of lung cancer treatment will be analyzed to determine how representative these datasets are. Through my analysis, I intend to find that there is a relationship between the financial standing of lung cancer patients and their likelihood of opting for treatments involving medical AI devices. In order to make sure that AI systems themselves do not propagate social biases amongst patients, healthcare officials and policy makers must standardize the use of these systems.

Literature Review

Prior research has established that fairness in healthcare AI systems is heavily influenced by the data that is available to train the models on. The ability to create representative datasets is hindered because many underserved communities do not have the means to digitize their health records (Schönberger, 2019). This digital divide contributes to the existing healthcare disparities amongst various socioeconomic groups. A lack of data on certain populations leads to sample size disparity amongst the AI system's training data, which was highlighted in a case study where African Americans were over diagnosed with schizophrenia (Schönberger, 2019). In addition to the problem of sampling bias, medical AI is tested and informed by clinical trials, which often do not match the target demographic of the system's users (Dankwa-Mullan & Weeraratne, 2022). Clinical trials have historically reported an underrepresentation of enrollment from older patients and African Americans, resulting in recruitment bias amongst training data. While researchers can mitigate this issue through an improved participant screening process, there has been little to no progress in the large-scale efforts to improve this process.

Several scholars agree that data labels are the key determinants of an AI algorithm's predictive quality and predictive bias. Models containing proxy labels disproportionately produce outcomes favoring those in the majority and disadvantage those in minority groups (Paulus & Kent, 2020). Proxy labels refer to labels defining the outcome of a model, which can skew a model's prediction to favor the outcome of interest if the labels are defined in a prejudiced manner. One notable study explored an algorithm that determines which patients require high-risk care management programs in order to provide additional resources to improve care for those with complex needs (Obermeyer et al., 2019). In this study, researchers found that white patients were being disproportionately favored over black patients. In their analysis, they discovered that the healthcare costs were very different between white patients and black patients, specifically because black patients had fewer surgical and specialist costs but more emergency care costs. Thus, the researchers deduced that one of the main driving forces of bias in this algorithm was the fact that patient need was based heavily on historical medical expenses, which is an example of labeling bias (Obermeyer et al., 2019). Another example demonstrated that exclusively using cancer diagnosis as a predicting factor for cancer incidence resulted in labeling bias because affluent communities tend to have more regular screening check-ups than underserved communities (Paulus & Kent, 2020). The above studies highlight that using a single aspect of health care to determine patient needs will likely contribute to labeling bias because there are several other external variables that need to be considered.

Lung cancer is one of the most researched cancer types with regards to AI development due to the availability of a vast number of CT/PET scans (Tunali et al., 2021). To provide some background on the disease, lung cancer is divided into two major groups: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC is more aggressive and less common than NSCLC because the abnormal cells that form in the lungs grow rapidly and uncontrollably (Blandin Knight et al., 2017). Lung cancer screening targets high risk populations, which include people who are typically in the age range of 55-74 with a specified minimum smoking history (Nooreldeen & Bach, 2021). Low-dose CT scans are being used for these routine screening methods; however, new screening methods along with the applications of AI systems are being investigated in hopes of improving early detection of lung cancer. Currently, medical AI devices are primarily used for lung cancer diagnosis and treatment by analyzing medical data and building models that can accurately predict patient responses to treatments and risk of tumor relapse. Radiomics, an emerging field of research, focuses on converting medical imaging into quantitative data which can be merged with other data sources and analyzed using AI to improve detection of pulmonary nodules (Tunali et al., 2021). In addition to tumor detection, AI in lung cancer imaging can improve the accuracy of tumor staging and assist doctors with treatment decisions by streamlining the process of data analysis (Pei et al., 2022). IBM's Watson for Oncology (WFO) system utilizes an AI-based compressed storage system to aid doctors in identifying crucial information regarding a patient's medical history and accordingly choose a mode of surgery (Pei et al., 2022). In a study conducted by You et al. (2020), researchers found the recommended treatment plans provided by Watson for Oncology were consistent with the hospital treatment regimen determined by the medical team in the study about 85% of the time. This situation illustrated that medical AI systems have the ability to make decisions similar to oncology professionals in a much more efficient manner.

While there are copious amounts of research demonstrating the benefits AI brings to lung cancer treatment, there is not much research on how AI contributes to the existing disparities amongst lung cancer patients. Lung cancer is known to be the leading cause of cancer deaths in low- and middle-income countries, which is partly due to patients' untimely hospital visits, implying that socioeconomic status as well as patient circumstances greatly impact the accessibility to adequate health care facilities (Lubuzo et al., 2020). Race and sex are also correlated with patients who are untimely treated for lung cancer, which can potentially skew the data available for AI models to train on (Shugarman et al., 2009). These sources detail the existing social disparities among lung cancer patients, but they fail to discuss them in the context of treatment involving AI and how this innovation has impacted these disparities, which is a problem that must be addressed.

I intend to analyze the push and pull of various actors in the healthcare system with a focus on the relationships between AI systems as non-human actors and doctors, patients, and insurance companies as the human actors. I will use Actor Network Theory (ANT) to get a better understanding of the relationships between AI systems, doctors, patients, and health insurance companies. ANT argues that human and non-human actors are enrolled in the construction of technological systems with an emphasis on generalized symmetry, which refers to the fact that human and non-human actors must be treated with equal importance (Latour, 2005). I hope that using ANT in this context will provide more insight into how interactions take place between the various actors and whether these interactions lead to any inequalities within the network.

Methods

I used mainly secondary sources by gathering journal articles focusing on the effects of AI systems on lung cancer patients. I focused on collecting sources published within the last 10 years because the role of AI in medicine has become more prevalent during this time. I also collected information from journal articles about the relationship between health insurance companies and lung cancer patients. I hope to draw connections between the thread on insurance companies and the thread focusing on the effect of AI systems on lung cancer patients so that I can better explore the relationships between these actors. I also studied articles detailing the specifics of AI training datasets such as the Lung Image Database Consortium (LIDC) and the Image Database Resource Initiative (IDRI) to evaluate how representative the AI systems trained on these datasets are.

Analysis

As discussed earlier, the development of AI systems used to diagnose lung cancer patients is heavily dependent on publicly available datasets which can lead to sampling bias amongst the results produced by these systems. It is typical for medical AI systems to rely on these public datasets because they contain well established sources of data, but this in itself is a problem. Most deep learning technologies have trained and tested their systems on the Lung Image Database Consortium and Image Database Resource Initiative (LIDR-IDRI) dataset, which can lead to overfitting and homogenous models (Li et al., 2019). The researchers of this paper conducted an analysis of various deep learning devices trained on datasets that were not from the LIDR-IDRI database and found that these models reached high levels of classification accuracy (Li et al., 2019). This indicates that other datasets can achieve the same or better results, but most medical AI systems continue to train on LIDR-IDRI datasets, presenting an unaddressed problem. Improvement of medical AI devices is greatly limited by the absence of unique large-sized and well-annotated datasets. Medical image datasets often only contain less than 10,000 images, of which only a small percentage is annotated by medical professionals (X. Chen et al., 2022). This is a relatively small amount of data compared to the datasets required to train other computer vision models used for purposes much less complex than the detection of lung cancer. Some contend that unsupervised deep learning methods can overcome this limitation, but this has led to the development of a single widely used public dataset, introducing sampling bias and community-wide overfitting. Even when unsupervised deep learning models are used to address the problem of overfitting, utilizing a single dataset to initially train the model can cause it to encode the training data too well, which will not generalize to new inputs (Zhang & Yang, 2019).

The lack of standardization policies regarding the use of AI systems in healthcare has led to minimal coverage of costs by insurance companies, which in turn affects the patients in need of these systems. Current laws dictate that doctors have the final say when it comes to interpreting and making decisions based on information produced by AI systems (Vogel, 2019). This makes the use of AI systems subjective and difficult to standardize without an overarching authority setting policies. The coverage of a medical service depends on the clinical utility aspect of AI devices in the context that they improve overall patient outcomes (Park et al., 2021). It was noted that medical professionals are responsible for determining whether AI devices are beneficial to patients and demonstrate its clinical utility (Park et al., 2021). This further increases the subjectivity of the process of validating the effectiveness of medical AI devices and makes it more difficult to obtain insurance coverage on these devices. It is evident that the doctors themselves play a crucial role when it comes to how the insurance companies evaluate the effectiveness of medical AI devices. When analyzing this relationship under Actor Network Theory, we may conclude that doctors are essentially defining the terms under which medical AI devices are clinically approved in the eyes of insurance companies, which affects the accessibility of these devices for those in need of them.

Financially burdened patients lack access to advanced treatment facilities, delaying lung cancer diagnosis until it is in the advanced stages. One of the biggest factors associated with financial burden is age. Lung cancer patients below the age of 65 have been shown to be in greater financial distress due to more volatile income sources and unaffordable private insurance rates (Ezeife et al., 2019). High out of pocket costs contribute to the financial distress of these patients and make them less likely to choose to pay for cutting edge lung cancer therapies. Researchers have demonstrated that uninsured patients "have longer delays in initiating curative treatments, which can impact survival in lung cancer," since presenting at later stages does not guarantee the standard treatment (Rice et al., 2020). Only 8% of uninsured patients underwent surgical resection as compared to the 45% of privately insured patients (Rice et al., 2020). If uninsured patients are unable to afford publicly funded therapies like surgery, their chances of undergoing breakthrough treatments involving AI systems is extremely low. Some may argue that there are many confounding factors present in the residential environments of patients of a lower socioeconomic status which invalidates the significance of this claim; however, the existing financial disparity is compounded in these environments and in fact makes this issue increasingly significant. This creates a never ending cycle: “poverty exacerbates poor health while poor health makes it harder to get out of poverty” (Lubuzo et al., 2020).

Actor Network Theory (ANT) is applicable to this situation due to the involvement of both human and nonhuman actors and the power imbalances between these actors, which contributes to the treatment bias involved with lung cancer patients. The main actors that I identified throughout my research were oncologists, lung cancer patients, medical AI systems, and health insurance companies. With respect to the patients, external factors such as socioeconomic status, smoking history, age, race, and gender can drastically impact their relationships with the other actors in this network. For example, the treatment plans recommended by Watson for Oncology had very poor concordance rates with doctor recommendations in the study investigating Chinese patients (Jie et al., 2021). Since Watson for Oncology was trained only on data from the Memorial Sloan Kettering Cancer Center (MSK), there was a built-in bias about how the system approaches treatment decisions. This was further highlighted when the system did not recommend the use of research drugs such as Icotinib and Endostar, which are primarily used in China due to differences in general physique and mutation rates in Chinese patients as opposed to American patients (Jie et al., 2021). This study indicates that it can be difficult for non-human actors to adapt to changing scenarios because of the limitations of training datasets, which puts the patients at risk while also making the doctors’ work more demanding as they would need to review each recommendation.

ANT has revealed that the doctors are typically in a higher position of power in this network because they have complete authority when it comes to making the final decisions about patient treatment in addition to being responsible for demonstrating the clinical utility of medical AI systems. While this power imbalance is not unreasonable, doctors have the ability to limit the use of AI systems, which can in turn negatively impact how insurance companies perceive the clinical utility of these systems.

Conclusion

It is evident that the evolution of medical AI systems is not being adequately regulated by oncology professionals, including the engineers developing these systems as well as the doctors using them, which has led to several problems aggravating the socioeconomic disparities existing in worldwide cancer care. The use of medical AI can be very subjective because it is left up to doctors and medical professionals to determine how to utilize the information produced from the AI systems, which further complicates the process of standardizing these practices. I intend for oncology researchers and medical professionals who are closely tied to this area of technological development to use my research when improving existing AI systems. I believe that future research could build off my work by exploring various data curation methods, such as federated learning (X. Chen et al., 2022), to build more representative datasets. I also hope my research will spark interest in establishing collaborative evaluation of medical AI devices by including actors like underserved patients and insurance companies in the discussion of these problems.

AI is shaping the way doctors go about cancer treatment because of innovative recognition technology that is improving the screening and detection of early-stage cancer. In order for these systems to continue advancing at the rate they are, it is essential that we as engineers discuss the potential overlooked problems and ramifications associated with AI systems to ensure that patients are being treated with high quality treatments.

References

Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., & Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, *7*(9), 170070. https://doi.org/10.1098/rsob.170070

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, *79*, 102444. https://doi.org/10.1016/j.media.2022.102444

Chen, Z.-H., Lin, L., Wu, C.-F., Li, C.-F., Xu, R.-H., & Sun, Y. (2021). Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Communications (London, England)*, *41*(11), 1100–1115. https://doi.org/10.1002/cac2.12215

Christie, J. R., Lang, P., Zelko, L. M., Palma, D. A., Abdelrazek, M., & Mattonen, S. A. (2021). Artificial Intelligence in Lung Cancer: Bridging the Gap Between Computational Power and Clinical Decision-Making. *Canadian Association of Radiologists Journal*, *72*(1), 86–97. https://doi.org/10.1177/0846537120941434

Dankwa-Mullan, I., & Weeraratne, D. (2022). Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias, and Data Diversity. *Cancer Discovery*, *12*(6), 1423–1427. https://doi.org/10.1158/2159-8290.CD-22-0373

Ezeife, D. A., Morganstein, B. J., Lau, S., Law, J. H., Le, L. W., Bredle, J., Cella, D., Doherty, M. K., Bradbury, P., Liu, G., Sacher, A., Shepherd, F. A., & Leighl, N. B. (2019). Financial Burden Among Patients With Lung Cancer in a Publically Funded Health Care System. *Clinical Lung Cancer*, *20*(4), 231–236. https://doi.org/10.1016/j.cllc.2018.12.010

Jie, Z., Zhiying, Z., & Li, L. (2021). A meta-analysis of Watson for Oncology in clinical application. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84973-5

Latour, Bruno (2005). Reassembling the Social: An Introduction to Actor-Network Theory. New York: Oxford University Press.

Li, D., Mikela Vilmun, B., Frederik Carlsen, J., Albrecht-Beste, E., Ammitzbøl Lauridsen, C., Bachmann Nielsen, M., & Lindskov Hansen, K. (2019). The Performance of Deep Learning Algorithms on Automatic Pulmonary Nodule Detection and Classification Tested on Different Datasets That Are Not Derived from LIDC-IDRI: A Systematic Review. *Diagnostics*, *9*(4), Article 4. https://doi.org/10.3390/diagnostics9040207

Lubuzo, B., Ginindza, T., & Hlongwana, K. (2020). The barriers to initiating lung cancer care in low-and middle-income countries. *The Pan African Medical Journal*, *35*, 38. https://doi.org/10.11604/pamj.2020.35.38.17333

Nooreldeen, R., & Bach, H. (2021). Current and Future Development in Lung Cancer Diagnosis. *International Journal of Molecular Sciences*, *22*(16), Article 16. https://doi.org/10.3390/ijms22168661

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Park, S. H., Choi, J., & Byeon, J.-S. (2021). Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence. *Korean Journal of Radiology*, *22*(3), 442–453. https://doi.org/10.3348/kjr.2021.0048

Paulus, J. K., & Kent, D. M. (2020). Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *Npj Digital Medicine*, *3*(1), Article 1. https://doi.org/10.1038/s41746-020-0304-9

Pei, Q., Luo, Y., Chen, Y., Li, J., Xie, D., & Ye, T. (2022). Artificial intelligence in clinical applications for lung cancer: Diagnosis, treatment and prognosis. *Clinical Chemistry and Laboratory Medicine*, *60*(12), 1974–1983. https://doi.org/10.1515/cclm-2022-0291

Rice, S. R., Vyfhuis, M. A. L., Scilla, K. A., Burrows, W. M., Bhooshan, N., Suntharalingam, M., Edelman, M. J., Feliciano, J., Badiyan, S. N., Simone, C. B., Bentzen, S. M., Feigenberg, S. J., & Mohindra, P. (2020). Insurance Status is an Independent Predictor of Overall Survival in Patients With Stage III Non–small-cell Lung Cancer Treated With Curative Intent. *Clinical Lung Cancer*, *21*(3), e130–e141. https://doi.org/10.1016/j.cllc.2019.08.009

Schönberger, D. (2019). Artificial intelligence in healthcare: A critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology*, *27*(2), 171–203. https://doi.org/10.1093/ijlit/eaz004

Shugarman, L. R., Mack, K., Sorbero, M. E. S., Tian, H., Jain, A. K., Ashwood, J. S., & Asch, S. M. (2009). Race and Sex Differences in the Receipt of Timely and Appropriate Lung Cancer Treatment. *Medical Care*, *47*(7), 774–781.

Tunali, I., Gillies, R. J., & Schabath, M. B. (2021). Application of Radiomics and Artificial Intelligence for Lung Cancer Precision Medicine. *Cold Spring Harbor Perspectives in Medicine*, *11*(8), a039537. https://doi.org/10.1101/cshperspect.a039537

Vogel, L. (2019). Rise of medical AI poses new legal risks for doctors. *CMAJ*, *191*(42), E1173–E1174. https://doi.org/10.1503/cmaj.1095823

You, H.-S., Gao, C.-X., Wang, H.-B., Luo, S.-S., Chen, S.-Y., Dong, Y.-L., Lyu, J., & Tian, T. (2020). Concordance of Treatment Recommendations for Metastatic Non-Small-Cell Lung Cancer Between Watson for Oncology System and Medical Team. *Cancer Management and Research*, *12*, 1947–1958. https://doi.org/10.2147/CMAR.S244932

Zhang, Y., & Yang, Q. (2019). Regularization methods for deep learning models. Neural Networks, 111, 280-296. https://doi.org/10.1016/j.neunet.2018.12.003