Undergraduate Thesis Prospectus

Navigating Lung Surgery Complications: Data-Driven Exploration of Demographic, Environmental and Intraoperative Factors

(technical research project in Computer Science)

Artificial Intelligence: Impartial Arbiter or Biased Judge?

(sociotechnical research project)

by

Keivon Chamanara

November 1, 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.

Keivon Chamanara

Technical advisor: Briana Morrison, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

How can health complications be better anticipated?

Late diagnoses and diagnostic errors are the 6th leading cause of death in the United States, contributing to around 80,000 deaths per year (Johns Hopkins Medicine, 2023; Tudor et al., 2016). Poor communication between primary and secondary care physicians, multiple comorbidities masking the real problem, and lack of patient awareness are just a few of the reasons resulting in this statistic (Tudor et al., 2016). Ensuring that health complications are better foreseen allows people to take preventive measures early, potentially avoiding serious health issues. Achieving this will require strengthening the healthcare system, fostering patient-doctor relationships, and leveraging technology for more effective monitoring and intervention.

Navigating Lung Surgery Complications: Data-Driven Exploration of Demographic, Environmental and Intraoperative Factors

How can technology be used to better analyze and predict post-operative complications?

This independent research project, supervised by Dr. Anil Vullikanti of the University of Virginia Biocomplexity Institute, aimed to assess how specific factors influence post-operative complications in lung surgery patients using different machine learning techniques. With the project's findings, the goal was to extend the models created to predict the post-surgical outcomes of future patients.

Prior research and models on post-lung surgery complications mostly focused on demographic and intraoperative factors, which provided a useful basis for this project. Kelkar (2015) investigates several common post-surgical complications and their likely causes. An example is post-operative hypoxemia, which results from the residual effect of anesthetic agents.

However, Vullikanti and other stakeholders, including an anesthesiologist at the University of Virginia hospital, wanted to extend existing predictive models and include additional variables to determine if model prediction would improve. The results from this project could help doctors better understand their patients' backgrounds and guide their decisions before, during, and after their procedures.

Since most previous research focused on the effect of intraoperative and health factors on complications after lung surgery, doctors could solely analyze these to influence their decision-making process. Previous models used this research to predict the likelihood of post-surgical complications, like those developed by Mao et al. (2021). Mao et al.'s (2021) statistical models found that patients with cardiovascular or respiratory comorbidities are more likely to develop post-pulmonary complications than those without. This served as a valuable foundation for our project, offering us insight into the factors that warrant consideration and allowing us to widen our scope to encompass additional variables, such as geospatial elements.

I analyzed a dataset containing details of numerous lung surgery patients from the University of Virginia Hospital. From this data, I utilized various machine learning classifiers to construct insightful predictive models. Tools used included Python, along with specialized libraries for data manipulation and machine learning, such as Pandas, NumPy, and Scikit-learn. For modeling purposes, I employed algorithms such as decision trees, logistic regression, and support-vector machines.

At the end of this project, the mission was to have a model that factored in a host of variables, from demographic to geospatial, to accurately predict the chance of post-operative complications in lung surgery patients. This model could be used by doctors and researchers alike to improve their knowledge and make more calculated decisions.

Artificial Intelligence: Impartial Arbiter or Biased Judge?

In the U.S., how are social groups competing to influence the extent of discriminatory bias in artificial intelligence tools?

With the increased adoption of artificial intelligence (AI) tools, there is growing concern over its ability to encode and exacerbate discriminatory biases in education, healthcare, and other areas. There is no doubt that AI has plenty of benefits, from improving collision avoidance technology in cars to more accurately labeling problematic lymph nodes (West & Allen, 2018). However, these advancements coupled with promises of increased productivity and economic gains overlook the possible risks associated with AI. Online loan lenders that incorporate AI into their products were found to be 80% more likely to reject Black applicants than their White counterparts (Hale, 2021). Akselrod (2021) gives an example of how AI tools used to evaluate possible tenants use court records with built-in biases stemming from centuries of systemic racism and sexism. Not to forget that the biases and prejudices of humans can be reflected in the AI tools they build. The contrast between AI's benefits and risks sheds light on the social groups vying to influence the role AI places in our society and the extent to which it harms already marginalized groups.

In their assessment of AI, Schmidt & Stephens (2019) discuss the details behind algorithmic discrimination, from how it emerges to how it can be reduced. They note that bias from these tools is not always evident, since many models nowadays use "alternative data". Alternative data is information that has not historically been used in model building in a given context. An example is factoring in the type of phone someone uses in predicting their credit trustworthiness (Schmidt & Stephens, 2019). In this case, questions can be raised over whether using this variable is unfair or irrelevant. In a study of one of Amazon's hiring tools, Lobacheva & Kashtanova (2022) attribute bias against women job candidates to training data that underrepresented women. The algorithm trained on resumes submitted over the past 10 years, which primarily came from men (Lobacheva & Kashtanova, 2022). This highlights that these tools can show signs of bias in many forms, whether it is directly from developers or the dataset used. Berk (2020), however, argues that discriminatory bias in AI has been overstated, and AI's advantages exceed its disadvantages. He mentions that human reasoning can be just as flawed, if not more, than that of AI (Berk, 2020). Together, these researchers offer insightful, yet opposing, viewpoints that allow a stronger understanding of all the social groups involved.

There are advocacies fighting the biases present in AI applications. The Algorithmic Justice League (AJL) publicizes and resists algorithmic biases in industry and policy (AJL, 2023). They use a variety of techniques, from research to artwork, to raise awareness of the harm unabated AI usage poses. Joy Buolamwini, the organization's founder, emphasizes that through these methods, "the Algorithmic Justice League will continue fighting for the excoded, those negatively impacted by AI" (AJL, 2023). One of their most impactful projects was "Gender Shades", which aimed to "evaluate the accuracy of AI powered gender classification products" (AJL, 2023). This builds off Buolamwini's research in analyzing the inaccuracies of facial detection systems, particularly against women and ethnic minorities. The project brought attention to this issue and resulted in major corporations, such as Google and IBM, to scale back or completely stop producing or selling facial recognition technologies.

Many tech companies, however, generally want to market their applications even when they may encode biases. For example, Equivant developed COMPAS, a tool that assesses a defendant's risk of reoffending. It argues its tools are more impartial than humans (Equivant, 2019). Some researchers found that with COMPAS, Black people are nearly twice as likely as White people to be mistakenly classified as higher risk without actually re-offending (Angwin et al., 2016). However, Equivant refutes these claims and points to opposing research that shows COMPAS performing “equally well for African American and Anglo defendants” (Equivant, 2019). Companies like Equivant acknowledge the concerns over AI’s potential to discriminate but downplay or deny the possibility of their products doing so.

Some engineers, turning against their employers, have objected to the indiscriminate use of AI. When Timnit Gebru, a computer scientist, denounced algorithmic biases, Google fired her (Gebru, 2021). Gebru claims that the most important way to safeguard against the unsafe use of AI is “curbing the power of the companies who develop it and increasing the power of those who speak up against the harms of AI and these companies’ practices” (Gebru, 2021). Being at Google gave her an inside look into how tech companies develop these tools and respond to the backlash to them. She has made a concerted effort to break big tech’s monopoly on AI research - so that we can see “technology that prioritizes the wellbeing of citizens” (Gebru, 2021).

Clients of tech companies often defend the AI programs they use. For example, London’s Metropolitan Police (2022) claims AI boosts operational efficiency without introducing bias. They assert that active measures are taken to mitigate risks related to bias. The New York State Division of Criminal Justice Services uses the COMPAS tool from Equivant. In its analyses of the product’s accuracy and chance of bias, it concluded that satisfactory predictive accuracy rates have been achieved (Lansing, 2012). These clients ensure that the tools they invest in are free from bias, even if that sometimes contradicts outside researchers’ claims.

These concerns have led the Biden administration to address the unchecked growth of AI. In their “Blueprint for an AI Bill of Rights”, The White House (2022) states that “among the great challenges posed to democracy today is the use of technology, data, and automated systems in ways that threaten the rights of the American public”. They acknowledge that AI systems developed over the past decade can have harmful consequences if not created and maintained carefully. Accordingly, the administration saw the need for a set of standards governing the responsible and fair use of AI. One of the primary principles of their Bill of Rights is protection from algorithmic discrimination, where they advocate for developers to “design systems in an equitable way” and to proactively consider the possibility of bias in their tools (The White House, 2022).

These social groups all aim to influence the extent of discriminatory bias in artificial intelligence tools. On one end are the groups exacerbating these biases, as opposed to those striving to thwart or mitigate them. There is a significant interplay between these groups, shaping the ethical landscape of AI and its place in society.

References

AJL. (2023). Algorithmic Justice League. *Algorithmic Justice League*. https://www.ajl.org

Akselrod, O. (2021, July 13). How Artificial Intelligence Can Deepen Racial and Economic Inequities. *American Civil Liberties Union*. https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016, May 23). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Berk, R. A. (2020). Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement. *Annual Review of Criminology*, *4*(1), 210. Web of Science

Equivant (2019). Practitioner's Guide to COMPAS Core. https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf

Gebru, T. (2021, December 6). For truly ethical AI, its research must be independent from big tech. *The Guardian*. https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebru

Hale, K. (2021, September 2). A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied. *Forbes*. https://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=1e1b0e2436fe

Johns Hopkins Medicine. (2013, April 23). Diagnostic Errors More Common, Costly And Harmful Than Treatment Mistakes. https://www.hopkinsmedicine.org/news/media/releases/diagnostic_errors_more_common_costly_and_harmful_than_treatment_mistakes

Kelkar K. V. (2015). Post-operative pulmonary complications after non-cardiothoracic surgery. *Indian Journal of Anaesthesia*, *59*(9), 599–605. https://doi.org/10.4103/0019-5049.165857

Lansing, S. (2012). New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy Criminal Justice Research Report. *New York State Division of Criminal Justice Services*. https://www.criminaljustice.ny.gov/crimnet/ojsa/opca/compas_probation_report_2012.pdf

Lobacheva, A., and Kashtanova, E. (2022). Social Discrimination in the Epoch of Artificial Intelligence. *WISDOM*, *1*(2). 99. Web of Science

Mao, X., Zhang, W., Ni, Y. Q., Niu, Y., and Jiang, L. Y. (2021). A Prediction Model for Postoperative Pulmonary Complication in Pulmonary Function-Impaired Patients

Following Lung Resection. *Journal of Multidisciplinary Healthcare*, *14*, 3187–3194. https://doi.org/10.2147/JMDH.S327285

Metropolitan Police. (2022). Machine Learning and Deep Learning Software. https://www.met.police.uk/foi-ai/metropolitan-police/d/july-2022/machine-learning-and-deep-learning-software/

Schmidt, N., and Stephens, B.E. (2019). An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination. *ArXiv, abs/1911.05755*.

The White House. (2022). Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Tudor Car, L., Papachristou, N., Bull, A., Majeed, A., Gallagher, J., El-Khatib, M., Aylin, P., Rudan, I., Atun, R., Car, J., and Vincent, C. (2016). Clinician-identified problems and solutions for delayed diagnosis in primary care: a PRIORITIZE study. *BMC Family Practice*, *17*(1), 131. https://doi.org/10.1186/s12875-016-0530-z

West, D., and Allen, J. (2018, April 24). How artificial intelligence is transforming the world. *Brookings*. https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world/