Development of a Ventricular Tachycardia Convolutional Neural Network Diagnostic System

(Technical Paper)

Gender Bias in AI-Powered Cardiac MRI Diagnostics: Examining Disparities in Medical Outcomes

(STS Paper)

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Civil Engineering

By

Jonathan Le

November 11, 2024

Technical Team Members:

Dr. Derek Bivonna

Dr. Kenneth Bilchick

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.

ADVISORS

Dr. Miaomiao Zhang, Department of Computer Science

Dr. Coleen Carrigan, Department of Engineering and Society

Introduction

Ventricular Arrhythmia (VA) is a dangerous cardiovascular condition that causes issues in electrical conductivity within the heart. As the leading cause of sudden cardiac death, It is estimated that there are over 300,000 deaths from VA every year (Foth et al. 2024). In order for the heart to pump oxygenated blood throughout the body, the heart requires a consistent rhythm of electrical signals to be directed towards cardiac tissue; VA disrupts these electrical pathways, leading to irregular contractions and inefficiencies in blood flow. VA is highly likely among patients that suffered previous acute conditions like heart attack where they have scar tissue

To treat these dangerous arrhythmias, clinicians have been using implantable cardioverter-defibrillators (ICDs) in order to treat such conditions. While these devices are vital for regulating heart rhythm, they come with significant costs - historically around $30,000 per device and associated procedures (Hlatky & Mark, 2007), representing a substantial healthcare expense that necessitates selective deployment to those that are highest at risk. Previous heuristics for determining whether to implant an ICD utilize an ejection fraction (proportion of blood that gets pumped out every cycle) threshold, where a patient with an $EF < 0.35$ is classified a high likelihood of VA and necessitates an implantation of an ICD (Kusumoto et al, 2014). But these metrics are often not comprehensive, leading to false positives and negatives. Such misclassifications lead to a waste of ICD use and death of VA prone patients, which necessitate an improved method for classifying which patients need an ICD.

For the technical portion of this paper, I will introduce a novel approach using a convolutional neural network that assesses the risk of VA by examining cardiovascular medical images. As this system will inform critical ICD deployment decisions, ensuring equitable performance across demographics is essential. Such behavior of a clinical decision system is heavily dependent on the quality of data, so minimizing bias in data against underrepresented demographics, especially women, is an important issue in cardiology. However, examining how biases are encoded into the data is a complex issue: multiple stakeholders with varying priorities influence the quality of data fed into models, where bias is difficult to identify. This necessitates investigating intersecting structures of power and how they encode bias into the data, leading to bias. Therefore, in the STS portion, I will investigate how power structures lead to AI models being biased against women when diagnosing cardiovascular conditions. Such questions on how best to decide and other sociological implications will be discussed in the STS portion of this prospectus. Together, these technical and social analyses aim to enhance VA prediction accuracy while ensuring responsible implementation, ultimately improving clinical outcomes for patients at risk.

Technical Topic

In healthcare, AI has been taking an increasing role in decision making. AI models have demonstrated impressive performance in supporting clinical decision making, from diagnosing diseases and predicting patient outcomes to recommending personalized treatment plans (Busnatu et al. 2022). One of the most significant advances is the development of Convolutional Networks, used for processing images. Convolutional Neural Networks are networks that process information through the use of a mathematical convolution operation. The particular use of Convolutional Networks and derivative architectures have achieved remarkable success in image processing applications; for instance, one network ChexNet achieved near expert level or surpassing expert level performance when classifying disease classifications (Rajpurkar et al. 2018).

In addition to the classification problem, recent advances have also been made in the deep learning problem of semantic segmentation. This problem for images is of interest because radiology often involves analyzing images where one would have to highlight non-linear regions of interest. For example, for identification of tumors on a medical image, one may want to create a model that specifically highlights a region of interest that corresponds to the tumor for downstream processing applications like classification. One of these deep learning architectures that performs well on this task of semantic segmentation is the Swin-U-Net, which is a model that allows for hierarchical processing of both granular and abstract features. This particular U-Net variant is particularly strong because this variant of the U-Net utilizes transformers in every layer to encode long-term dependencies that typical convolution operations would not be able to encode (Cao et al. 2021).

In my project, I aim to develop a new architecture that takes the Cardiac Magnetic Resonance (CMR) Images short-axis slices of the Left Ventricle and outputs a prediction on the risk of Ventricular Arrhythmia.

In order to do this, I will use a pre-trained Swin U-Net model designed to segment in all of the slices the scar regions of the image. Using the output, I will likely add a feedforward network that will output a probability score from 0 to 1 assessing the risk of VA.

The idea behind using the CMR images of the Left Ventricle of the heart is that this data of tissue is a significant indicator for the risk of VA. Given a Late Gadolinium Enhancement (LGE) when acquiring the image with an MRI machine, one can obtain images that highlight scarred portions of the image, which correspond to dead myocardial cells (heart tissue). More dead myocardial cells decreases conductivity of electric signals as dead cells act as insulators, increasing the likelihood of VA.

This approach is significant because this model provides a new more comprehensive way of evaluating whether a patient has a high likelihood of VA and therefore needs an ICD, more so than the Ejection Fraction heuristic. Such advances in decision making will enable better prioritization of ICDs, ensuring that those with the highest risk are treated.

STS Topic

AI systems for diagnosing cardiovascular diseases are not perfect and are subject to bias against women from a multitude of factors during development. One particular factor is the type of data that are used in models: current research suggests that a majority of the cardiac knowledge and data is representative of men (Hamid et al., 2024). For instance, AI algorithms may prioritize detecting classic male heart attack symptoms like chest and arm pain while missing women's more common symptoms such as fatigue and jaw pain. If models are trained on data that represents men better than women (the data emphasizes male centric features explaining disease), the models can potentially have a biased performance against women. This is significant because a biased result perpetuates previously occurring biases, leading to negative outcomes such as delayed diagnosis, improper treatment plans.

Furthermore, these decision making systems behave primarily to the interests of those in power, primarily men responsible for creating and managing these systems. For example, healthcare institutions may prioritize developing AI tools that streamline diagnosis for conditions with established male symptom patterns, as this aligns with existing clinical workflows and profit models, rather than investing in systems that could better detect women's presentations.

Similarly, Crawford (2021, p. 8) argues that "due to the capital required to build AI at scale and the ways of seeing that it optimizes AI systems are ultimately designed to serve existing dominant interests." Having AI systems serving those in power especially in interests in resource allocation for development means that the priority of male interests will emerge over that of women, occurring particularly in clinical trials where male participants are prioritized, leading to male centric cardiac knowledge and data. This implies power structure dynamics as important to causing negative outcomes for women in cardiovascular healthcare.

Therefore, in emphasizing the issue of unfair treatment of women cardiovascular healthcare, I present the following research question and elaborate on the methods to answer such question:

How do medical decision-making systems in AI, used to diagnose cardiovascular conditions, reinforce gender bias and cause poorer outcomes for women than men?

Answering such a question will help identify key insights and dynamics that cause AI to be biased, therefore creating insight on nuanced solutions to this complex problem.

Methods

Examining the dynamics for how AI systems cause poorer outcomes requires a careful analysis of key power structures influencing AI systems. Intersecting power structures are important because in medicine one must recognize the social organization as well as the distributions of power that exist between medical institutions and women. For instance, in clinical trials for treatments in heart conditions, researchers often prioritized studying male subjects to better analyze the anatomy as in women, hormonal changes had varying effects on cardiac function which was hard to analyze (Liu et al. 2016). In positions of leading research, where researchers hold power to the epistemic framing of cardiology, researchers prioritize goals of efficacy in clinical trials for treatments; in overemphasizing men and assuming similar cardiac function in women, they assume limited structures of explanation for cardiac function in women.

This has negative effects not only in informing precise treatments of women, but also in informing of the dataset curation, acquisition that are essential for informing AI clinical decision supporting systems of the right diagnosis (Deo, 2015). For instance, a study found that only 17% of cardiologists correctly identified women as having greater risk for heart disease than men (Cirillo et al., 2020). Indeed, physicians are typically trained to recognise patterns of angina and myocardial infarction (heart attack) that occur more frequently in men, resulting in women being typically under-diagnosed for coronary artery disease. Consequently, training an algorithm on available data on diagnosed cases could be influenced by an implicit gender bias as cardiologists labelling datasets could under diagnose women.

Additionally, using intersectionality is a key tool to trace how power structures affect women of different races. Most notable is the demographic of Black women, who not only face gender discrimination but also racial discrimination. Due to (historical) institutions that force black women in low economic conditions, black women often have difficulty accessing healthcare, and are therefore underrepresented in clinical trials (Nanavati et al., 2024). Underrepresentation in clinical trials leads to less clinical knowledge on specific manifestations of disease in black women, leading to poorer representation of conditions in data and worse outcomes for the AI models.

Therefore, identifying and addressing bias in AI cardiovascular diagnostic systems requires understanding the complex web of power structures that shape medical knowledge production. From clinical trial design to healthcare access barriers, these institutional forces create data gaps that AI systems inevitably inherit. Recognizing these systemic issues is crucial for developing more equitable AI systems that can effectively serve all populations.

Conclusion

It is clear that AI continues to play an increasing role in decision-making medicine, where models can spot decisions that humans cannot. In cardiac imaging, as demonstrated through the technical project of developing a CNN model for VA prediction, there is enormous potential for such technologies to improve patient care through disease diagnosis and risk assessment. However, in ensuring the best clinical outcomes, examining the gender biases that are encoded in cardiac AI models remains an important issue for later informing practices in data/knowledge curation as well as AI development. From my research into the sociotechnical problem, findings such as the dynamics of clinical trials affecting data for specific diseases can not only identify better key practices in development such as better data curation, and training techniques to compensate for biased data. Through working on both the VA prediction model and examining the social implications of AI-assisted clinical decisions, I hope to create insights that guide the optimal use of these technologies in healthcare. It is not only the development of the technologies to satisfy accuracy metrics that is essential, but also ensuring reliable performance across both men and women.

Word Count: 1870

References

Al Hamid, A., Beckett, R., Wilson, M., Jalal, Z., Cheema, E., Al-Jumeily OBE, D., Coombs, T., Ralebitso-Senior, K., & Assi, S. (n.d.). Gender Bias in Diagnosis, Prevention, and Treatment of Cardiovascular Diseases: A Systematic Review. Cureus, 16(2), e54264. https://doi.org/10.7759/cureus.54264

Busnatu, Ștefan, Niculescu, A.-G., Bolocan, A., Petrescu, G. E. D., Păduraru, D. N., Năstasă, I., Lupușoru, M., Geantă, M., Andronic, O., Grumezescu, A. M., & Martins, H. (2022). Clinical Applications of Artificial Intelligence—An Updated Overview. Journal of Clinical Medicine, 11(8), 2265. https://doi.org/10.3390/jcm11082265

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation (No. arXiv:2105.05537). arXiv. https://doi.org/10.48550/arXiv.2105.05537

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. Npj Digital Medicine, 3(1), 1–11. https://doi.org/10.1038/s41746-020-0288-5

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Deo, R. C. (2015). Machine Learning in Medicine. Circulation, 132(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Foth, C., Gangwani, M. K., Ahmed, I., & Alvey, H. (2024). Ventricular Tachycardia. In StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK532954/

Hlatky, M. A., & Mark, D. B. (2007). The high cost of implantable defibrillators 28(4), 388–391. https://doi.org/10.1093/eurheartj/ehl311

Kusumoto, F. M., Calkins, H., Boehmer, J., Buxton, A. E., Chung, M. K., Gold, M. R., Hohnloser, S. H., Indik, J., Lee, R., Mehra, M. R., Menon, V., Page, R. L., Shen, W.-K., Slotwiner, D. J., Warner Stevenson, L., Varosy, P. D., & Welikovitch, L. (2014). HRS/ACC/AHA Expert Consensus Statement on the Use of Implantable Cardioverter-Defibrillator Therapy in Patients Who Are Not Included or Not Well Represented in Clinical Trials. Heart Rhythm, 11(7), 1270–1303. https://doi.org/10.1016/j.hrthm.2014.03.041

Liu, K. A., & Mager, N. A. D. (2016). Women's involvement in clinical trials: Historical perspective and future implications. Pharmacy Practice, 14(1), 708. https://doi.org/10.18549/PharmPract.2016.01.708

Nanavati, H. D., Andrabi, M., Arevalo, Y. A., Liu, E., Shen, J., & Lin, C. (2024). Disparities in Race and Ethnicity Reporting and Representation for Clinical Trials in Stroke: 2010 to 2020. Journal of the American Heart Association, 13(6), e033467. https://doi.org/10.1161/JAHA.123.033467

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P., Patel, B. N., Yeom, K. W., Shpanskaya, K., Blankenberg, F. G., Seekins, J., Amrhein, T. J., Mong, D. A., Halabi, S. S., Zucker, E. J., … Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine, 15(11), e1002686. https://doi.org/10.1371/journal.pmed.1002686