Undergraduate Thesis Prospectus

Using Deep Learning to Segment Coronary Artery CTAs, Hypertrophic Cardiomyopathy MRIs, and Hydrocephalus Ventricle CT Scans

(technical research project in Biomedical Engineering)

The Response to Algorithmic Bias in Disease Diagnosis

(STS research project)

By

Wenxuan (Sharon) Zheng

December 11, 2020

Project Team Members

Ramiz Akhtar

Rohan Patel

Vignesh Valaboju

Wenxuan (Sharon) Zheng

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.

signed: ____ date: _________________

approved: ____ date: _________________

approved: ____ date: _________________

General Research Question - *How to improve American healthcare using AI?*

The healthcare industry would account for around 20% of the United States GDP, and the healthcare spending would increase to $4.01 trillion in 2020, estimated by a federal agency (Keehan et al., 2020). The enormous need and opportunities encourage the emergence of new technologies to improve the efficiency and quality of current healthcare systems and increase the accessibility of healthcare to individuals with difficulties. With the advancement of AI technology, it seems very likely that hospitals will be facilitated with automatic software in the future. Scientific research has indicated the AI can do better in gaming, creating entertainment, and can be as good as human experts in diagnosing diseases using medical images (Anwar et al., 2018).

The technological thesis focuses on using deep learning to segment coronary artery CTAs, hypertrophic cardiomyopathy MRIs, and hydrocephalus ventricle CT Scans. The project will utilize the state-of-art AI algorithm for medical image processing, aiming to replace the laborious manual segmentation with reliable software based on AI algorithms. The implementation of AI can alleviate the need for more healthcare professionals and increase efficiency at a low cost (Davenport & Kalakota, 2019). The healthcare quality can be significantly increased by AI, which also has the potential to be implemented on a large scale.

The successful implementation of new technologies requires the coproduction of science and social order, as higher efficiency due to technological advancements does not guarantee equality. The STS thesis will discuss the algorithmic bias in disease diagnosis and provide insights into future policymaking by identifying the source of algorithmic bias and discussing ways to encourage diversity.

Automatic Segmentation of Ventricular Volumes from CT Scans in Hydrocephalus Patients Using Convolutional Neural Network (CNN)

In the clinical workflow, medical images from computed tomography (CT) or magnetic resonance imaging (MRI) are commonly acquired in diagnosis. Most current clinical processes rely on qualitative interpretation of the images by radiologists. Quantitative evaluation, on the contrary, can improve the accuracy and reliability of diagnosis by extracting imaging biomarkers, which are measurable indicators of a disease state, to provide insights such as probability for disease progression, the intensity of disease presence, and other numerical features about the physiology and anatomy of the image (Kim, Park, Goo, Wildberger, & Kauczor, 2015). We identify three disease states of interest that require medical image analysis.

Coronary artery disease is classified by the narrowing of the arteries via a buildup of plaque, limiting the blood flow to the heart, and often leading to a heart attack. Coronary artery disease is the most common type of heart disease, killing 365,914 people in 2017, and affecting 18.2 million adults above the age of 20 in the United States ("Heart Disease Facts | cdc.gov"). Current identification of coronary artery disease mainly consists of manually examining CTAs (Computed Tomography Angiography) for the presence of calcification and plaque residues (Arnett et al., 2019). This approach of identification suffers from being too time-consuming; however, through automated CTA segmentation, the narrowing and calcification of the coronary arteries can be quantified in a timely manner.

Hypertrophic cardiomyopathy (HCM), a disease with a prevalence of 1 in 200, is characterized by unexplained enlargement of the heart's left ventricle (LV), misalignment of cardiac fibers, and cardiac scarring, all of which can lead to heart dysfunction (Kramer et al., 2015; Olivotto et al., 2011; Semsarian, Ingles, Maron, & Maron, 2015). Most HCM cases are genetically inherited and can gradually progress towards heart failure and cardiac arrest (Watkins, Ashrafian, & Redwood, 2011). Current identification methods for HCM include genetic testing and manual segmentation of cardiac magnetic resonance (CMR) images for unexplained left ventricular wall thickness and scarring (Wigle et al., 1985). Automating CMR image segmentation would allow for more efficient patient diagnosis and condition monitoring, ultimately improving personalized patient care.

Hydrocephalus affects 1 million people in the United States; it is very common in children as 1.1 in every 1000 infants will develop hydrocephalus (Tully & Dobyns, 2014). It is characterized by the buildup of fluid in the ventricles deep in the brain due to increased pressure which can damage brain tissues and cause severe disabilities (Hamilton, Gruen, & Luciano, 2016). Image segmentation, an essential toolkit for hydrocephalus diagnosis, is used to extract areas of brain ventricles. Automating image segmentation from the images can more efficiently estimate ventricular abnormality and can potentially distinguish between chronic and acute hydrocephalus.

In each of the three disease states discussed above, manual segmentation of the medical images is arduous and time consuming for physicians. The manual segmentation process also leads to inter-observer biases and causes poor diagnoses. Innovation through the creation of an automated process is required for improved segmentation efficiency. There have been several recent advances in using artificial intelligence to automatically segment medical images. Deep learning convolutional neural networks (CNNs) allow for image classification. Images can be fed through a series of convolutions to extract image features for classification. For segmentation, different regions of the images are classified.

The goal of the project is to **develop and validate deep learning-based segmentation algorithms for regions of interest in CT and MRI scans as well as implementing these algorithms into a medical software package.** Medical images for each of the aforementioned disease states will be segmented for their regions of interest using a CNN backbone. For coronary artery segmentation, innovative multi-channel inputs about the prior shape of blood vessels will be fed into the CNN. For HCM scar and LV wall segmentation, the goal will be to evaluate the patient population's generalizability of an improved segmentation algorithm. For hydrocephalic brain ventricle segmentation, most automatic segmentation algorithms are done using MRIs; however, we will segment CT scans for increased spatial resolution. Automated segmentation results will be compared against manual image segmentation as a metric of success.

Considering medical professionals must undertake a multitude of tasks, they must maintain high efficiency when completing any given task. Presenting the aforementioned automated segmentation algorithms to physicians in their raw form would create the potential for unintended modification of the algorithms, thus creating technical issues, reducing efficiency, and compromising patient care. To alleviate this concern, we also propose creating a universal software application to house and black-box our underlying automated segmentation algorithms. Creating a user interface (UI) that prioritizes simplicity, intuitive use, and extensive user control would grant physicians the power to see biomarker information in as much depth as needed while avoiding common problems, such as electronic medical record click fatigue, that plague physician interactions with software (Collier, 2018). Therefore, a software application for automated segmentation algorithms would make them more likely to see use in clinical and clinical training environments.

The Response to Algorithmic Bias in Disease Diagnosis

Coproduction of Science and Social Order

Artificial Intelligence (AI) has recently been intensely studied to solve complicated medical problems. The popularity of AI in healthcare roots from its potential to reduce human errors, which is estimated to be 5.08% in the United States and affect ~12 million adults every year (Singh, Meyer, & Thomas, 2014). On the other hand, scientists have proven that well-trained AI algorithms classify skin cancer at comparable accuracy to dermatologists (Esteva et al., 2017). They also believe that AI algorithms can improve the accuracy of diagnosis without any human bias, especially in diagnosis using radiology, where a robust AI algorithm can be built based on a large number of medical images.

However, AI has a long documented record of low diversity, and previous applications of AI in healthcare demonstrate the prevalence of algorithmic bias, which systematically creates unfair outcomes that cause unintentional harm, such as the exclusion of disadvantaged groups. For example, AI predictions of no-show appointments use models that consider the features of a patient. As a result, patients with lower income, pre-existing conditions, and addiction problems will be regarded as a low priority due to the high correlation with no-shows (McCullough, n.d.).

Therefore, in order to improve the excellence and equality of healthcare, the STS thesis will focus on the social responsibility to minimize algorithmic bias in healthcare using the framework coproduction of science and social order. Without adequate social order that regulates changes, AI algorithms can inherit prejudice from prior works and reflect or even exacerbate current bias in the society. The deterministic nature of AI can create more inequalities as the algorithm can be hard for the patients to communicate with, making underprivileged people more helpless unless new regulations are implemented.

The Demand for New Policy Regulations

Currently, there is a lack of regulation on AI-based algorithms used in the healthcare system. The most influential social force for the regulation of new technology implementation in the healthcare system is the Food and Drug Administration (FDA), which is responsible for the regulation of anything related to public health, including drugs, biological, and medical devices. Any “software intended to be used for one or more medical purposes without being part of a hardware medical device” will be defined as a medical device by the FDA, including the software used for diagnosis, treatment, and prevention. However, current FDA regulations on medical devices do not apply to AI-based software. Therefore, the first FDA proposal in 2019 on AI-based software introduces a Total Product Life Cycle (TPLC) to access the quality and organizational quality of the company to have reasonable assurance of high-quality AI-based software development, testing, and performing monitoring.

Until recently, only 29 AI-based software was approved by the FDA for the application of disease diagnosis that is mostly focused on radiology, cardiology, and general medicine (Benjamens, Dhunnoo, & Meskó, 2020). This software fulfills the requirements of quality required by the FDA, but trials that ensure equal treatments for the patients are not required. In order to promote equality of AI healthcare, FDA has to make a few changes, and new policies are required to accommodate the technological and social change.

The prerequisite of new policies on diagnostic AI algorithms will include identifying potential sources for algorithmic bias and recognizing ways to promote diversity. Discussions in this prospectus focus on similar applications in healthcare using AI, which will lay the foundation for the more in-depth analysis in the context of disease diagnosis in the thesis paper.

Potential Source of Algorithm Bias

First of all, from a scientific perspective at an early development stage, AI algorithms can create bias because they are based on pre-existing training data, which means that if there is any bias in the data of the scientists' choice, the algorithms would reflect the bias when facing a patient. Especially when the scientific experiments are mostly concerned with isolated variables, the generalizability in a social setting is concerning due to the choice of scientists.

Secondly, clinical trials can be an equality killer when they mostly focus on the privileged group. One example is the gender imbalance in the trials for cardiovascular diseases, which comprises a population that is 85% male and mostly postmenopausal females (Dougherty, 2011). The underlying reason is the fear of the disruption of standardized results by female menstruation cycle, and the intentional uneven distribution of male and female has negative impacts on disease treatments for females, where some of them are more likely to receive lower levels of treatment for cardiovascular diseases (Bugiardini, Estrada, Nikus, Hall, & Manfrini, 2010).

Finally, an essential factor that contributes to unequal healthcare is biological. Multiple studies have shown the genetic differences that result in an increase in risk factors for minorities. One report suggests that the similarity of traits linked to asthma in European Americans and African Americans is only 5%, which makes African American children 10 times more likely to die from asthma compared to non-Hispanic white children (White et al., 2016).

The aforementioned sources may contribute to the algorithm bias in disease diagnosis. Therefore, the thesis will explore the sources of bias specifically in the context of the approved diagnostic AI algorithms. Focused on a few representative cases, the thesis will look at the choice of datasets by scientists during algorithm development, compare available open-source databases with current demographics, and determine the variances among different demographics within each diagnostic application. Factors that are specific to disease diagnosis using AI will be identified, and the main questions to address include: (1) what datasets did the scientists choose, (2) which demographics are most underrepresented in the currently available database, and (3) to what extent does the underrepresented group different from focused groups?

Encouraging Greater Diversity

AI algorithms need a diverse deep learning experience, which can be promoted in a few ways. First is the inclusion of a diverse group of people and increasing the awareness of diversity during the development stage. Nowadays, most people in AI-related fields are males - 80% of AI professors are males, and 80%-90% of the staff in big AI technology companies are also male - and very few are minorities ("Gender, Race, and Power in AI"). More diversities in the development team can bring a more comprehensive worldwide perspective to the algorithm to prevent any bias.

Secondly, when designing clinical trials, researchers should include a diverse population of study participants, recruit participants from different practice settings, and collect data on a broad range of health outcomes (Tunis, Stryer, & Clancy, 2003). The cutting-edge biomedical research should not happen in isolation, and more resources should be allocated to translational research to encourage collaboration to effectively decrease the disparity created by independent scientific research based on a homogenous population.

Finally, a diverse collection of algorithms and a group of physicians should be encouraged to work together during the implementation phase. The different results by multiple algorithms can help scientists identify aspects that have been previously overlooked and provide significant feedback for improvements. Additionally, Identifying the formation of algorithmic bias is not only important for creating equality, but also to provide insights to how physicians and AI algorithms can work together. There can be cases where AI algorithms make most of the decisions for the majority of the privileged population, where physicians can be more focused on interventions for the minority population.

The goal of the thesis is to study how to encourage diversity in the context of diagnostic AI algorithms. For the previous rechosen presentative cases, the thesis will look at the authors for online publications related to each case, datasets used for clinical trials, and compare the intra- and inter-variability of algorithms and physicians for each diagnostic application. There three major questions include (1) what is the demographic of the development team, (2) what is the demographics of the testing datasets, and (3) what is the strength or weakness of each algorithm in terms of treating a diverse population?

The Future of Healthcare Policy Making

No algorithm can solve a problem that has been rooted in human society, but it has the potential to make the best judgment based on the highest level of knowledge on the constantly evolving social sciences, laws, and ethics.

The future policy can follow the guidance from the source of algorithmic bias and ways to improve diversity by emphasizing varieties of datasets during development and testing, differences of individual patients, diversion of demographic backgrounds within a team, and corporations between multiple entities.

The goal of policies is not to impede AI development in disease diagnosis, but to encourage more creations that promote equality and fix the current AI diversity crisis. If we know that certain algorithms favor the mainstream population, we should still use the algorithm, but only on the population it favors.

Conclusions

New technical developments offer solutions to current challenges as well as causes of social changes. This thesis incorporates the co-production of science and social order and aims to contribute to the successful implementation of AI in the healthcare industry. In the technical report, the implementation of AI in the automatic segmentation of 3 different applications - analyzing myocardial scarring in hypertrophic cardiomyopathy patients, diagnosing coronary artery disease, and segmenting ventricle area for the application of hydrocephalus - indicates the potential of AI in other applications in the healthcare industry. Meanwhile, the social dimensions of technological advancement can facilitate development, such as identifying new sources of algorithmic bias and determining methods to promote diversities. In conclusion, AI is the potential solution to current challenges in the healthcare industry, and social accommodation is part of the solution for the successful implementation.

Bibliography

Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., & Khan, M. K. (2018). Medical Image Analysis using Convolutional Neural Networks: A Review. *Journal of medical systems*, *42*(11), 226.

Arnett, D. K., Blumenthal, R. S., Albert, M. A., Buroker, A. B., Goldberger, Z. D., Hahn, E. J., Himmelfarb, C. D., et al. (2019). 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: A report of the american college of cardiology/american heart association task force on clinical practice guidelines. *Circulation*, *140*(11), e563–e595.

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*, *3*, 118.

Bugiardini, R., Estrada, J. L. N., Nikus, K., Hall, A. S., & Manfrini, O. (2010). Gender bias in acute coronary syndromes. *Current vascular pharmacology*, *8*(2), 276–284.

Collier, R. (2018). Rethinking EHR interfaces to reduce click fatigue and physician burnout. *Canadian Medical Association Journal*, *190*(33), E994–E995.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), 94–98.

Dougherty, A. H. (2011). Gender balance in cardiovascular research: importance to women's health. *Texas Heart Institute Journal*, *38*(2), 148–150.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.

Gender, Race, and Power in AI. (n.d.). .

Hamilton, M., Gruen, J. P., & Luciano, M. G. (2016). Introduction: Adult hydrocephalus. *Neurosurgical Focus*, *41*(3), E1.

Heart Disease Facts | cdc.gov. (n.d.). . Retrieved December 11, 2020, from https://www.cdc.gov/heartdisease/facts.htm

Keehan, S. P., Cuckler, G. A., Poisal, J. A., Sisko, A. M., Smith, S. D., Madison, A. J., Rennie, K. E., et al. (2020). National Health Expenditure Projections, 2019-28: Expected Rebound In Prices Drives Rising Spending Growth. *Health Affairs (Project Hope)*, *39*(4), 704–714.

Kim, H., Park, C. M., Goo, J. M., Wildberger, J. E., & Kauczor, H.-U. (2015). Quantitative computed tomography imaging biomarkers in the diagnosis and management of lung cancer. *Investigative Radiology*, *50*(9), 571–583.

Kramer, C. M., Appelbaum, E., Desai, M. Y., Desvigne-Nickens, P., DiMarco, J. P., Friedrich, M. G., Geller, N., et al. (2015). Hypertrophic Cardiomyopathy Registry: The rationale and design of an international, observational study of hypertrophic cardiomyopathy. *American Heart Journal*, *170*(2), 223–230.

McCullough, C. (n.d.). Who is least likely to attend? An analysis of outpatient appointment DNA data in NHS Greater Glasgow & Clyde.

Olivotto, I., Girolami, F., Sciagrà, R., Ackerman, M. J., Sotgia, B., Bos, J. M., Nistri, S., et al. (2011). Microvascular function is selectively impaired in patients with hypertrophic cardiomyopathy and sarcomere myofilament gene mutations. *Journal of the American College of Cardiology*, *58*(8), 839–848.

Semsarian, C., Ingles, J., Maron, M. S., & Maron, B. J. (2015). New perspectives on the prevalence of hypertrophic cardiomyopathy. *Journal of the American College of Cardiology*, *65*(12), 1249–1254.

Singh, H., Meyer, A. N. D., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ quality & safety*, *23*(9), 727–731.

Tully, H. M., & Dobyns, W. B. (2014). Infantile hydrocephalus: a review of epidemiology, classification and causes. *European Journal of Medical Genetics*, *57*(8), 359–368.

Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *The Journal of the American Medical Association*, *290*(12), 1624–1632.

Watkins, H., Ashrafian, H., & Redwood, C. (2011). Inherited cardiomyopathies. *The New England Journal of Medicine*, *364*(17), 1643–1656.

White, M. J., Risse-Adams, O., Goddard, P., Contreras, M. G., Adams, J., Hu, D., Eng, C., et al. (2016). Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. *Immunogenetics*, *68*(6–7), 391–400.

Wigle, E. D., Sasson, Z., Henderson, M. A., Ruddy, T. D., Fulop, J., Rakowski, H., & Williams, W. G. (1985). Hypertrophic cardiomyopathy. The importance of the site and the extent of hypertrophy. A review. *Progress in Cardiovascular Diseases*, *28*(1), 1–83.