Optimization of Probiotic Manufacturing with Computational Modeling

(Technical Paper)

A Policy Analysis of Regulations on Artificial Intelligence for Diagnosis of Disease

(STS Paper)

A Thesis Prospectus Submitted to the

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

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

Samantha Clayton
Fall, 2019

Technical Project Team Members
Caroline Bereuter
Lily Lin

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

Introduction

Most Americans will receive a medical misdiagnosis at least once in their lifetime (Rue, 2019). Artificial intelligence (AI) has the potential to change the rate of misdiagnoses. There has been significant research over the past few decades into using artificial intelligence for medical and healthcare applications. One focus of this development is using artificial intelligence systems to diagnose diseases. Diagnostic errors have contributed to approximately 10% of patient deaths in the U.S. AI could be utilized to help with diagnosis of diseases by revealing previously hidden trends in data, and thus has the potential for substantial impact both at the individual patient and system level (Panch, Szolovits, & Atun, 2018). However, these systems prompt concerns related to the privacy of patient data, the quality and safety of these algorithms, and their impacts on the role of physicians and the healthcare system at large. The proposed STS paper aims to analyze the regulations surrounding artificial intelligence for diagnosis of diseases and their impacts on the concerns listed previously.

As exemplified by AI's potential benefits for healthcare applications, computational approaches can provide significant advantages to a variety of fields. Computational modeling of metabolomic data was designed to evaluate the metabolic reactions of individual bacterial strains to test yield under different conditions. Bacteria production is important to a number of applications including probiotics and nutrition. Approaches to treat malnutrition are currently limited by the cost and scalability of manufacturing and later combining individual microbial bacterial strains to approximate the complexity of the microbiome. Co-culturing these microbial bacteria could lower costs, be more robust to environmental variations, and be more resistant to invading pathogens. However, these advantages are not universal to all combinations, thus a technique must be devised to find the best combinations and conditions. The proposed technical project aims to produce a computational pipeline to predict optimal combinations of species with metabolomic data. This paper examines the impact of computational approaches to health problems, focusing specifically on the development of a computational framework to optimize co-cultures for the manufacturing of biotherapeutics for nutrition and the impacts and regulation of artificial intelligence applied to the diagnosis of diseases.

Technical Topic

In 2018, 21.9% of children under the age of five were affected by stunting, defined as below minus two standard deviations from median weight for height of reference population, and an additional 7.3% of children were affected by wasting, defined as below minus two standard deviations from median height for age of reference population (UNICEF, n.d.). These afflictions are the result of poor nutrition, beginning in utero and continuing into early childhood (United Nations Children's Fund (UNICEF), World Health Organization, & The World Bank, 2019). Malnutrition is considered a pressing global health concern, a contributing factor for almost half of all childhood deaths (Walson & Berkley, 2018), and an agent in developmental impediments. Gut dysfunction and altered gut microbiota have been linked to clinical outcomes of infantile malnutrition (Subramanian et al., 2014). Thus, nutritional rehabilitation therapies have been developed to help restore a healthy gut microbiome.

Current therapies have proven to be ineffective for sustainable growth. These therapies involve the administration of probiotics, which are most often composed of only a single bacterial species that is grown in a bioreactor and freeze-dried before it is delivered to a human (Fenster et al., 2019). A promising new strategy is the transfer of multi-species probiotics that consist of live gut microbes that can restore the gut microbiome. Although these multi-species probiotics could serve as a potential replacement for current solutions, they have not yet been scalable for large scale administration due to the complexity of the metabolic interactions associated with co-culturing multiple bacterial strains. To become feasible for large-scale administration, new strategies must be employed to increase manufacturing yield of these human gut bacteria (O'Toole, Marchesi, & Hill, 2017). This technical project aims to predict optimal combinations of strains of gut microbes, leveraging metabolic data to understand how strains interact. Funding for the project is provided by a Bill & Melinda Gates Foundation grant.

As a strategy to improve the yield of manufacturing human gut bacteria, Caroline, Lily, and I will develop a computational pipeline that can be applied to various probiotic strains to develop metabolic models, and then use these models to determine optimal strain combinations. The timeline for the generation of the computational framework is illustrated by Figure 1. The

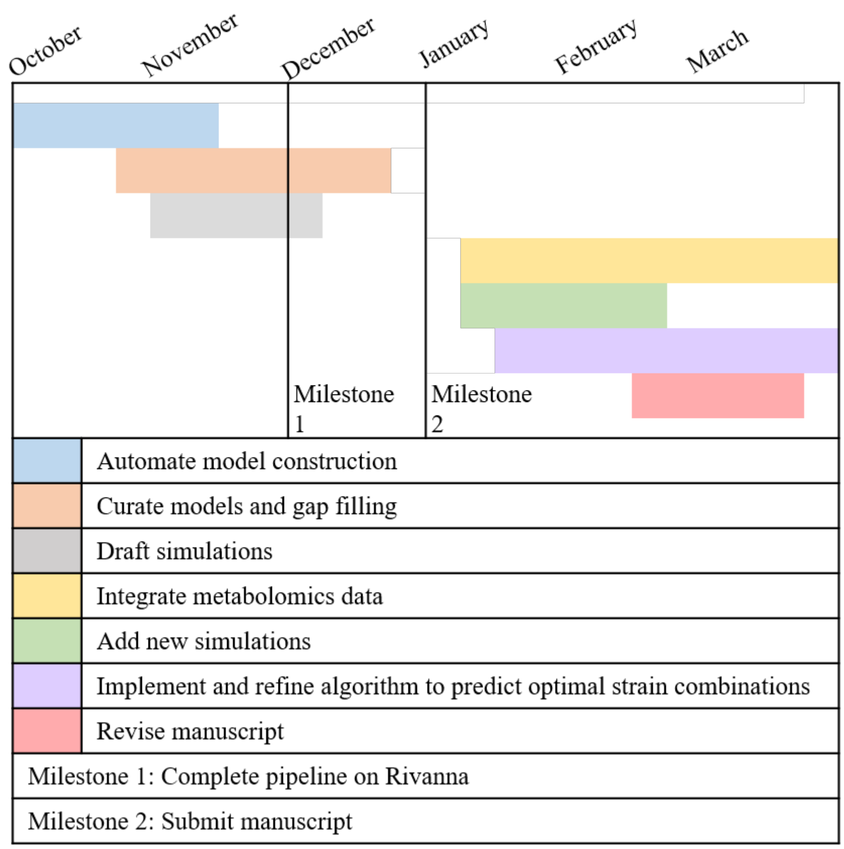

Figure 1: Timeline of Capstone Project: The timeline depicts the schedule for completion of the creation of the models, simulations, integration of metabolomic data, and generation of predictions (Created by Bereuter, Clayton, & Lin, 2019).

pipeline will generate metabolic models for each strain, then refine these models with metabolomic data. The metabolomic data is currently being collected by our technical advisor, Dr. Gregory Medlock, from experimental analysis of 10 different probiotic strains from human gut-derived microbes. Following creation and refinement of the metabolic models, nutrient preferences and secreted products will be simulated. These computational predictions of nutrient preferences and metabolites will be integrated into a machine learning-based process to predict optimal species combinations. We will be using a novel algorithm, Growth Optimization by Packing Metabolomes (Go-PacM), developed by our technical advisor, to predict optimal co-cultures. To validate our findings, we will grow microbes in both optimal and random combinations and evaluate yield to measure the effectiveness of our predictions.

Accomplishing these aims will serve as a way to predict the optimal combinations of any bacterial strains to be used in the development of probiotics. The optimal combinations of bacterial strains, generated from this computational framework, can be co-cultured in order to improve biomass yield. This co-culturing process will serve to lower costs by reducing nutrients required per batch and improve scalability by utilizing a more robust combination of strains. The more efficient process of co-culturing in manufacturing with gut microbial strains will help to make live biotherapeutic strategies more accessible to those in need.

STS Topic

Artificial intelligence (AI) is defined as "a broad discipline that aims to understand and design systems that display properties of intelligence, emblematic of which is the ability to learn, or derive knowledge from data" (Panch et al., 2018). In 1956, the development of a program mimicking human problem solving called The Logic Theorist, widely considered the first AI program, showed the feasibility of AI systems (Anyoha, 2017). AI has been researched for a variety of applications ranging from automated assistants to facial recognition software. Artificial intelligence in medicine dates back to 1972 with Stanford University's MYCIN system, an AI prototype program used to identify bacteria causing blood infections and recommen antibiotics (Shortliffe, 1974). In medicine alone, artificial intelligence has been researched to apply to everything from interpreting images for radiology, creating management plans for patient care, and supporting the diagnosis of diseases.

1 in 20 adults experience a diagnostic error every year, with about half of these errors being potentially harmful (National Academies of Sciences, Engineering, and Medicine, 2015). The large amounts of misdiagnoses due to human error opens the door for the use of AI for diagnosis. This technology also has the potential to significantly reduce costs. In 2017, analysts predicted that AI applied for preliminary and automated image diagnosis could save $8 billion by 2026 from healthcare spending (HealthITAnalytics, 2019). AI use for diagnosis has been the focus of a significant amount of venture funding, with 19 private digital health companies receiving $330.4 million from 2011-2017 (Zweig & Tran, 2018). AI research has focused on diagnosing a variety of diseases from various types of cancers to pneumonia (Rajpurkar et al., 2017). In a review of studies, AI programs performed as well as, or slightly better, than human healthcare professionals overall (Liu et al., 2019). With more research into the advancement of the technology over time, the accuracy of these systems can improve even more.

However, there are several possible drawbacks associated with AI's use in this way that impact and affect the stakeholders of this technology. These stakeholders are the patients, the physicians, the scientists and engineers, and the politicians and policy makers. The patients will be the focus of this advancement of technology for care and thus their safety and privacy must be considered a priority. These systems would require access to significant amounts of data, which would have to be provided from patients for use in not only their own examinations but to assist with the development of the technology (Longhurst, Harrington, & Shah, 2014). This use of patient data may in part conflict with HIPAA regulations and need to be examined further (Longhurst et al., 2014). There has also been research into how secure these systems are that showed that they are susceptible to adversarial attacks, which could put patient data further at risk and call in question the results provided by these systems (Finlayson et al., 2019). Additionally, this application could significantly change the day-to-day work of a physician and the required skills they would need to be successful. Physicians would require significant training on similar technology so that they could use it effectively without errors (Cabitza, Rasoini, & Gensini, 2017). The development of these systems by scientists and engineers could introduce their implicit biases into the system, either through their design or the choice of datasets, which could put minority communities at risk for worse treatment (Crigger & Khoury, 2019; Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

I will apply the STS framework of political technologies to my analysis of artificial intelligence for diagnosis of disease. The theory of political technologies posits that there are two ways artifacts are political, either by being developed to settle an issue within a community or by correlating with political relationships (Winner, 1980). The latter group is referred to as inherently political technologies. Applications of this theory have been critiqued on the basis of whether the technologies described are in fact political or if they are being misjudged (Joerges, 1999). Under this framework, I will treat this application of artificial intelligence as an inherently political technology due to its impact on the power relationships within the healthcare system and risk of further disadvantaging minority groups.

Research Question and Methods

This thesis will seek to answer where is the need for future policies surrounding artificial intelligence for diagnosis to protect patient privacy while providing the best possible care.
I will use documentary research methods and policy analysis to understand scientific, medical, political, and public perceptions of this technology. Documentary research methods will involve an extensive literature review to assess the current state of this technology, the possible future benefits of its adoption, and the potential dangers that scholars have noted. This research will include looking at databases, such as PubMed, Web of Science, Congress.gov, and Public Affairs Information Service International (PAIS Index). I will also look at journals, such as the AMA Journal of Ethics, Health Affairs, and Nature, to analyze perspectives, progress and policies on artificial intelligence use for diagnosis. Policy analysis will be applied by reviewing current and proposed policies surrounding artificial intelligence applied to medical purposes. Among others, this will include current policies, such as Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the American Medical Association's (AMA's) policy on Augmented Intelligence released in 2018 (Crigger & Khoury, 2019; Office for Civil Rights, 2009). Additional policies general to the use of artificial intelligence, such as the proposed Algorithmic Accountability Act of 2019 and the California Consumer Privacy Act, will also be reviewed to discern how other applications of this technology have been dealt with and if these policies could be effective for regulating artificial intelligence for diagnosis (Clarke, 2019; Palsan, 2018). This research and analysis will take place during the Spring 2020 semester.

Conclusion

At the end of the academic year, the goal for the technical project is to have a computational pipeline that can predict the optimal bacterial species combinations and media conditions. A manuscript will be submitted for peer review following the successful creation of this pipeline. This pipeline will be able to aid in the manufacturing of a microbiome that is more cost-efficient, scalable, and robust to assist in the treatment of malnutrition. This more efficient process of manufacturing with gut microbial strains will help to make live biotherapeutic strategies more accessible to those in need.

The completion of the STS thesis paper at the end of this academic year will explain the positive and negative potential impacts of the use of artificial intelligence for diagnosis of disease and analyze the effects of current policies and regulations on its development and implementation. The completed thesis will also propose additional policies that would consider each of the stakeholders of this technology with a primary focus on the protection of the patients.

References

Anyoha, R. (2017, August 28). The history of artificial intelligence. Retrieved October 30, 2019, from Science in the News website: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA*, *318*(6), 517–518. https://doi.org/10.1001/jama.2017.7797

Clarke, Y. D. (2019, April 10). H.R.2231 - 116th Congress (2019-2020): Algorithmic Accountability Act of 2019 [Webpage]. Retrieved October 30, 2019, from https://www.congress.gov/bill/116th-congress/house-bill/2231

Crigger, E., & Khoury, C. (2019). Making policy on augmented intelligence in health care. *AMA Journal of Ethics*, *21*(2), 188–191. https://doi.org/10.1001/amajethics.2019.188.

Fenster, K., Freeburg, B., Hollard, C., Wong, C., Rønhave Laursen, R., & Ouwehand, A. C. (2019). The Production and Delivery of Probiotics: A Review of a Practical Approach. *Microorganisms*, *7*(3). https://doi.org/10.3390/microorganisms7030083

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, *363*(6433), 1287–1289. https://doi.org/10.1126/science.aaw4399

HealthITAnalytics. (2019, July 29). Future AI opportunities for improving care delivery, cost, and efficacy. Retrieved October 24, 2019, from HealthITAnalytics website: https://healthitanalytics.com/news/future-ai-opportunities-for-improving-care-delivery-cost-and-efficacy

Joerges, B. (1999). Do Politics Have Artefacts? *Social Studies of Science*, *29*(3), 411–431. https://doi.org/10.1177/030631299029003004

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., … Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, *1*(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Longhurst, C. A., Harrington, R. A., & Shah, N. H. (2014). A ‘green button’ for using aggregate patient data at the point of care. *Health Affairs*, *33*(7), 1229–1235. https://doi.org/10.1377/hlthaff.2014.0099

National Academies of Sciences, Engineering, and Medicine. (2015). *Improving Diagnosis in Health Care*. Retrieved from https://doi.org/10.17226/21794

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

Office for Civil Rights. (2009, November 20). Summary of the HIPAA Security Rule [Text]. Retrieved October 24, 2019, from HHS.gov website: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html

O’Toole, P. W., Marchesi, J. R., & Hill, C. (2017). Next-generation probiotics: The spectrum from probiotics to live biotherapeutics. *Nature Microbiology*, *2*(5), 17057. https://doi.org/10.1038/nmicrobiol.2017.57

Palsan, S. (2018). The California consumer privacy act: Getting a head start on compliance. Retrieved October 30, 2019, from Ogletree Deakins website: https://ogletree.com/insights/2018-10-08/the-california-consumer-privacy-act-getting-a-head-start-on-compliance

Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of Global Health*, *8*(2). https://doi.org/10.7189/jogh.08.020303

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., … Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv:1711.05225 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1711.05225

Rue, N. (2019, January 30). AI and healthcare: The battle against misdiagnosis. Retrieved October 24, 2019, from Medium website: https://becominghuman.ai/ai-and-healthcare-the-battle-against-misdiagnosis-7eb3742ad1f7

Shortliffe, E. H. (1974). *MYCIN: A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection.* (No. AIM-251). Retrieved from Stanford Univ., Calif. Dept. of Computer Science website: https://apps.dtic.mil/docs/citations/ADA001373

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M. A., … Gordon, J. I. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*, *510*, 417. Retrieved from https://doi.org/10.1038/nature13421

UNICEF. (n.d.). Definitions: Nutrition. Retrieved October 31, 2019, from UNICEF website: https://www.unicef.org/infobycountry/stats_popup2.html

United Nations Children's Fund (UNICEF), World Health Organization, & The World Bank. (2019). *Levels and trends in child malnutrition: Key findings of the 2019 Edition of the Joint Child Malnutrition Estimates.*

Walson, J. L., & Berkley, J. A. (2018). The impact of malnutrition on childhood infections. *Current Opinion in Infectious Diseases*, *31*(3). Retrieved from https://journals.lww.com/coinfectiousdiseases/Fulltext/2018/06000/The_impact_of_malnutrition_on_childhood_infections.5.aspx

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136. Retrieved from https://www.jstor.org/stable/20024652

Zweig, M., & Tran, D. (2018). The AI/ML use cases investors are betting on in healthcare. Retrieved October 22, 2019, from Rock Health website: https://rockhealth.com/reports/the-ai-ml-use-cases-investors-are-betting-on-in-healthcare/