I first developed an interest in Biostatistics while working with Professor Brian Caffo on volumetric analysis of MRIcloud output. By that time, I had completed enough statistics and computer science courses to be able to meaningfully contribute to the “MRIcloudT1volumetrics” package he was building. With a goal to create tools that can increase researcher productivity and improve package quality, I joined the development team and have improved classification models using principal component analysis (PCA) and multinomial regression. Also, I built functions visualizing weight of loadings on each principal component to make analyzing data much easier. This package may help find new biomarkers of aging. Through this experience, I have strengthened my statistics skills and become well-versed in good research culture and practice.

Most of my real-world machine learning experience was gained during my internship at Johns Hopkins Bloomberg School of Public Health. I worked in the Wearable and Implantable Technology (WIT) research group and was there introduced to Professor Ciprian M. Crainiceanu. One of our most interesting analyses has been the National Health and Nutrition Examination Survey (NHANES). I led a research project addressing a problem intrinsic to the high-dimensional nature of our data, which would complicate analyses in time series forecasting. Using cross-validation and ANOVA test, my regression analysis results and model revealed important associations between daily activity and health outcomes such as BMI and mortality. Ultimately, I submitted a first-author full paper to the American Journal of Epidemiology on this work, which is currently under review [1].

My motivation to pursue a PhD comes from my experience working on several substantial research projects. So far, I’ve worked in groups at North Eastern University, Hong Kong Polytechnic University, and Johns Hopkins Bloomberg School of Public Health. Through these experiences, I have improved my research skills and published several first-author papers [2,3]. I have found that I am most engaged when working on ambitious data science projects. Although my actual data engineering work has been educational, what really sparked my interest in research was experiencing the first-hand challenges that data engineers face daily. I am interested in projects which, if successful, have the potential to impact a large number of people by improving the state of the art in a particular area.

I am currently studying Biomedical Engineering at Johns Hopkins University as a master’s student, doing research under Dr. Brian Caffo. My objective is to pursue a PhD in Biostatistics with a focus on applied machine learning and data science. My career aspiration is to lead a research laboratory in industry after completing my PhD, where I will strive to make usable, effective technology with a positive societal impact. Although I am open to a variety of research, there are several professors at Johns Hopkins University whose projects are especially appealing to me: Brian Caffo (time series analysis), Ciprian M. Crainiceanu (WIT), and Scott L. Zeger (Bayesian statistics). After reading several papers in each of these groups, I see a clear fit for my skills and interests at Johns Hopkins University and am confident that it is a great place for me to pursue a PhD.

Thank you for your consideration of my application.

References

[1]

[2]

[3]