



The Ohio State University Department of Biomedical Informatics Summer Internship Program

About the Program

The Department of Biomedical Informatics hosts an annual in-person internship program each summer, which provides opportunities for current undergraduate and graduate students to pursue research projects in the field of biomedical informatics under the guidance of research and operational staff and renowned faculty mentors. Participants learn useful tools and technologies used in biomedical and clinical research and attend weekly seminars to learn more about the various fields and interdisciplinary interactions biomedical informatics facilitates. Interns also gain experience in scientific presentation through regular lab and programmatic presentation opportunities, including an end-of-program poster session.

Many students who have participated in this program have gone on to pursue doctoral degrees in biomedicine, biomedical informatics, computer science, and electrical engineering, or to complete degrees in data, medicine, or the health sciences.

All student internship positions in the Department of Biomedical Informatics are paid, with the hourly wage depending on student experience levels. In some cases, students may have the opportunity to continue employment in the lab into the academic year.

This year the program will run from May 20, 2024 to August 9, 2024.

Please contact internship coordinator, Gabrielle Kokanos, gabrielle.kokanos@osumc.edu, with questions.

- Interns are expected to work in-person but will be able to work remotely on Fridays.
- There are weekly program sessions in which all interns are expected to attend and participate, prioritizing this time as a cohort.

- All interns will prepare a research poster and present their findings during a poster session in the last week of the internship.
- All interns will be able to work up to 38 hours a week, this includes times at the weekly program sessions. Interns may work less than 38 hours if they are registered for courses at OSU (28 hours/week limit) or choose to work fewer hours due to other commitments.
- *Undergraduate students will be paid \$13/hour and graduate students will be paid \$14/hour*

2024 Projects

A. Courtney Hebert, MD, MS

--seeking 1-2 interns--

The Hebert lab focuses on the secondary use of electronic health record data to improve care and help with clinical decision making, specifically in the area of infectious diseases. Interns will be involved in projects which include cleaning data, analyzing data, and working within interdisciplinary teams.

Targeted Skills: Familiarity with basic statistics and programming and an interest in clinical research or health sciences is recommended.

B. Lijun Cheng, PhD

--seeking 1-2 interns--

Precision medicine, including gene therapy and gene editing technology, is rapidly advancing across the health landscape, as reflected in the growing research community in drug development, and increased public awareness. Patients along with their physicians are hopeful and eager to welcome new, innovative therapies that can potentially bring about unprecedented clinical success in genetic/genomic diseases for which there have been no effective treatments. The research in Dr. Cheng's lab focuses on computational precision medicine to identify optimal targets for drugs. She has developed several computational tools to elucidate the molecular action of drugs. Dr. Cheng hopes to recruit 1-2 students to continue to work in the precision medicine field. **Targeted Skills:** Students should be competent programming in Python, R, or MATLAB coding

C. Anjun Ma, PhD

--seeking 1 intern--

Potential student can select one or multiple projects:

Project 1: Data integration and analysis for diffuse large B cell lymphoma single-cell omics data.

Project 2: Develop an online platform for patient omics and manufacturing data collection and display.

Project 3: Single-cell transcriptomics analysis visualization for GPT prompting.

Targeted Skills: Potential students should be familiar with either R or Python, and have experience in using GitHub, high-performance computing, and the Linux environment.

D. Fode Tounkara, PhD

--seeking 1-2 interns--

This project focuses on addressing challenges in analyzing high-dimensional data (HDD), prevalent in fields such as genomics, biomedical, and clinical research. The primary aim is to develop innovative statistical strategies to handle datasets where the number of variables surpasses the sample size. We will explore and improve upon existing penalized regularization methods for variable selection and estimation. The project will specifically tackle the identification and quantification of sparse, strong, and weak-moderate signals in HDD, aiming for a realistic and effective framework in model selection and post-estimation processes. **Targeted Skills:** Basic to moderate statistical skills; Linear regressions and familiarity with regularization methods such as Lasso, Ridge, or Elastic Net; Coding experience in R; Good standard writing skills; Ability to work effectively as part of a team.

E. Qin Ma, PhD

--seeking 1 intern--

Potential students should expect to be developing a bioinformatics software package that utilizes in-house algorithms, methods, or deep learning models related to the single-cell data analysis field. Rather than developing a novel algorithm for a particular biological question, the focus of the project would be implementing modern software development practices to an application and extension of one of the Ma lab's published methods. The final product could be either an R or Python package, followed by maintainable code, well-documented tutorials, user-friendly APIs, and intuitive visualization functions. **Targeted Skills:** Potential students should be familiar with at least one of the following: R, Python, GitHub, high-performance computing, and the Linux environment. Special skills may be required for the specific project, such as package or library development experiences. The students are encouraged to have a fundamental knowledge of biology or microbiology.

F. Xia Ning, PhD

--seeking 1-2 interns--

Large language models (LLMs) have been well-developed for applications that involve natural languages. For example, ChatGPT, which is trained over natural languages, can generate new texts for conversation purposes. Recently, LLMs have been applied and adapted to design new biological sequences (e.g., protein sequences, peptide sequences) using the “biological languages.” This project will explore how LLMs can be used for protein sequence generation and protein structure generation using protein amino acid sequences, and will leverage and compare with alphafold for protein design. **Targeted Skills:** Proficiency in python (pytorch) programming, good communication skills. Biology background is preferred but not required.

G. Yuan Xue, PhD

--seeking 2 interns--

Possible projects include:

1. Digital pathology image analysis including synthesis, diagnosis, and quality assurance. Students are expected to learn and help implement the latest neural networks including foundation generative models.
2. Infant movement video motion analysis. Results will be used for early detection cerebral palsy and prediction of infant’s motor development.

Targeted Skills: Potential students should have basic knowledge about machine learning, deep learning, and programming such as Python. A background in computer vision and medical image analysis is highly beneficial. Additionally, candidates with experience in other areas of artificial intelligence, including natural language processing, multi-modality learning, and generative models, are encouraged to apply.

H. Xiaoli Zhang, PhD

--seeking 1-2 interns--

Immunotherapy has showcased remarkable advancements in cancer treatment over recent decades. Despite the improvement, not all patients exhibit positive responses to immunotherapy, and a subset may experience immunotherapy-related adverse events (irAEs) during treatment. Hence, a comprehensive understanding of the factors and mechanisms contributing to treatment failures and irAEs is paramount for enhancing treatment efficacy and mitigating adverse effects. This project aims to undertake a

thorough literature review and/or leverage real-world patient electronic health record (EHR) data to explore the impact of patient clinical factors and molecular biomarkers on immunotherapy response and development of irAEs. Through this investigation, we seek to uncover crucial insights into potential mechanisms underlying patient resistance to immunotherapy and identify strategies for broadening the scope of patients who can benefit from this innovative treatment approach. **Targeted Skills:** Potential students should have basic training of programming in SAS, R, or other languages.

I. Qing Wu, PhD

--seeking 3 interns--

Project 1: We are excited to announce an innovative research project to revolutionize medical meta-analysis by applying advanced machine learning and AI technologies. This project involves developing a sophisticated GPT-based model to automate and enhance the process of conducting meta-analyses in medical research. The primary objective is to create an AI tool capable of efficiently searching, summarizing, and abstracting data from various medical research papers, thereby significantly improving the speed and accuracy of meta-analyses. As a graduate student working on this project, you will be crucial to its success. Your responsibilities include meticulously identifying and analyzing existing meta-analysis papers to extract relevant data and insights. This data will form the backbone of our training set for the machine learning model. You will work closely with a team of AI, data science, and medical research experts, gaining hands-on experience in cutting-edge AI applications in healthcare, data processing, and analysis. This project offers an opportunity to contribute to a significant advancement in medical research methodology. It provides a platform for you to develop your AI and data science skills in a real-world, impactful context.

Project 2: We are developing a new machine learning-based PFRAX model to improve fracture risk prediction compared to the standard FRAX tool. FRAX estimates a 10-year risk of major osteoporotic and hip fractures to guide treatment decisions but has limitations we aim to address with PFRAX. Our interdisciplinary team is incorporating novel risk factors like fall history and social determinants of health into advanced machine learning algorithms. We will validate PFRAX and compare its performance to FRAX using the rich cohort data and outcomes in the Women's Health Initiative. This project offers an excellent opportunity to gain experience applying cutting-edge machine learning methodology to solve an impactful medical prediction challenge. As a graduate student intern, you would support the validation of model performance, subgroup analyses, and direct comparisons of modeling approaches.

You will learn highly relevant skills in health data science, predictive analytics, and optimal communication of technical results. Our team culture emphasizes mentorship, collaboration, and professional development. This internship promises invaluable preparation for your future research career and for translating biomedical data science methods into better patient care.

SPECIFIC SKILLS: Demonstrated skills with **Python** machine learning and a strong foundation in data science principles would prepare a graduate student to contribute and learn quickly. We prioritize finding interns excited to expand their data science and predictive modeling capabilities. Our team follows coding best practices, and an emphasis on mentorship will help elevate an intern's skills.

J. John Bridges, PhD

--seeking 4-6 interns--

The Patient-centered Outcome Research Lab is headed by Dr. John F P Bridges and is predominantly focused on the measurement of the priorities and preferences of patients and other stakeholders in medicine. We are world leaders in several theory-driven methods used to measure patient preferences including discrete-choice experiments and best-worst scaling. Our group has been a leader in the development and dissemination of best practices for using these methods and we have worked with decision makers around the world to advance and apply these methods. We are seeking summer interns to work on a variety of projects that we are leading in our lab and/or collaborating with other on a wide array of topics including HIV, cancer, mental health, pediatrics, organ donation, medical devices, artificial intelligence, and other therapies. Dr. Bridges has mentored over 100 undergraduate/postgraduate students and junior researchers/clinicians and his former mentees have gone on to graduate studies or positions at many top universities, which many other have been successful in industry, government, and non-governmental agencies. A surprising number of his mentees still work in measuring preferences or have incorporated these skills into their research. Finally, some of summer scholars have stayed on to work or study with us, and we are always looking for bright students to join our lab.

PCOR Lab Projects

- a) **Best-Worst Scaling (BWS) Quality Evaluation:** Past systematic reviews of best-worst scaling (BWS) studies have struggled to evaluate the quality of BWS studies due to a lack of consensus on quality and limitations with quality evaluation instruments. The **BWS Quality Evaluation**

project aims to apply an updated quality assessment instrument to evaluate the quality of BWS studies ever conducted in all fields. The project is designed to provide candidates with an opportunity to engage in team science and contribute to scientific products, including a scientific manuscript.

- b) Best-Worst Scaling (BWS) Policy Relevance Assessment:** There is growing emphasis on designing and conducting best-worst scaling (BWS) studies to have an impact on policy making. However, there is no comprehensive review of the policy relevance of existing literature that uses BWS across all fields. The **BWS Policy Relevance Assessment** project involves conducting a systematic assessment of the policy relevance of all studies that apply BWS following rigorous and transparent methods. The project is designed to provide candidates with an opportunity to engage in team science and contribute to scientific products, including a scientific manuscript.
- c) Best-worst scaling (BWS) in the context of cancer screening perceptions and barriers:** Preventative screening and improved access to healthcare play major roles in reducing cancer incidence and enhancing outcomes. Tackling disparities in healthcare access is paramount; individuals facing economic challenges often encounter heightened prevalence of cancer and associated mental health issues, such as depression, with their health needs often poorly understood by medical professionals. This project involves conducting three short online surveys to gauge knowledge levels, perceptions, and barriers related to cancer screening and could inform the development of tailored interventions, educational initiatives, and policies.
- d) Patient-Centered Outcomes Research (PCOR) Lab Policy Impact:** The PCOR lab led by Dr. John Bridges has been a leader in the development and dissemination of best practices for using patient preference methods and we have worked with decision makers around the world to advance and apply these methods. The **PCOR Lab Policy Impact** project will involve assessing and characterizing the impact of the lab's research in areas such as regulatory science, patient advocacy, and public health. The project is designed to provide candidates with an opportunity to engage in team science and to contribute to external communication materials.
- e) Secondary Analysis of a Suicide Prevention Discrete-Choice Experiment (DCE):** Information from credit bureaus such as bankruptcy and foreclosure might improve the accuracy of machine learning models used by health systems for suicide risk prediction. A discrete-choice experiment (DCE) was conducted to assess patients' preferences for linking credit bureau data to healthcare

records for suicide prevention. The **Secondary Analysis of a Suicide Prevention DCE** project will involve conducting secondary analysis of the data, including comparisons between covariates such as age, race, ethnicity, sex, income, and educational attainment. The project is designed to provide candidates with the opportunity to engage in team science and to contribute to scientific products, including a scientific manuscript.

- f) **Investigative Genetic Genealogy (IGG) Discrete-Choice Experiment (DCE):** Investigative genetic genealogy (IGG) can assist law enforcement with identifying perpetrators of crime and unidentified human remains. While previous survey research has demonstrated general acceptance of the use of IGG in criminal investigations, it has not proved acceptability of specific practices. The **IGG DCE** project will involve assisting with the completion of the corresponding manuscript, including conducting literature searches. This project is designed to provide candidates with an opportunity to engage in team science and contribute to scientific products, including a scientific manuscript.

K. Mohamed Elsaid, PhD

--seeking 1-2 interns--

Hepatocellular Carcinoma (HCC) survival rates are among the lowest in the United States. Localized HCC tumors show substantially improved patient outcomes, emphasizing the need for effective early detection strategies for at-risk individuals. Current guidelines recommend semiannual surveillance of patients with cirrhosis. However, overall adherence rates to HCC surveillance guidelines in cirrhosis patients are very poor, and studies show that social determinants of health exacerbate these rates. We aim to examine health disparities in HCC outcomes by using data on communities most impacted by health disparities to develop predictive analytics tools to optimize surveillance adherence among cirrhosis patients. Results from this work will help identify and characterize high-risk patients who would most benefit from targeted interventions, leading to improved outcomes and reduced health disparities. **Targeted Skills:** Basic programming skills preferred, basic statistical background, interest in health disparities research

L. Xiaokui (Molly) Mo, PhD

--seeking 1 intern --

Biomarkers for Early Detection of Low-Dose Radiological Exposure in Biological Samples

Our research aims to discover early biomarkers for low-dose radiological exposure in biological samples. Using spectroscopy platforms and sequencing technology, we measure samples exposed to radiation. Biomarker identification combines traditional statistics methods (ANOVA, multiple regression, multiple

logistic regression) and advanced machine learning, including the powerful random forest algorithm. Validation, ensuring robust predictability, includes Receiver Operating Characteristic (ROC) analysis, assessing specificity (true negative rate) and sensitivity (true positive rate). Comprehensive analysis is conducted using SAS, R, and other relevant software tools. **Targeted Skills:** Know the basics of R

M. AI Challenge under the supervision of Ping Zhang, PhD

--seeking 5 interns --

Potential Projects: Attend KDD Cup 2024 and/or PhysioNet Challenge 2024 as the [Team Buckeye AI](#). Last summer, the topic was to predict neurological recovery from coma after cardiac arrest based on signal data from EEG, ECG, and/or other clinical time-series data. We usually develop large language models (LLM) and large multimodal models (LMM) to tackle real-world healthcare problems.

Targeted Skills: Ideal candidates should be self-motivated and passionate about conducting original research. Students from related disciplines such as computer science, software engineering, informatics, statistics, mathematics, automation, and electrical engineering are strongly encouraged to apply. Programming experience in either Python, R, MATLAB, Java, or C/C++ is required. Prior knowledge of and research experience in either data mining, machine learning, or biomedical informatics is preferred.

N. No Preference

If you have no preference on projects, your application will be shared with all of the faculty mentoring students this year.