

This is the first part of the application process for the internship program in biomedical informatics at OHSU.

The Department of Medical Informatics and Clinical Epidemiology (DMICE) will be hosting its College Undergraduate Biomedical Informatics and Data Science Internship Program again in the summer of 2024.

These internships provide full-time experiences in biomedical informatics and data science to college undergraduates students to increase their knowledge and skills in these areas. The program's goals are to provide exposure to biomedical informatics and data science with respect to best practices, impact of research on medicine and health, as well as possible graduate study and career paths. In addition to participation in a faculty-led research project, the experience will include:

- Knowledge of the motivations, activities, and challenges for biomedical informatics and data science**
- Skills development in data science through the use of standard data sets and tools**
- Training in responsible conduct of research**
- Student-led and faculty-facilitated health data ethics and equity seminar**
- Summer interns will also attend weekly Informatics Research**
- Conferences and present their research at the end of the internship at a mini-symposium.**

The overarching goal of this summer program is to support educational activities that encourage individuals from diverse backgrounds, including those from groups underrepresented in the biomedical and behavioral sciences, to pursue further studies or careers in research. Due to the funding agency requirements, this program is open to U.S. Citizens and Permanent Residents only. Some of the positions will be funded by a new program from the US National Library of Medicine to enhance diversity in biomedical informatics and data science. This is a 10 - 12 week opportunity with projects starting June 17, 2024. Interns will work 40 hours per week and will be paid \$16.70 per hour.

Once you complete the online application, the next step will be to provide the following documents:

- 1. Transcripts - unofficial transcripts acceptable, email pdf to ilgan@ohsu.edu**
- 2. One letter of recommendation, emailed to ilgan@ohsu.edu**

* 1. Please enter the following information to start your application

Name:

Address:

City/Town:

State:

ZIP:

Country:

Email Address:

Phone Number:

2. Date of Birth (mm/yyyy)

3. Preferred pronouns

4. Please describe your race/ethnicity.

Please note: Demographic information is being collected for evaluative purposes and will be used along with other information and metrics to help determine the effectiveness of the training program. This information will not be used in intern selection.

☐ American Indian or Alaska Native

☐ Asian

☐ Black or African American

☐ Hispanic or Latino

☐ Native Hawaiian or Other Pacific Islander

☐ White

☐ Prefer not to state

☐ Other (please specify)

* 5. Country of Citizenship

Academic Information

6. Please provide us with the following academic information

School attending	<input type="text"/>
Major	<input type="text"/>
Minor	<input type="text"/>
Dates attended	<input type="text"/>
Current class in school	<input type="text"/>
Completed credit hours	<input type="text"/>
Current cumulative GPA	<input type="text"/>
Anticipated graduation date	<input type="text"/>

7. Please provide information about experience/classes you have had in the following areas

Programming experience (R, Python, etc.)

Statistical Analysis (t-test, ANOVA, etc.)

Databases (sql, etc.)

Command Line Interface (Terminal on MAC OS X or Powershell on Windows)

Version Control (Git, Mercurial, etc.)

Project Opportunities

* 8. Please rank your top 3 internships. In each question, the faculty lead(s), the description, and the skills required are listed.

1st 2nd 3rd

Causal Fairness Analysis

Faculty Mentor: Dr. Mohammad Adibuzzaman

Fairness in data-driven decision machines and algorithms is an emerging point of discussion in the scientific, political, and policymaker communities. Common reasons for algorithmic biases include (not limited to) changes in data distribution, real-world interactions, user behavior, and shifts in data capture and management practices. Distinct computational methods are being rigorously developed to tackle this issue; however, there still exists controversy around estimating algorithmic biases and instigating algorithmic fairness.

Our research project investigates the causal pathways to identify, quantify, and address algorithmic bias. The research aims to diminish predictive biases (algorithmic inaccuracies in producing estimates that significantly differ from the underlying truth) and social biases (systemic inequities in care delivery leading to suboptimal health outcomes for specific populations). Using theories of causal inference, we explore structural causal and fairness models to disentangle complex causal puzzles and ways to mitigate these biases. Primarily, we are exploring computational approaches to identify predictive and social bias, point of bias generation, and ways forward for follow-up investigations. Additionally, we are looking for consistent evaluation and assessments of the algorithm over time and for all patient population cohorts. For this exploration, we are using existing benchmark datasets (COMPAS recidivism dataset) and the Cosmos population cohort in Epic. Computer Science/Informatics background recommended and algorithms course a plus!

☐☐☐

Analyzing the impact of Large Language Models on Automated Classification and Manual Annotation

Faculty Mentor: Dr. Aaron M. Cohen

Large language models (LLMs) are a new natural language technology that is poised to change almost all tasks current performed with traditional machine learning. However, there are many unanswered questions in using LLMs for biomedical text classification problems, such as how do we best ensure good performance and how does providing examples improve results? In this project the intern will perform a comprehensive study on the interactions between amount of manual annotation, performance assessment and uncertainty, and no-shot vs. few-shot prompting for LLMs. Some background in Python programming and statistical analysis is strongly recommended.

☐☐☐

High Blood Pressure patient-facing clinical decision support

Faculty Mentor: Dr. David Dorr

High blood pressure is one of the most common chronic conditions in adults older than 50, and the most common contributing factor for heart attacks and strokes. Significant evidence exists about both pharmacologic and non-pharmacologic methods to lower blood pressure, but they require substantial shared decision making and patient motivation. The intern will help us with our patient-facing HBP application; depending on their skill set, help programming, testing, or evaluating the tool with patients and care teams.

☐☐☐

Evidence Synthesis for Stakeholder Groups

Faculty: EPC Core Investigators

This position supports evidence synthesis research and related projects for the Pacific Northwest Evidence-based Practice Center (<http://www.ohsu.edu/epc>) which conducts systematic reviews on health care topics for federal and state agencies, professional associations, and other organizations. These reviews present the evidence from research studies and the quality of that evidence for use by clinicians, employers, policymakers, researchers, and others in making decisions about the provision of health care services and health research. It is anticipated that an intern will engage in the development and/or application of advanced computational methods to enhance the efficiency and accuracy of systematic reviews. This may include utilizing AI to semi-automate literature search, data extraction, and analysis, and applying machine learning algorithms to identify patterns and insights within datasets. Depending on skillset and interests, the intern will participate in project meetings, import and de-duplicate text files of electronic database search results; code and manage EndNote library citations, perform abstract review, data abstraction, contribute to, format, and track report content. This is a great opportunity to learn more about healthcare research, systematic reviews, and the cutting-edge use of AI and machine learning in the field.

☐☐☐

Improving Matching of Patients to Clinical Studies

Faculty Mentors: Drs. William Hersh and Steven Bedrick

Medical research advances when people volunteer to participate in clinical trials and other studies. One challenge is that patients are not identified or otherwise aware of studies in which they may take part. Our work focuses on using data from the electronic health record to identify patients who might be candidates for clinical studies. Python programming experience recommended.

☐☐☐

Voice Biomarkers and AI

Faculty Mentor: Drs. Bill Hersh and David Dorr

Voice biomarkers and Artificial Intelligence show potential for advancing healthcare by building predictive models to assist in screening, diagnosis, and treatment of a broad range of diseases, including several diseases with unmet clinical needs. The Voice AI Summer School provides a unique and innovative hands-on opportunity for students in health, informatics, data, computer science and related fields to collaborate and learn about the creation of voice biomarkers using Artificial Intelligence. Curriculum will include processing voice audio, extracting and choosing acoustic and linguistic features, training and testing voice algorithms, as well as the social and ethical implications of working with voice data. Applicants with non-clinical backgrounds (e.g., computer or data science, public health) will be matched in pairs with a clinical intern. Pairs will work together complete assignments and solve challenges throughout the program, finalizing the internship with a Hackathon event at University of South Florida in mid-August (travel expenses will be paid by the program). Applicants should have training or experience in data science, computer science, or informatics. There are up to 6 summer internship positions for this project. This internship project starts June 17 and ends on August 16, 2023.

☐☐☐

Data Standardization

Faculty Mentor: Michelle Hribar

In order for clinical data to be used in large-scale projects for AI and machine learning, it has to be standardized. Data collected during ophthalmology exams has yet to be fully standardized, limiting its reuse for data science. Currently, there is a national effort to address this. This project will involve working with ophthalmologist and vision researchers at the National Eye Institute and other prominent academic institutions on any steps of the data standardization process: demonstration of the need for standards,

☐☐☐

identification of ophthalmic data elements that need standardization, consensus building on about standard definitions, and/or validation of the standards. [Note: there could possibly multiple projects here]

Telehealth

Faculty Mentor: Dr. Michelle Hribar

Many age-related eye diseases that result in irreversible vision loss have no advanced symptoms before vision loss, but can be detected early through eye exams and imaging. Unfortunately, access to eye care across Oregon can be limited due to geographic and socioeconomic factors. Casey Eye Institute is partnering with community health clinics throughout Oregon to establish telehealth screening programs to identify eye disease early. This ambitious project is in the pilot phase and will require informatics and data science work to monitor the effectiveness of screenings, patients' follow-up with eye care specialists, screening workflow efficiencies, and technological evaluation as the program is scaled to more clinics.

☐☐☐

Understanding the Human Microbiome

Faculty Mentor: Dr. Lisa Karstens

Humans live in a symbiotic relationship with hundreds of microorganisms. These bacteria, fungi, and viruses that make up the human microbiome are essential for understanding human health and, more importantly, disease. To study the human microbiome, researchers often generate large datasets containing sequencing or metabolic information that is then associated with clinical and demographic information to address a clinical question. Intern projects include developing, testing, and improving the pipelines for handling these data for a variety of projects investigating the microbiome's role in relation to human disease, including bladder disorders, rheumatic disease, and cancer. Data include survey and questionnaire data from REDCap, 16S rRNA gene sequencing data, and metabolomics data. The projects will provide experience of analysis and biological interpretation of so-called 'big data' that arises from the rich and complex datasets generated by high throughput techniques used in basic research. Excellent record-keeping skills and self-motivation are essential. Some familiarity with programming and statistical analysis in R are preferred but not essential.

☐☐☐

Land Ho! Mapping clinical informatics competencies by navigating DMICE courses - it's not quite Magellan's voyage, but it is a journey of discovery especially if you are interested in pursuing a career in clinical informatics.

Faculty Mentor: Dr. Vishnu Mohan

Our core clinical informatics (CI) courses are utilized by students in our graduate certificate, masters, PhD and clinical informatics subspecialty fellowship program. We want to map the content of these core CI courses to defined competencies in the field, and develop an updated matrix that will help us understand how the courses we teach meet the competencies defined for clinical informatics. As an intern, you will survey eight CI courses, and review their syllabi, learning materials and associated content. You will get a comprehensive, in-depth exposure to one of the largest and most innovative CI programs in the nation. Plus, you will help to improve the training of informaticians, which is always a good thing (especially if you intend to be one of those informaticians trained!)

☐☐☐

AI for Cancer Imaging

Faculty Mentor: Xubo Song

Our group focuses on applying artificial intelligence to cancer imaging, to better understand how cancer develop and evolve, and for precision early cancer detection and precision treatment. Example projects include applying machine learning and generative AI models for data normalization, domain adaptation, image super-resolution, image denoising, cross-modal translation, multimodal integration, image segmentation, representation learning, spatial

☐☐☐

and dynamics modeling. There are multiple imaging modalities including histopathology, immunofluorescent imaging, electron microscopy, and mammography. The approaches include convolutional neural networks, transformers, diffusion models, variational autoencoders, and graph models. Students familiar with programming and knowledge/experience with machine learning are preferred.

Assessing the impact of Oregon's new Nurse Staffing Law

Faculty Mentor: Dr. Dana Womack

In 2023, the Oregon Legislature passed HB 2697, which makes significant changes to Oregon's hospital staffing law. HB 2697 establishes minimum standards for staffing plans for direct care registered nurses (RNs), and imposes penalties for failure to adopt and comply with staffing plans. This historic change presents opportunities for pre/post analysis of multiple aspects of care delivery and nursing work experiences. We seek a student to participate in collection and analysis of nursing and care delivery data as part of OHSU's effort to evaluate the impact of this new law. Desired skills: experience with data collection, data validation, and data analysis tools. Understanding of the nurse staffing process and workflow is desirable but not required.

☐☐☐

Applying chatGPT and other large language models (LLMs) to automate the Reactome pathway annotation

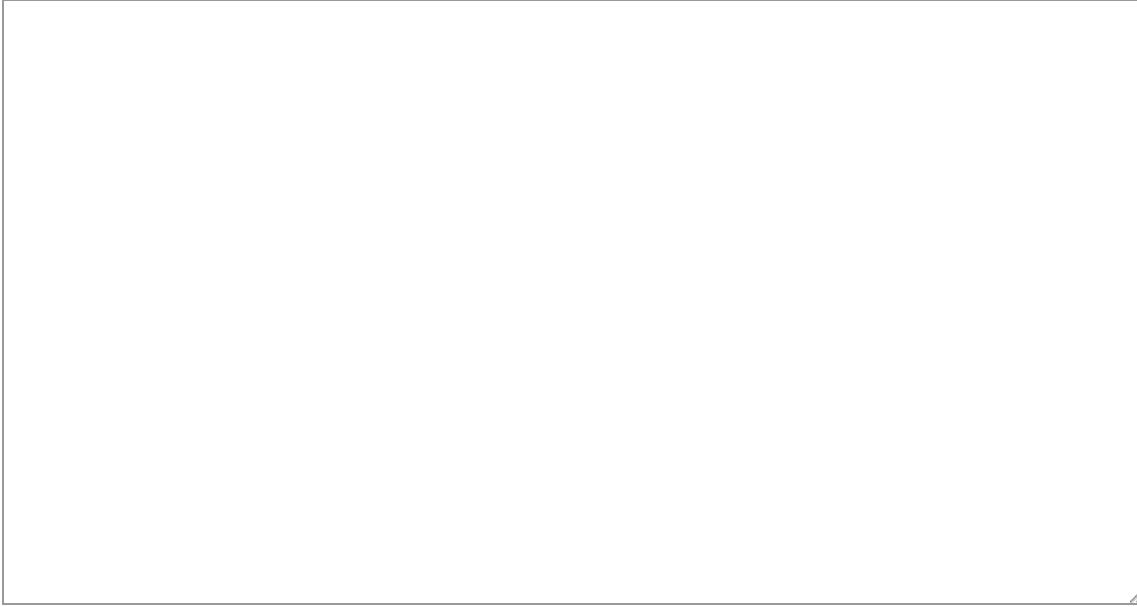
Faculty: Dr. Guanming Wu

Reactome is one of the most popular open-source biological pathway knowledgebases, widely used in the community for large scale data analysis and visualization. The content in Reactome is manually curated to ensure high quality. However, manual curation is laborious and time consuming. In this project, our aim is to explore the feasibility of leveraging chatGPT and other LLMs for automating the curation process, thereby enhancing curation efficiency. A background in biology is required, and experience with Python programming will be advantageous for contributing to this project.

☐☐☐

Essay Question

* 9. Essay Question: Please tell us how your education, experience, skills, and interest make you the best candidate for your top (few) choices.

A large, empty rectangular box with a thin black border, intended for the user to write their essay response. The box is positioned below the question text and occupies a significant portion of the page layout.

Finishing up the application process

* 10. How did you hear about this program?

11. Are you planning to pursue graduate education? If so, in what field.

Thank you for applying to the summer undergraduate internship program in biomedical informatics at OHSU.

The application deadline is February 16, 2024. Candidates will be notified on or before March 22, 2024 regarding final decisions.

Please remember to submit the additional documents to complete your application.

1. Transcripts - unofficial transcripts are acceptable, email pdf to ilgan@ohsu.edu
2. One letter of recommendation, emailed to ilgan@ohsu.edu

Specific questions can be directed to Andrea Ilg at ilgan@ohsu.edu