College Undergraduate and Graduate Biomedical Informatics and Data Science Internship Program
Sample of Projects for Summer 2026 Internship

A common longitudinal intensive care unit data format (CLIF) for critical illness research
Faculty Mentor: Dr. Patrick Lyons
Critical illness threatens millions of lives annually. Electronic health record (EHR) data are a source of granular information that could generate crucial insights into the nature and optimal treatment of critical illness. The multi-center Common Longitudinal ICU data Format (CLIF) project aims to overcome key EHR data standardization, management, and security, barriers to large-scale critical illness EHR studies. Depending on the intern's skill set and interest, they may help with consensus-based data element standardization, development of internal validation and quality assessment tools, or a disease-focused subproject.

ADHD Risk Prediction
Faculty Mentor: Dr. Michael Mooney
Neurodevelopmental conditions, like attention-deficit hyperactivity disorder (ADHD), are influenced by a complex interplay between multiple genetic and environmental factors. Our work aims to better understand individual risk factors and how they jointly contribute to the course of ADHD-related traits and outcomes over time. Internship projects involve examining relationships between genomic/epigenomic factors, environmental exposures, and behavioral measures in longitudinal data sets to aid the development of risk prediction models. Experience with programming and statistical analysis in R or Python is recommended.

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!

Leveraging Large Language Models for Automated Extraction of Pathway Information from Literature
Faculty: Dr. Guanming Wu
Reactome is the most comprehensive open-source biological pathway knowledgebases, widely used in the community to support large-scale data analysis and visualization. Its content is manually curated from literature to ensure accuracy, but this process is time-consuming. This project aims to explore the use of large language models to automatically extract pathway-relevant information, including molecular interactions and regulatory relationships, from biomedical literature. The goal is to assess the feasibility of integrating LLM-based extraction into Reactome's curation workflow to improve efficiency while maintaining quality. A background in biology is required, and experience with Python programming will be advantageous.

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.

Implementing an Electronic Care Plan for People with Multiple Chronic Conditions – eCarePlan

Faculty Mentor: Dr. David Dorr

This project intends to understand the best method to implement the eCarePlan (presented to the patient as "MyCarePlanner") application in clinical care by studying the way Fast Healthcare Interoperable Resources (FHIR) can be used to pull multiple sources of data in order to prioritize, summarize, and display it for care coordination and care planning. There is a focus on behavioral health conditions and settings. Clinicians and patients are participating in focus groups, interviews and surveys to evaluate the application. The intern will help with evaluating the tool with patients and clinicians.

AI for Cancer Imaging

Faculty Mentor: Dr. 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 harmonization, 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, immuno-fluorescent imaging, electron microscopy, and radiological images. The approaches include convolutional neural networks, transformers, diffusion models, variational autoencoders, graph models, foundation models and multimodal large language models. Students familiar with programming and knowledge/experience with machine learning are preferred.

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 demographical 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.

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.

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!)