About the Program

The Department of Biomedical Informatics hosts an annual internship program each summer, which provides opportunities for graduating high school seniors and 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 very useful presentation abilities 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 medicine and nursing.

All student internship positions in the Department of Biomedical Informatics are paid, with the hourly wage depending on student experience levels. Many positions have the ability to turn into academic year student employment opportunities.

This year the program will run May 22 to July 28, 2023.

Please contact internship coordinator, Gabrielle Kokanos, gabrielle.kokanos@osumc.edu, with questions.
2023 Internship Opportunities

A. Lijun Cheng, PhD
Assistant Professor, Biomedical Informatics

--seeking 3 interns for up to 20 hours per week--

Potential Projects: 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. Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated nuclease 9 (Cas9), as a powerful genome editing system, in the identification of synthetic viability/resistance and lethality mechanisms and in overcoming drug resistance in the most frequent solid tumors. Dr. Cheng’s research focuses on artificial computational models to identify optimal targets or drugs by CRISPR screening data guidance in precision cancer medicine. Her AI computational tools were developed are to elucidate the molecular basis of drug resistance, and identification of resistance-related genes, and improve clinical outcomes integrating with CRISPR/Cas9 system. Dr. Cheng hopes to recruit several students continue to work on this project.

Targeted Skills: Python, R, or MATLAB coding

B. Norah Crossnohere, PhD
Assistant Professor, Internal Medicine

--seeking 3 interns for up to 20 hours per week--

Potential Projects: We have two exciting projects related to measuring the preferences of patients and other stakeholders in medicine, both of which will have opportunities for students to co-author publications.

The first project is about measuring preferences for lung cancer screening. Despite being a leading cause of cancer-related death in the US, lung cancer is often not routinely screened in primary care. In this study we will be exploring what the preferences of patients and doctors are for administering lung cancer screening. Knowing preferences for lung cancer screening can ultimately be used to improve uptake of lung cancer screening. In this project the intern would be responsible for assisting with a literature review of previous research on preferences for lung cancer screening and be involved in the design of a new experiment to measure the preferences of patients and clinicians for lung cancer screening.

The second project is more methodological in nature and is to develop a new measure to evaluate the quality of preference studies. Preference studies in medicine typically describe the priorities, preferences, or needs of stakeholders to inform medical decision making. There are unique methods to measure preferences, but the current tools that are used to evaluate how well these studies are
2023 Internship Opportunities

Conducted are lacking. This project is developing a consolidated set of standards for the evaluation of preference studies in health. In this project the intern would be responsible for assisting in the development and conduct of a literature review to catalogue all current standards to evaluate preference studies, and to use this information to inform the development of our new consolidated measure.

Targeted Skills: Literature reviews, Research design, Analysis of quantitative data (our preferred package is Stata), Organizational and administrative skillsets

It is okay if interns do not have these skillsets but must be willing to learn.

C. Mohamed Elsaid, PhD
Assistant Professor-Clinical, Biomedical Informatics

--seeking 2 interns for up to 30 hours per week--

Potential Projects: Advances in cancer control across the disease spectrum depend on testing and validating new intervention strategies and treatment modalities in clinical trials. Although enrollment in clinical trials have increased in the United States during recent years, participation rates have not been equal across all racial and ethnic groups. This racial disparities limits progress against efforts to advance treatment modalities in all population groups. In this project, we aim to conduct a review of all cancer screening clinical trials conducted in United States to assess and examine racial and ethnic enrolment disparities.

Targeted Skills: Basic research and computer skills. Experience in research related to cancer prevention is preferred but not required.

D. Naleef Fareed, PhD
Assistant Professor, Biomedical Informatics

--seeking 2 interns for up to 10 hours per week--

Potential Projects: I am an assistant professor in biomedical informatics, with a specialized focus in clinical informatics and public health informatics. The Fareed lab has developed data pipelines and data science tools that can transform health care decision making through better care delivery and resource distribution. Dr. Fareed also conducts extensive work with clinical, patient generated data, and publicly available data to generate clinical insights.

Targeted Skills: Students are encouraged to have fundamental knowledge in areas such as statistics and programming. A training plan will be established at the start of the program to facilitate the professional development of the student.
2023 Internship Opportunities

E. Courtney Hebert, MD, MS

Associate Professor, Biomedical Informatics

--seeking 1 intern for up to 20 hours per week--

Potential Projects: This project will involve working on an interdisciplinary team to analyze electronic health record data on infections within the hospital.

Targeted Skills: Prior course work in statistics and familiarity with statistical software is preferred. Will require completing IRB training on human subject protection.

F. Lang Li, PhD

Professor and Chair, Biomedical Informatics

--seeking 3-5 interns for up to 38 hours per week--

Potential Projects: We will have three types of projects. (1) Data curation for pharmacotherapy data in cancer, pediatric, and maternal patient population. (2) Data curation for machine learning papers. (3) Text mining and natural language processing analysis using deep learning and machine learning packages.

Targeted Skills: Data curation work would require biology, pharmacology, epidemiology, and/or biostatistics skills. Text mining work will require some programming skills in python.

G. Qin Ma, PhD

Associate Professor, Biomedical Informatics

--seeking 1 interns for up to 30-38 hours per week--

Potential Projects: 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 programming: R, Python, GitHub, familiarity with high-performance computing skills, 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 and microbiology.
2023 Internship Opportunities

H. Xia Ning, PhD
Associate Professor, Biomedical Informatics and Computer Science & Engineering
--seeking 2 intern for up to 38 hours per week—

Potential Projects: We are building recommendation engines for various applications in biomedicine. Those recommendation engines will leverage the cutting-edge techniques from e-commerce applications (e.g., Amazon) so as to personalize information items with respect to each users (e.g., physicians, patients). The recommendation engines will also utilize advanced representation techniques (e.g., embeddings via deep learning) to represent information items (e.g., ICD codes, symptoms) to facilitate fast and effective information search and retrieval. The recommendation engines will be used in various applications in clinical settings.

Targeted Skills: Python programming including pytorch

I. Fode Tounkara, PhD
Assistant Professor-Clinical, Biomedical Informatics
--seeking 3 interns for up to 35 hours per week--

Potential Projects: Project #1: Rapid growth in the size and scope of data sets in a host of disciplines have created a need for innovative statistical strategies analyzing such data. For example, many private and public agencies are using sophisticated data mining strategies and/or big data analytics to reveal patterns based on collected information. Some examples of big data that have prompted demand are gene expression arrays, social network modeling, clinical, genetics and phenotypic data. This project focuses on estimation of model parameters and prediction based on high dimensional data (HDD). In a regression context, we define HDD as when variables are larger than the sample size. There are techniques available to analyze HDD and they mostly rely on penalized regularization methods to do variable selection and estimation simultaneously. These methods do well when the model has a few strong signals and with sparse signals. In this project we consider model selection and post-estimation problem in a more realistic framework as follows:

a. Most of the features may not have any influence (sparse signals) on the response of interest
b. Some of the features may have strong influence (strong signals) on the response of interest
c. Some of the features may have weak-moderate influence (weak-moderate signals) on the response of interest.

Project #2: Unstructured longitudinal electronic health records (EHR), such as free text clinical notes, inherently contain rich information about patient’s health and outcomes over time. The right analytical tools capable of capturing sequential information can therefore maximize utilization of longitudinal EHRs and can be valuable for supporting clinical decisions and prognostic models. In this project, the
2023 Internship Opportunities

Successful candidate will focus on a different methodology to address the problem of capturing sequential information contained in longitudinal electronic health records (EHR).

**Targeted Skills:** Basic to moderate statistical skills; coding in R and Python; standard writing skills, team player

J. Ping Zhang, PhD
Assistant Professor and Vice Chair, Biomedical Informatics and Computer Science & Engineering

--2 interns for up to 38 hours per week--

**Potential Projects:**

Project #1: Integrate multiple data sources for drug effect prediction

Project #2: Explore electronic medical records and claims data for clinical risk prediction

**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. Extensive 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 required. Experience in publishing first-author research articles in peer-reviewed technical conferences (such as NeurIPS, ICML, ICLR, KDD, ICDM, SDM, WWW, IJCAI, AAAI, AMIA, BIBM or similar) is highly preferred.

K. Xiaoli Zhang, PhD
Associate Professor-Clinical, Biomedical Informatics

--seeking 2 interns for up to 20 hours per week--

**Potential Projects:** Immune checkpoint inhibitor (ICI) has revolutionized cancer treatment in a broad variety of tumor types including lung cancer and multiple types of hematologic malignancies (HMs, i.e., lymphoma, leukemia, and multiple myeloma). However, due to patient heterogeneity, only a small proportion of patients respond to the ICIs and majority of them eventually develop drug resistance. Therefore, the potential project is to analyze the immunotherapy database and/or do literature review to understand effects of patient demographics, treatment history, and clinical biomarkers/characteristics on patient ICI treatment outcomes/adverse events. This will provide critical information to understand potential mechanisms of patient resistance to checkpoint blockade, and possible strategies for expanding the number of patients who can benefit from immune checkpoint directed therapies.

**Targeted Skills:** Potential students should have basic training of programming in SAS, R, or other languages.
2023 Internship Opportunities

L. AI Challenge under the supervision of Ping Zhang, PhD
--seeking 5 interns for up to 20 hours per week--

Potential Projects: Attend KDD Cup 2023 and/or PhysioNet Challenge 2023 as the Team Buckeye AI.

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. Extensive 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 required. Junior or Senior year students with "A" grades on AI/ML/DM courses are highly preferred. Junior or Senior year students who would like to apply for PhD program in computer science are highly preferred.