Research Statement

My research focuses on developing novel AI techniques for biomedical applications, with an emphasis on translating biomedical data science research to benefit both clinical practice and biological investigations. I possess a unique blend of expertise in AI, bioinformatics, high-performance computing, and biomedical physics (BMP) and imaging. My passion for translational research positions me to conduct cutting-edge AI research and lead innovations in AI-augmented biomedical physics, establishing connections with cancer biology and radiation therapy to make significant contributions to the field. The following summarizes my research experience and achievements and outlines the future directions of my research endeavors.

A. Research achievements

A1. Medical imaging and image analysis: I developed innovative experimental and computational techniques for imaging the elastic and poroelastic properties of cancer tissues using ultrasound during my PhD study. A novel mathematical framework for reconstructing the elastic and fluid transport properties (vascular permeability, interstitial permeability, and interstitial fluid pressure) of cancer tissues was established [1-3]. Additionally, an accurate method to simultaneously estimate elastic modulus and compressibility from ultrasound images was developed within the framework. The work enhanced our understanding of the complex mechanical behavior of cancer tissues and offered valuable insights for cancer diagnosis and therapeutic assessment.

A2. High-performance and interpretable AI for BMP applications: Deep neural networks (DNNs) extract thousands to millions of task-specific features during model training for inference and decision-making. While visualizing these features is critical for comprehending the learning process and improving the performance of the DNNs, existing visualization techniques work only for classification tasks. For regressions, the feature points lie on a high dimensional continuum having an inherently complex shape, making a meaningful visualization of the features intractable. Given that the majority of deep learning applications in BMP are regression-oriented, developing a conceptual framework and computational method to reliably visualize the regression features is of great significance. I introduced a manifold discovery and analysis (MDA) method for DNN feature visualization [4], which involves learning the manifold topology associated with the output and target labels of a DNN. MDA provides insightful visualizations of the DNN features, highlighting the appropriateness, generalizability, and adversarial robustness of a DNN. Another of my recent contributions is the development of a Contrastive Feature Analysis (CFA) framework, detailed in Ref.[5]. These works allow for a deeper understanding of the DNN 'black box,' enabling the design of more efficient neural network architectures. The effectiveness of CFA and MDA in improving DNN performance and interpretability across multiple medical imaging applications has been found to be remarkable.

A3. Deep learning for genomics data analysis: Remarkable advances in single-cell genomics have presented unique challenges and opportunities for interrogating a wealth of biomedical inquiries. High-dimensional genomic data are inherently complex because of intertwined relationships among the genes. Existing methods, including emerging deep learning-based approaches, do not consider the underlying biological characteristics during data processing, which greatly compromises the performance of data analysis and hinders the maximal utilization of state-of-the-art genomic techniques. I developed an entropy-based cartography strategy to contrive the high dimensional gene expression data into a configured image format, referred to as genomap, with explicit integration of the genomic interactions[6,7]. This unique table-to-image conversion casts the gene-gene interactions into the spatial configuration of genomaps and enables us to extract the deep genomic interaction features and discover underlying discriminative patterns of the data. I show that, for a wide variety of applications (cell clustering and recognition, gene signature extraction, single cell data integration, cellular trajectory analysis, dimensionality reduction, and visualization), the proposed approach drastically improves the accuracies of data analyses as compared to the state-of-the-art techniques.

A4. Applications of AI in radiation therapy tasks: I have a strong interest in the translation of AI techniques in clinical applications. In the past few years, I have had strong collaborations with many researchers and done substantial work in advancing patient care technologies. For example, working with Vasudevan et al [8], we have developed an implicit neural representation (NeRP) for patient treatment plan. Clinically, a concise representation of the radiation dose distribution is of great value in facilitating treatment planning and downstream applications. With the NeRP technique, we were able to reduce the necessary storage space of the treatment plans substantially. We have also applied a content-based image retrieval method for retrieving dose distributions of previously planned patients based on anatomical similarity[9]. Retrieved dose distributions from this method can be incorporated into automated treatment planning workflows to improve the iterative planning process. I have also leveraged my background in deep learning and worked with my peers on image segmentation and super-resolution imaging for improved precision radiation therapy [10,11]. Recently, I have worked on developing an AI model with Liu et al [12] to predict whether a particular patient will respond to immunotherapy based on the patient's tumor gene mutation information.

B. Future research directions

I plan to establish a translational AI research program to improve biomedical research and clinical decision-making by integrating multi-modality and multi-scale data into the deep learning analysis pipeline. Specifically, my lab will focus on (1) developing interpretable and trustworthy AI techniques for biomedical applications; (2) investigating deep learning methods for improved analysis of tabular data (such as omics and clinical trials) and integration of multi-modal data; (3) integrating large language models into clinical research and applications; and (4) advancing medical imaging and radiation treatment planning decision-making by utilizing state-of-the-art AI techniques.

B1. Developing interpretable and trustworthy AI techniques for biomedical applications: AI has demonstrated remarkable potential in advancing biomedical research and healthcare delivery. However, the widespread adoption of AI requires the enhancement of interpretability and trustworthiness. I will leverage the CFA [5] tool to create highly interpretable AI models with mechanisms to dive into the model's feature space to better understand the decision-making process of DNNs. I will explore the models' feature space to identify the relevant biomarkers for various applications such as biomarkers for cancer diagnosis and prognosis. I will also investigate a strategy for improving the robustness of the model against adversarial attacks and noises by using the CFA, ensuring the reliability of AI systems in real-world settings. During the course of this research, I will collaborate closely with clinicians and computer scientists across campus, fostering a multidisciplinary approach that combines clinical knowledge with AI expertise. I will validate the developed techniques through rigorous experimentation on diverse biomedical datasets, spanning genomics, medical imaging, electronic health records, and clinical trials. The expected outcomes of this research include interpretable AI models that enhance biomedical data analysis, improved trustworthiness of AI-driven decision support systems, and increased acceptance of AI technologies in clinical practice.

B2. Investigating deep learning methods for improved analysis of tabular data (such as omics and clinical trials) and integration of multi-modal data: I plan to extend the previously developed table-to-image conversion technique[6] to other important data analysis applications in cancer biology and clinical studies, such as RNA expressions, gene mutation profiles, radiomics, proteomics, circulating tumor DNA (ctDNA), and clinical trials. In addition to addressing some task specific issues (e.g., developing strategy for handling missing data and embedding strategies for categorical data in the case of clinical trials), I will also investigate the use of emerging deep learning techniques such as foundation models for gaining novel insights into cancer biology and accurate clinical decision-making. Finally, methods for seamless integration of data from different modalities (e.g., genomics, radiomics, ctDNA, and medical images) will also be investigated for precision oncology.

B3. Integrating large language models into clinical research and applications: Large Language Models (LLMs) have emerged as powerful tools for natural language understanding and generation.

However, the fine-tuning of these models for specific medical tasks currently relies on heuristic approaches, lacking a systematic framework for optimizing their performance. I will develop innovative techniques to understand and interpret the inner workings of LLMs by delving into the high-dimensional (HD) feature space of these models. By exploiting the feature space using previously developed tool [4], I will investigate an efficient fine-tuning strategy to adapt pre-trained foundation models for personalized AI-augmented clinical decision-making. This approach will enhance model performance while significantly reducing the need for extensive training data in deep learning inference. To demonstrate the potential of the proposed data-efficient AI models, I will conduct extensive retrospective and prospective clinical evaluations. The outcomes of this project will encompass transformative and data-efficient medical AI strategies, accompanied by software tools for visualizing DNN feature space data and facilitating reliable cancer data analysis and decision-making.

B4. Advancing medical imaging and radiation treatment planning decision-making by utilizing state-of-the-art AI techniques: I anticipate that I will continue participating in the development of state-of-the-art deep learning models for improved image segmentation, radiation dose calculation, and super-resolution to advance patient care. Segmentation of tumor target structures and sensitive organs for dose calculation are essential tasks in radiation therapy treatment planning. Super-resolution holds potential to enhance image quality while reducing the cost of imaging system. Research in these areas will not only improve the quality, efficiency and cost-effectiveness of our patient care but also help to advance AI technologies. I will also explore the synergy between these deep learning tasks by developing a single model for improved accuracy and efficiency. I will employ state-of-the-art deep learning architectures, and training strategies to ensure the robustness and generalizability of the proposed system across different clinical scenarios.

C. Research aspirations and collaboration

My career goal is to become an academic leader in biomedical AI with extensive extramural funding support. I aim to establish a premier academic program that leverages my expertise in AI, imaging physics, image analysis, and bioinformatics to advance patient care. While my dedication to an independent research program is unwavering, I deeply value interdisciplinary approaches and collaborative endeavors. I eagerly anticipate dedicating a significant portion of my time to collaborating with esteemed colleagues in different departments across the university. These collaborations are poised not only to enhance my research but also to make substantial contributions to translational AI research within the biomedical community.

References:

1 Islam, M. T., Chaudhry, A., Tang, S., Tasciotti, E. & Righetti, R. A new method for estimating the effective Poisson's ratio in ultrasound poroelastography. *IEEE transactions on medical imaging* **37**, 1178-1191 (2018).
2 Islam, M. T. *et al.* Non-invasive imaging of Young's modulus and Poisson's ratio in cancers in vivo. *Sci Rep* **10**, 7266 (2020).
3 Islam, M. T., Tasciotti, E. & Righetti, R. Estimation of vascular permeability in irregularly shaped cancers using ultrasound poroelastography. *IEEE Transactions on Biomedical Engineering*, 1-1 (2019).
4 Islam, M. T. *et al.* Revealing hidden patterns in deep neural network feature space continuum via manifold learning. *Nature Communications* **in press** (2023).
5 Islam, M. & Xing, L. Deciphering the feature representation of deep neural networks for interpretable and high performance biomedical AI. *IEEE Trans Pattern Anal Mach Intell*, under review (2023).
6 Islam, M. T. & Xing, L. Cartography of genomic interactions enables deep analysis of single-cell expression data. *Nat Commun* **14**, 679 (2023).
7 Islam, M. T. *et al.* Leveraging data-driven self-consistency for high-fidelity gene expression recovery. *Nature Communications* **13**, 7142 (2022).
8 Vasudevan, V. *et al.* Implicit neural representation for radiation therapy dose distribution. *Phys Med Biol* **67** (2022).
9 Huang, C. *et al.* Learning image representations for content-based image retrieval of radiotherapy treatment plans. *Phys Med Biol* **68** (2023).
10 Sang, S., Zhou, Y., Islam, M. T. & Xing, L. Small-object sensitive segmentation using across feature map attention. *IEEE Trans Pattern Anal Mach Intell* **45**, 6289-6306 (2023).
11 Ye, S., Shen, L., Islam, M. T. & Xing, L. Super-resolution biomedical imaging via reference-free statistical implicit neural representation. *Phys Med Biol* **68** (2023).
12 Liu, J., Islam, M. T., Sang, S., Qiu, L. & Xing, L. Biology-aware mutation-based deep learning for outcome prediction of cancer immunotherapy with immune checkpoint inhibitors. *npj Precis. Onc.* **7**, 1-10 (2023).