Research Statement

Ling Tong

My primary research lies at the intersection of health informatics and machine learning. Both disciplines require medical data. Electronic Health Records show an unprecedented opportunity for the use of machine learning in health sciences. Specifically, I have explored the following directions in my doctoral study using machine learning techniques:

1. Predicting Patient Prognosis and Fall Outcomes Using Machine Learning
2. Analyzing biomedical images for osteosarcoma detection
3. Using Machine Learning classification techniques to find adverse events from clinical trials
4. The Interpretation of Machine Learning Diagnostic Models

I will only show the first two pieces of research in this statement because of the page limits.

Topic 1: Patient prognosis of elderly falls and interpretation of machine learning model

Elderly-focused care has become a significant issue for the healthcare system due to the aging population [1]. Elderly falls are one of the major causes of mortality and morbidity among people over 60 years of age [2]. There is an imperative need for an approach to predict and prevent elderly falls in clinical practice. Clinical factors play an important role in clinical decision support systems. Key clinical factors include age, gender, impaired balance and gait, polypharmacy, and a history of previous falls [3]. A clinical decision-making model predict the risk of falls using these clinical features. A machine learning prediction model can quantify the risk of falls for the elderly [4] and prevent elderly falls. One of my conference proceeding [5] shows the effectiveness of a machine learning model to predict the risk of falls for the elderly.

Predicting the risk of falls has significant implications for their prevention. Physicians can suggest precautions for high-risk patients. Proactive alerts reduce the risk of serious injuries. In a broader context, the machine learning algorithm can be used clinically in medical facilities, nursing homes, and other assisted living facilities. However, the adoption of diagnosis-based prediction remains low because of the lack of interpretability. If patients and healthcare providers are unable to understand the model, they will not make informed decisions and adopt the new forms of care.

Despite the importance of model interpretability in the healthcare area, research and application in the medical domain is still far from mature. There is a wide gap between physicians' need to understand the prediction and the model's interpretability. Our ongoing study focuses on the interpretability of machine learning models to address such issues. This includes using SHAP [6] and LIME [7], two interpretable methods proposed as a way to interpret machine learning models from internal data. Therefore, we are working on a paper explaining how the machine learning model in the medical domain works for the layperson and clinical experts. The transparency of the model will be a key factor in encouraging the use of AI-assisted medical decision-making and may benefit the medical community in the future.

Topic 2: Analyzing Osteosarcoma on Histological Images

Primary bone tumors account for 5–10% of all new pediatric cancer diagnoses. Osteosarcoma is the most common form of malignant primary bone tumor under the category of bone tumors. Despite the limited approximately 1,000 new cases every year in the United States, the prognosis of osteosarcoma remains a challenging issue [8]. If not treated, Osteosarcoma often results in metastasis to other parts of the body such as at lungs, other bones, and soft tissues [9].

Diseases diagnosed based on radiological and histological images often get more help in Computer-Aided technologies. Although all diseases are possible with the help of CAD technology, this technique is specialized in pattern recognition in images. Therefore, CAD technology has its own strength in interpreting medical images in both the radiological department and lab tests. The accuracy is also high in both radiological and histological images.

Fortunately, Machine learning shows promising results in detecting Osteosarcoma from histological images. Machine learning algorithms use computer-extracted features, which are called learning materials. Several machine learning techniques have been applied in the past, including linear discriminant analysis, support vector machines, decision trees, random forests, and neural networks [10]. For images, the data is often encoded with RGB-encoding schemas, and sometimes with additional features. With appropriate features, the important histological and radiological image information can be integrated into the above algorithms.

My publication shows that (1) deep learning-based tools are capable of detecting osteosarcoma malignancy with high accuracies based on a public dataset. (2) The proposed deep learning method provides accurate detection, discovering possible features in the field of detecting histological images. The idea is to complement the internal drawbacks of machine learning models. Unlike widely used AI-based techniques such as face recognition and object detection, clinical data sets require significantly higher accuracy to prevent the cost of failure. Based on these thoughts, my lab also explored in clinical trials adverse event detection [16] and interpreting how a machine learning model will make a decision based on doctor's notes [17].

Summary

The AI Leader of Stanford university, Andrew Ng, who proposed that AI will become the new electricity. They will integrate to every part of our life. For example, AI is quietly recommending your favorite Twitter posts, and AI is quietly filtering out email spams. Many Computer Science developments are on the frontier of innovation, yet its application on clinical field always seems to be behind. In current era, the use of AI in medicine are still limited to researcher's domain. In the next decades, I believe we can reach a point that AI will seamlessly integrate into the wide public. The healthcare AI may remind you to take pills on time, helping you to make a doctor's appointment smartly. On the doctor's side, when you take an x-ray picture in radiology department, AI will closely monitor your lung image and check for any problems in case they arise. The outlook of AI application in medical area is clear and promising.

Future work:

In the short term, I will continue to investigate the performance of machine learning algorithms in the clinical domain and find the internal weaknesses, such as lack of interpretability and the need for feature selections. We will need to overcome these problems before applying the machine learning algorithms to the clinical domain.

In the long term, I will explore the possibility of deploying this model in traditional healthcare settings. More importantly, how can deploying such models impact the health outcomes of patients? In many settings, predictive performance cannot easily be translated into causal effects. My approach to this problem will be to map the outputs of a machine learning model to a treatment policy and then to estimate average patient outcomes under that policy using causal estimation methods. Looking for physicians' advices on how to improve the model will be a key part of my future work. The adoption and popularization of AI in healthcare will improve the health outcomes of the public and improve people's quality of life. This is a difficult commitment, and I will strive for the best.

Undergraduate Research Considerations

Machine learning projects are always a hot topic these years. My projects are well-suited for an undergraduate research student, especially for those looking for a Computer Science minor. It is an invaluable opportunity for undergraduate students to learn domain applications in a lab environment. It also has minimal device requirements. First, the necessary equipment only involves computers and databases, and the traditional onsite lab training can be transferred to an online environment. Through working in my lab, Students will master data analysis techniques. I can provide undergraduate researcher positions for the community, develop students' skill sets, and empower the community to contribute to the conversion of digital health in the future. I have specifically designed research projects that are accessible at the graduate level. In addition to undergraduate students, I have successfully trained graduate students under my guidance in several of these statistical techniques. The students I mentor will gain experience in developing their own research plans, generating hypotheses, evaluating data, and preparing posters, talks, and papers. My proposed research will complement existing research and offer projects for both undergraduate and graduate students to perform exciting research in health informatics.

References: [1] S. R. B. L. Shrivastava, P. S. Shrivastava, and J. Ramasamy, "Health-care of Elderly: Determinants, Needs and Services," Int. J. Prev. Med., vol. 4, no. 10, p. 1224, 2013,

[2] W. H. Organization, WHO global report on falls prevention in older age. 2008.

[3] A. F. Ambrose, G. Paul, and J. M. Hausdorff, "Risk factors for falls among older adults: A review of the literature," Maturitas, vol. 75, no. 1, pp. 51–61, May 2013, doi: 10.1016/J.MATURITAS.2013.02.009.

[4] E. H. Shortliffe and J. J. Cimino. Computer applications in health care and biomedicine, 4th ed. Springer, 2014.

[5] Tong, L., Luo, J., et al (2022, July). A Clustering-Aided Approach for Diagnosis Prediction: A Case Study of Elderly Fall. 2022 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC).

[6]. Scott Lundberg. SHAP doc. Accessed 7/15, 2021.

[7]. Mikahail Korobov, ELI5 doc. Accessed 7/15, 2021.

[8] A. J. Chou, D. S. Geller, R. Gorlick, Therapy for osteosarcoma, Pediatric Drugs 10 (2008) 315–327.

[9] P. P. Lin, S. Patel, Osteosarcoma, in: Bone Sarcoma, Springer, 2013, pp. 75–97.

[10] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J. Aerts, Machine learning methods for quantitative radiomic biomarkers, Scientific Reports 5 (1) (2015) 1–11.

[11] A.R. Jamieson, M.L. Giger, K. Drukker, L.L. Pesce, Enhancement of breast CADx with unlabeled data a, Medical Physics 37 (8) (2010) 4155–4172.

[12] A.R. Jamieson, M.L. Giger, K. Drukker, H. Li, Y. Yuan, N. Bhooshan, Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and-SNE, Medical Physics 37. (1) (2010) 339–351

[13] L. Huang, W. Xia, B. Zhang, B. Qiu, X. Gao, MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images, Computer Methods and Programs in Biomedicine 143 (2017) 67–74.

[14] R. Shen, Z. Li, L. Zhang, Y. Hua, M. Mao, Z. Li, Z. Cai, Y. Qiu, J. Gryak, K. Najarian, Osteosarcoma Patients Classification Using Plain X-Rays and Metabolomic Data, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 690–693.

[15] Z. Li, S.R. Soroushmehr, Y. Hua, M. Mao, Y. Qiu, K. Najarian, Classifying osteosarcoma patients using machine learning approaches, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 82–85.

[16] Tong, L., Luo, J., Cisler, R., & Cantor, M. (2019, July). Machine learning-based modeling of big clinical trials data for adverse outcome prediction: A case study of death events. In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC).

[17] Tong, L., Luo, J., Adams, J., Liu, X., Osinski, K., & Friedland, D. Interpretable Machine Learning Text Classification for Computed Tomography Reports–A Case Study of Temporal Bone Fracture.[preprints], SSRN, 4034059.