DEEP LEARNING RISK PREDICTION OF BLOODSTREAM INFECTION IN THE INTENSIVE CARE UNIT

PATIENT TRUST OF AI IN CRITICAL CARE SITUATIONS

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

By

Zachery Boner

October 27, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Kent Wayland, Department of Engineering and Society

Rich Nguyen, Department of Computer Science

General Research Problem: Developing Patient and Clinical Trust in AI

How do we develop patient and clinical trust in AI systems deployed for healthcare applications?

Artificial intelligence and machine learning fields have transformed since the introduction of deep neural networks. Deep neural networks of many architectures have fundamentally changed the scale and power at which models can learn from data. A natural application of these models is in the field of healthcare. The United States' push to standardize Electronic Health Record (EHR) systems resulted in hospitals collecting a massive amount of patient data. This data offers an enticing opportunity to develop models that improve diagnosis, deliver more targeted care, and improve patient outcomes. Examples of such applications include detecting mammogram lesions, analyzing brain MRIs, and identifying risk factors in a patient's EHR. However, researchers must address concerns with models and data before deployment in a clinical setting. Such problems include distributional bias in the data, overfitting in the model, lack of robustness to noisy data, failure to generalize out of distribution, and the difficulty of model interpretation.

These issues contribute to a lack of trust by both clinicians and patients in AI systems. However, model-side problems are not the only factor contributing to this trust breakdown. Focus groups indicate that patients are worried about potential higher costs associated with AI diagnosis, discriminatory use, and a reduction in both patient and doctor autonomy, in addition to technical concerns about AI models (Richardson et al., 2021). In a series of experiments analyzing trust in AI diagnosis relative to human doctors, researchers found that people will trust human doctors over an algorithm even if informed that an AI doctor is more accurate than a human doctor (Juravle et al., 2020). Therefore, AI systems in the healthcare field must meet two criteria: first, the model must outperform human capabilities and minimize bias; second, measures must be taken in each deployment stage to understand and address both clinicians' and patients' points of distrust.

Technical Research Problem: Deep Learning Risk Prediction of Bloodstream Infection in the Intensive Care Unit

Can state-of-the-art deep learning algorithms be used in conjunction with modern EHR data to help doctors identify bloodstream infections in the ICU earlier?

Bloodstream infections are associated with high mortality risk, prolonged hospital stays, and expensive treatment (Rudd et al., 2020). Patients in the intensive care unit (ICU) are at exceptionally high risk of bloodstream infection, given their already critical conditions and the common usage of intravenous catheters in their treatment. The most common pathogens related to bloodstream infection are bacteria and other microbes. The primary treatment option is broad-usage antibiotics, which are becoming decreasingly effective as general antibiotic resistance grows. Furthermore, the detection and diagnosis mechanism for bloodstream infection is a blood culture, which has a turnaround time of up to several days and a risk of contamination, invalidating the result (Bates et al., 1990; Bates et al., 1991; Rupp et al., 2017).

The standard technique for identifying BSI is the Sepsis Inflammatory Response Syndrome (SIRS) criteria; this method, which is used by doctors at the bedside with no computer aid. The SIRS criteria asks whether body temperature, heart rate, respiratory rate, and partial pressure of CO_2 are abnormal; if at least two of them are unusually high or low, the doctor typically recommends a blood culture to be drawn. This system is moderately effective, but it does not factor in all the available data for a patient. Notably, it does not consider any chemical lab tests, or any prior data collected for a patient. Previous studies have demonstrated promising results predicting the presence of bloodstream infection using non-temporal patient data (Pai et al., 2021), predicting mortality of patients diagnosed with a bloodstream infection (Zoabi et al., 2021), and identifying pathological signatures of bloodstream infection (Zimmet et al., 2020). This research project proposes to build on these results using state-of-the-art deep learning approaches to learn from a multivariate time-series dataset collected from the University of Virginia EHR (UVA). We will use full longitudinal data for each patient to make predictions, which is only plausible using deep learning approaches.

This research project uses data from the UVA EHR in the time window 2011-2021. It is longitudinal data with 38 predictor variables, a mix of lab tests, and continuously monitored vital sign data, sampled hourly over a 168-hour episode, and centered around the drawing of a blood culture: 96 hours before the draw and 72 hours after. Models will only have 96 hours of data access before the draw. There are 66,260 total episodes representing 41,681 unique patients in the dataset. There is a significant amount of missing data in this dataset; to this end, this research project will explore methods for imputing data in longitudinal clinical data. There is a high potential for models derived from the data to be biased toward patients from the Charlottesville area specifically. To address this bias, models will be validated using an external dataset from the University of Pittsburgh (Pitt).

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers are the candidate models in this study. These three classes of deep learning architectures are the most widely used in the fields of computer vision and timeseries learning and apply to our data very naturally. CNNs are famously successful in the image domain in accomplishing a breadth of tasks in computer vision. They encode relationships between nearby pixels using the convolution operation and encode longer-distance relationships with pooling operations. RNNs are specifically designed for time-series data and leverage an unrolling procedure to learn both cross-sectional and temporal dynamics in the data. They are explicitly designed for learning from timeseries but are expensive to train and suffer from overfitting. Transformers are a more recent development in the machine learning domain and use a construct called the Attention mechanism to replace both the convolution and recurrence operations (Vaswani et al., 2017). Transformers have demonstrated abundant success in the computer vision field and have superseded both CNNs and RNNs on the well-known ImageNet challenge. This project will use the TensorFlow framework and the service Weights and Biases to implement models, run experiments, and track results to compare models.

This technical research aims to develop a predictive model for bloodstream infection and then extend that result toward a continuous monitoring tool for evaluating the risk of bloodstream infection in a time-aware environment. Once a successful model has been demonstrated and validated externally with data from Pitt, this project could move toward deployment in a real ICU. A successful deployment would help save actual patients' lives by assisting doctors and nurses in the ICU to identify patients at high risk of BSI. This project could lead to future work analyzing the performance of an operational deep learning system in an ICU and improving on such performance.

STS Research Problem: Trust Mechanisms in Automated Patient Monitoring

How do critical care doctors develop trust in automated patient monitoring systems?

An AI system deployed carelessly could substantially harm the patient-doctor relationship. AI in a clinical care setting does not exist in an algorithmic vacuum – unlike in the purely data-driven development environment, decisions made by a deployed algorithm affect actual patients. Alongside these decisions is the relationship between a patient and their doctor. In the acute care setting, this relationship does not have time to develop as in a lifetime care setting; patients in critical condition have a diminished ability to disagree with their doctor or seek a second opinion. Most patients' trust in their doctor is grounded in role-based trust (LaRosa et al., 2018). An AI system that has learned from data but has no human experience in the acute care context can violate role-based trust. It is imperative that, for an AI system to be deployed successfully in a critical care environment, it must not violate the patient-doctor relationship.

LaRosa et al. (2018) argue that there are three primary avenues for patients to develop trust in their doctor. The authors present ways a deployed AI system could undermine all three avenues. The focus group study of Richardson et al. (2021) substantiates this work through group interviews with potential patients. The first route is the license and certification of doctors to practice; this gives patients an implicit understanding of the qualifications and expertise of their care provider. An AI system could ground decision-making in an algorithm rather than a certified doctor. This altered decision-making process would damage role-based trust in the doctor's qualification. Asan et al. (2020) proposed a trust calibration mechanism between a doctor and the AI system to counteract this trust degradation. A second route proposed is the support of a patient's values through the social role of the doctor. An AI decision-making tool could shift a doctor's social role to be simply a user or analyst of AI rather than a "distinctive repository of knowledge and skills" (LaRosa et al., 2018). The third proposed route is the patient's experience with their doctor and the gradual development of two types of trust: reliability and understanding trust. An AI system could sever the line of communication between the patient and the doctor, exposing the patient to a diagnosis they have not had the opportunity to discuss thoroughly.

A series of experiments performed by Juravle et al. (2020) offer a path forward in developing patient trust in AI without harming their relationship with their care provider. These experiments were conducted by walking participants through simulated diagnoses in various scenarios with human and AI doctors. They found that, even if informed that an AI doctor is more accurate than a human one, participants indicated a decreased likelihood of following the AI's recommended treatment plans. However, if nudged towards an AI plan and offered a choice to adopt it, rather than being prescribed a plan and told it was from AI, they found that participants chose an AI's recommendation more often and indicated a higher likelihood of following the plan. This study suggests that patient empowerment is essential in developing trust in AI. This patient empowerment originates in the care provider's interactions with the patient.

The STS research will explore trust mechanisms between doctors and automated health monitoring systems and will focus on the CoMET (Continuous Monitoring of Event Trajectories) score system deployed in the UVA Medical ICU (Ruminski, et al., 2019). It will use this case study to investigate the ways in which doctors develop and lose trust in predictive monitoring systems and explore the effect of predictive monitoring on doctors' interactions with their patients. This ethnographic research will be conducted in three stages: first, we will conduct interviews with doctors, biomedical engineers, and ICU staff to understand each of these groups' unique perspectives on the CoMET score. Next, we will observe real-time interactions of doctors and nurses in the ICU by shadowing in the ICU. Finally, we will conduct interviews with ICU doctors after the observation to ask follow-up questions about what we learned.

Conclusion

The STS research project aims to understand how doctors mediate trust with predictive monitoring systems and how their trust in those systems influences their interactions with their patients. By understanding this trust mechanism, the research will derive insight into the efficacy and adoption by doctors of existing successful predictive monitoring solutions. The technical research aims to develop a predictive algorithm to identify bloodstream infections in ICU patients to help inform clinical action. It will do so by leveraging a large dataset of patients in the UVA EHR, alongside state-of-the-art approaches in deep learning, to learn patterns indicating BSI in those patients. Each of these projects addresses a component of the larger general problem presented: the STS research will seek to understand trust mechanisms associated with a deployed predictive monitoring system, and the technical research will aim to develop a successful system that might be eventually adopted.

References

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research*, *22*(6), e15154. https://doi.org/10.2196/15154

Bates, D. W. (1990). Predicting Bacteremia in Hospitalized Patients: A Prospectively Validated Model. *Annals of Internal Medicine*, *113*(7), 495. https://doi.org/10.7326/0003-4819-113-7-495

Bates, D. W. (1991). Contaminant Blood Cultures and Resource Utilization: The True Consequences of False-Positive Results. *JAMA*, *265*(3), 365. https://doi.org/10.1001/jama.1991.03460030071031

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare* (pp. 25–60). Elsevier. https://doi.org/10.1016/B978-0-12-818438-7.00002-2

Juravle, G., Boudouraki, A., Terziyska, M., & Rezlescu, C. (2020). Trust in artificial intelligence for medical diagnoses. In *Progress in Brain Research* (Vol. 253, pp. 263–282). Elsevier. https://doi.org/10.1016/bs.pbr.2020.06.006

LaRosa, E., & Danks, D. (2018). Impacts on Trust of Healthcare AI. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 210–215. https://doi.org/10.1145/3278721.3278771

Pai, K.-C., Wang, M.-S., Chen, Y.-F., Tseng, C.-H., Liu, P.-Y., Chen, L.-C., Sheu, R.-K., & Wu, C.-L. (2021). An Artificial Intelligence Approach to Bloodstream Infections Prediction. *Journal of Clinical Medicine*, *10*(13), 2901. https://doi.org/10.3390/jcm10132901

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, *27*(3), 491–497. https://doi.org/10.1093/jamia/ocz192

Richardson, J. P., Smith, C., Curtis, S., Watson, S., Zhu, X., Barry, B., & Sharp, R. R. (2021). Patient apprehensions about the use of artificial intelligence in healthcare. *Npj Digital Medicine*, *4*(1), 140. https://doi.org/10.1038/s41746-021-00509-1

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., Colombara, D. V., Ikuta, K. S., Kissoon, N., Finfer, S., Fleischmann-Struzek, C., Machado, F. R., Reinhart, K. K., Rowan, K., Seymour, C. W., Watson, R. S., West, T. E., Marinho, F., Hay, S. I., … Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *The Lancet*, *395*(10219), 200–211. https://doi.org/10.1016/S0140-6736(19)32989-7

Ruminski, C. M., Clark, M. T., Lake, D. E., Kitzmiller, R. R., Keim-Malpass, J., Robertson, M. P., Simons, T. R., Moorman, J. R., & Calland, J. F. (2019). Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *Journal of Clinical Monitoring and Computing*, *33*(4), 703–711. https://doi.org/10.1007/s10877-018-0194-4

Rupp, M. E., Cavalieri, R. J., Marolf, C., & Lyden, E. (2017). Reduction in Blood Culture Contamination Through Use of Initial Specimen Diversion Device. *Clinical Infectious Diseases*, *65*(2), 201–205. https://doi.org/10.1093/cid/cix304

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (Arxiv). https://doi.org/10.48550/ARXIV.1706.03762

Zimmet, A. N., Clark, M. T., Gadrey, S. M., Bell, T. D., Zimmet, A. M., Moorman, J. R., & Moore, C. C. (2020). Pathophysiologic Signatures of Bloodstream Infection in Critically Ill Adults. *Critical Care Explorations*, *2*(10), e0191. https://doi.org/10.1097/CCE.0000000000000191

Zoabi, Y., Kehat, O., Lahav, D., Weiss-Meilik, A., Adler, A., & Shomron, N. (2021). Predicting bloodstream infection outcome using machine learning. *Scientific Reports*, *11*(1), 20101. https://doi.org/10.1038/s41598-021-99105-2