Prospectus

Deep Learning for Bloodstream Infection Detection
(Technical Topic)

The Lab to Clinic Roadmap for Computer Aided Diagnosis
(STS Topic)

By

Matthew Pillari

March 29, 2022

Technical Project Team Members: Rich Nguyen's BSI Research Team

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.

Signed: Matthew P. Pillari
Technical Advisor: Rich Nguyen, Department of Computer Science
STS Advisor: Caitlin D. Wylie, Department of Engineering and Society

Introduction

Medical error is the third leading cause of death in the United States. This equates to 250,000 deaths every year (Rodziewicz et al., 2022). It is safe to assume that hundreds of thousands more have suffered needlessly or have become permanently disabled due to preventable mistakes in diagnosis and care. This paper explores diagnostic error, which causes more harm than all other medical errors combined according to Johns Hopkins professor Makary (2016) in “Medical error—the third leading cause of death in the US”. One technological solution to widespread diagnostic error is computer-aided diagnosis (CAD). CAD, for the purposes of this paper, is a software tool that automates the diagnostic process. Due to recent developments in deep learning, CAD has begun to surpass human radiologist performance for an increasing number of pathologies. In addition to better accuracy, CAD has other major benefits over human radiologists. CAD produces a diagnosis nearly instantly, compared to 5-15 minutes for a radiologist, and since CAD is software, it can be deployed at any hospital, allowing for even the most rural facilities access to the best diagnosis available. However, CAD faces one major roadblock that will take decades to overcome at the current pace: Adoption. The technology to diagnose many diseases exists, the national digitization of medical data is nearly complete, but nearly every diagnosis is still done by un-aided radiologists (Taylor, 2019). In order to connect the engineering of CAD to the human reality of medical care, an STS perspective is imperative. In this paper specifically, a technical case study of a blood stream infection detection model will lead into an STS discussion of how to bring a CAD from the lab into widespread clinical adoption.

Deep Learning for Bloodstream Infection Detection

Bloodstream infections (BSI) are a common and deadly occurrence in intensive care units worldwide. A major factor in their deadliness is the difficulty of detecting an infection, as they manifest very quickly. BSIs cause about 85,000 deaths per year (Ring, 2018). In the intensive care unit (ICU) of major hospitals, about six percent of patients acquire a BSI (Ring, 2018). About a quarter of BSI patients die due to the infection. This means BSI takes more lives than heart attacks or breast cancer (Ring, 2018). The reason BSI is so deadly is that it can kill an otherwise healthy person in as little as 12 hours, with very few effective treatments. BSI has very few understood warning signs, and requires near-immediate use of antibiotics as the risk of death increases drastically as the hours pass. One promising technical solution to BSI deaths in the ICU is to create a model or metric that can help doctors predict the onset of BSI based on continuously recorded vital signs. The model would then notify doctors as early as possible of a potential BSI and antibiotics can be administered to eliminate the infection before it reaches a critical level. The current state of the art for this problem has an AUC, a measure of the tradeoff between true positive and false positive where closer to 1 is better, of 0.83 (Zoabi et al., 2021). Improvement over the state-of-the-art may require a deep-learning approach.

The UVA-BSI research team at UVa, led by Dr. Nguyen, attempts to improve over Zoabi et al. (2021). UVA-BSI aims to improve over prior art by increasing the amount of data used to create the model, applying a deep-learning approach, while allowing for a better understanding of the reasoning behind the diagnosis. The UVA-BSI team works closely with doctors from the ICU at UVA's Hospital, as well as the division for infectious diseases. UVA-BSI uses data from UVa, Pitt, and other institutions in order to create a larger and more diverse dataset than previous efforts. The deep learning model-type of greatest interest is based on the idea of a "long short-term memory" (LSTM) module. LSTM models have performed excellently on other complex problems that involve continuous, time-related data like vital signs. Additionally, LSTM research by Gau et al. (2019) provides a method for transparency into the reasoning that takes place in traditionally opaque deep learning models. All of these improvements may result in a new state of the art for bloodstream infection detection.

The Lab to Clinic Roadmap for Computer Aided Diagnosis

Machine learning and large scale medical data has allowed computer-aided diagnostics to take massive technical strides, yet clinical adoption remains nearly non-existent (Cai, 2019). The patient data exists on hospital servers, the diagnostic models to process this data exist in research labs, but the journey from the lab to the clinical setting seems to never quite happen. The major inhibitors are regulation, incentives, and trust (Pinto, 2018). The FDA is the regulatory authority for medical devices, including medical software like CAD. In the past, the FDA has had little issue with evaluating medical software that automates simple tasks. However, deep learning has presented a new issue for the FDA reviewers; there is typically no way to get to the root of the reasoning behind a deep learning model's diagnosis. Deep learning models are black boxes; medical data is provided as input, a diagnosis is output, but the steps in-between are an inscrutably complex math equation, akin to a plinko board with millions of pegs. A small change in the input, like a few pixels of difference in an X-ray, can drastically change the output. With few exceptions—see Grzybowski's (2021) work with diabetic retinopathy—black-box models have been rejected by FDA review.

The dynamics of ideation, development, and deployment of AI powered CAD are highly technical as well as highly interpersonal and social. Actor network theory (ANT) gives us a lens to view the issue by breaking down the dynamic into a more understandable format (Law, 1992). When we look at the issue with a broad ANT lens, we see our actors are developers, regulators, doctors, and patients. This paper takes on the perspective of a developer who wants to create a medical AI software product that is widely used and improves patient outcomes. However, to create such a product, the developer must understand the perspectives of each other actor, as well as how their own perspective may introduce bias into the work. Developers are typically unfamiliar with the intricacies of medicine. While a developer may be very familiar with the technicalities of creating a deep learning diagnostic model, medical diagnosis is not straightforward, and as such, a medical professional who is practiced in whatever diagnosis is being automated is crucial to creating a useful piece of software. Without a developer-doctor relationship during the development stage, a product may be solving a problem that does not exist, or potentially worse, solving a real problem incorrectly. Further connections between actors at different stages of the development process all contribute to a complete ANT understanding of creating and deploying an AI medical device.

Medical students, who represent the first class of doctors who are arriving into hospitals at the same time as automated diagnostic software, are also not completely sold on the idea. Pinto et al. (2018) investigatigation of medical student attitude toward AI medical devices found that medical students have little to no understanding of medical AI devices. Amann et al. (2020) acknowledge the power of AI models and their approaching superiority over human physicians, but warn that such systems need to be developed with great care for transparency and openness, or they will be rejected by doctors, patients, and regulators.

On top of FDA hesitation, doctors are understandably unwilling to offload life-critical decisions to a black box. Even though radiologists are understaffed and overworked compared to diagnostic demand, surveys show they are still largely against the adoption of AI in their practice (Ahuja, 2019). The solution to both FDA and doctor hesitancy is to provide sufficient diagnostic explainability as well as demonstration of performance. One idea to build trust with radiologists is to run a CAD model on a radiologist's prior diagnosis to demonstrate missed cases.

Trust is discussed as one of the "Three Ghosts of Medical AI" by Quinn et al. (2021). The three ghosts represent three major challenges that blackbox models have in the medical space. The authors point out that the main issues preventing adoption by regulators and doctors are that blackbox models "(1) lack quality assurance, (2) fail to elicit trust, and (3) restrict physician-patient dialogue" (Pinto, 2018, p. 1). The only software that can successfully traverse the road from the lab to successful implementation is software that can overcome the above listed shortcomings of medical AI products.

This psychological approach to trust building is just one example of using STS thinking to bring an engineering solution to a wider audience. The result of this STS research will be an outline of a realistic pathway that a deep learning engineer could take to bring a deep learning model from concept to life-saving clinical adoption.

Conclusion

The steady march of technological progress in medicine is beginning to break into an all out sprint. However, the high stakes involved in medical care have fostered an appropriately hesitant attitude towards the increased pressure to adopt new technologies. Doctors rely heavily on their own senses and intuition, as well as established and peer reviewed methods to carry out care. Now that deep learning powered computer aided diagnostics are surpassing human diagnostic ability, medical practice has no choice but to begin adopting these technologies. But accuracy is not the only metric that matters. In the end, patient wellbeing is at stake, and non-technical metrics like patient-doctor communication, explanation of the reasoning behind a diagnosis, as well as repeated and provable quality are just as important to the successful integration of CAD technologies.

References

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-01332-6

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359206

Grzybowski, A., & Brona, P. (2021). Analysis and Comparison of Two Artificial Intelligence Diabetic Retinopathy Screening Algorithms in a Pilot Study: IDx-DR and Retinalyze. *Journal of Clinical Medicine*, *10*(11), 2352. https://doi.org/10.3390/jcm10112352

Law, J. (1992). Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity. *Systems Practice*, 5, 379–393.

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., & Potter, H. (2018). Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences*, *115*(45), 11591–11596. https://www.pnas.org/content/115/45/11591

Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ*, *353*(353), i2139. https://doi.org/10.1136/bmj.i2139

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

Park, H. J., Jung, D. Y., Ji, W., & Choi, C.-M. (2020). Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study. *Journal of Medical Internet Research*, *22*(8), e19512. https://doi.org/10.2196/19512

Pinto dos Santos, D., Giese, D., Brodehl, S., Chon, S. H., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2018). Medical students' attitude towards artificial intelligence: a multicentre survey. *European Radiology*, *29*(4), 1640–1646. https://doi.org/10.1007/s00330-018-5601-1

Quinn, T. P., Jacobs, S., Senadeera, M., Le, V., & Coghlan, S. (2021). The Three Ghosts of Medical AI: Can the Black-Box Present Deliver? *Artificial Intelligence in Medicine*, *124*, 102158. https://doi.org/10.1016/j.artmed.2021.102158

Quinn, T. P., Senadeera, M., Jacobs, S., Coghlan, S., & Le, V. (2020). Trust and medical AI: the challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association*, *28*(4), 890–894. https://doi.org/10.1093/jamia/ocaa268

Rodziewicz, T. L., Houseman, B., & Hipskind, J. E. (2022). Medical Error Reduction and Prevention. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK499956/

Straw, I. (2020). The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future. *Artificial Intelligence in Medicine*, *110*, 101965. https://doi.org/10.1016/j.artmed.2020.101965

Taylor, J., & Fenner, J. (2019). The challenge of clinical adoption—the insurmountable obstacle that will stop machine learning? *BJR|Open*, *1*(1), 20180017. https://doi.org/10.1259/bjro.20180017

Wang, B., Jin, S., Yan, Q., Xu, H., Luo, C., Wei, L., Zhao, W., Hou, X., Ma, W., Xu, Z., Zheng, Z., Sun, W., Lan, L., Zhang, W., Mu, X., Shi, C., Wang, Z., Lee, J., Jin, Z., & Lin, M. (2020). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. *Applied Soft Computing*, *98*, 106897. https://doi.org/10.1016/j.asoc.2020.106897

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nature Medicine*, *27*(4), 582–584. https://doi.org/10.1038/s41591-021-01312-x

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). https://doi.org/10.1038/s41598-021-99105-2