Determining predictive models of bloodstream infection by using big data and deep learning
University of Virginia Global Infectious Diseases Institute; Collaborative Seed Grant Proposal

Investigators. University of Virginia: Christopher C. Moore, MD (Contact PI), Associate Professor, School of Medicine, Division of Infectious Diseases and International Health; J. Randall Moorman, MD (Co-PI), Professor, School of Medicine, Departments of Medicine and Biomedical Engineering, Division of Cardiovascular Medicine; Nhat Nguyen, PhD (Co-PI), Assistant Professor, School of Engineering and Applied Science, Department of Computer Science; Amanda Zimmet, PhD (Post-doctoral research scientist), Department of Medicine, Division of Cardiovascular Medicine; Alex Zimmet, MD (Medical resident trainee), Department of Medicine; Bobby Andris, BS (Research Assistant), Department of Medicine, Division of Cardiovascular Medicine; Computer Science undergraduate students, to be determined.
University of Pittsburgh: Gilles Clermont, MD (Co-investigator), Professor, Department of Critical Care

Significance. Sepsis is a syndrome of critical illness defined as life-threatening organ dysfunction due to a dysregulated host response to infection. It is the leading cause of global mortality causing approximately 1 out of every 5 global deaths[1,2]. Bloodstream infections (BSIs) are a frequent cause of sepsis and are associated with a high mortality rate[3–5]. Clinical signs and symptoms of BSI are nonspecific, and published guidelines do not provide clear indications for when it is most appropriate to obtain blood cultures[6]. Therefore, clinicians often have a low threshold to perform blood culture testing[7]. This leads to a low diagnostic yield, with true positive rates of only 4%-7%[7,8]. In addition, false positives (*i.e.* contaminants) lead to unnecessary antibiotic use and associated toxicities, increased cost, and length of hospital stay[9–13]. We systematically reviewed 25 studies assessing 256 unique predictors of BSI. Attempts to compose multivariable scoring systems and clinical prediction rules for BSI primarily involved categorical variables (*e.g.* heart rate >90 bpm or <90 bpm), utilized relatively small data sets, and led to only moderate predictive efficacy[14–17]. However, routine physiological monitoring data contain dynamic multivariable signatures of critical illness states, and there is growing recognition of the heterogeneity of sepsis physiology, including the recent identification of four physiological phenotypes of sepsis[18,19]. Applying this type of detailed physiologic modeling to BSI prediction could improve earlier recognition of infected patients, maximize diagnostic stewardship of blood cultures, measure the effectiveness of treatments, and identify opportunities for development of novel therapeutics.

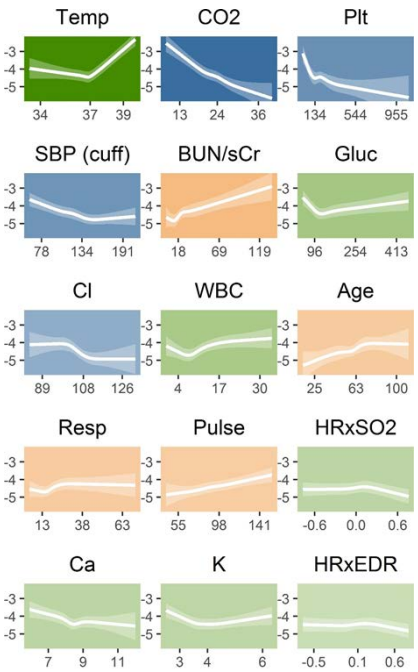

Figure 1. Physiological features comprising a signature of BSI in 9,954 critically ill patients, 2011-2015. Each tile plots the value of the feature on the x-axis against the log-odds of BSI on the y-axis. The translucent ribbon represents the 95% CI. Background hue represents the direction of association with BSI - orange indicates positive association, blue indicates negative association, and green indicates nonlinear or biphasic association. Background color saturation indicates the strength of the association with BSI.

Proof-of-principle. We tested the hypothesis that routine monitoring data may identify a detailed physiological phenotype of BSI in critically ill adult patients. We analyzed 9,954 UVA intensive care unit (ICU) admissions from 2011-2015 with over 144 patient-years of vital sign and cardiorespiratory monitoring (CRM) data, totaling 1.3 million hourly measurements. These admissions included 15,577 episodes of blood culturing leading to 1,184 instances of BSI. The multivariable physiological signature of BSI was characterized by abnormalities in 15 different physiological features, and the area under the receiver operating characteristic curve for the logistic regression model was 0.78 (**Figure 1**). **Figure 2** shows the predicted risk of BSI as a function of time to blood culture and time to systemic antimicrobial administration. For patients with positive cultures, the average predicted risk began to rise 24-48 hours prior to the time of culture (**panel A**). The predicted risk peaked three hours after cultures were obtained – with a greater than twofold increase in risk compared to baseline – and slowly dropped over 48 hours. Similarly, the relative risk of BSI increased up until the time of antibiotic administration and was then slowly reduced over 48 hours (**panel B**).

Approach. We propose to capitalize on the large amounts of continuous physiological data captured in the ICUs of multiple hospitals to develop stronger predictive models of BSI than previously published models. With access to an updated UVA dataset from 2015 to the current day, and a large dataset from the University of Pittsburgh Health System (Pitt), we

plan to expand this work to externally validate our previously derived predictive models and to create new models of BSI. One solution to the problem of creating an accurate predictive monitoring tool is to exploit machine learning, a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed. Instead, pattern recognition, predictions, and generation of data-driven hypotheses from massive amounts of multidimensional data can be done with automatically learning algorithms, where computers build models that are not explicitly programmed in advance, based on the data being fed to them[20]. These models can learn extremely complex relationships between data inputs and outcomes, and have exceeded human abilities in performing classification tasks similar to identifying patients at risk for BSI[20,21].

Deep learning is different from traditional machine learning in how representations are learned from the raw data. In deep learning, computational models are composed of multiple processing layers based on neural networks to learn representations of data with multiple levels of abstraction. Every layer of a deep learning system produces a representation of the observed patterns based on the data it receives as inputs from the layer below. For classification tasks such as identifying BSI, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations. Our deep learning algorithms will change with the input of new clinical data, and will integrate a variety of clinical parameters such as continuous cardiorespiratory monitoring data (CRM), comorbidities, and the trend of clinical and laboratory parameters. Once established, these techniques can also be used to identify specific types of BSI including Gram positive (GP) and Gram negative (GN) bacteria, and fungi; drug resistant bacteria; other types of clinical infection leading to bloodstream infection, *e.g.* pneumonia and intra-abdominal infection; and other adverse outcomes such as ICU transfer, emergent intubation, and hemorrhage. The large size of our dataset will allow us to be the first group to apply deep learning techniques to BSI prediction.

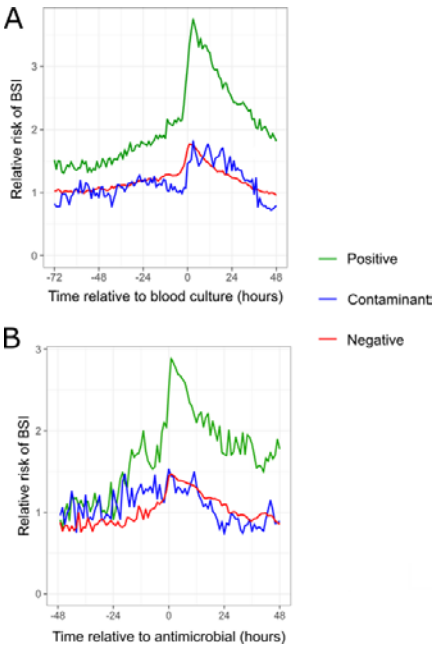

Figure 2. Predicted risk of bloodstream infection according to a multivariable logistic regression model as a function of time relative to blood culture (**panel A**) and time relative to antimicrobial administration (**panel B**) in 9,954 critically ill patients, 2011-2015. Patient data were grouped as positive, negative, or contaminant based on the ultimate result of the culture instance obtained at time 0. Abbreviations: BSI, bloodstream infection.

Our hypothesis is that our deep learning algorithms will lead to significantly better predictions of BSI than both conventional representation learning methods (e.g. Principal Component Analysis, k-means) and logistic regression with only physiological features.

Aim 1. Externally validate previously derived predictive models of bloodstream infection. To externally validate our prior models, we will apply our prior models of BSI to the updated UVA dataset and the Pitt dataset. The updated UVA dataset will include roughly 52,000 patients, 75,000 blood cultures, and 3,000 BSI episodes. The Pitt dataset includes multiple hospitals, including regional and tertiary care academic hospitals, totaling ~200,000 patients, ~266,000 blood cultures, and ~10,000 episodes of BSI. Initially, we will collate and clean the data from both UVA and Pitt for validation of our prior models and deep learning.

Aim 2. Create deep learning algorithms for predicting bloodstream infection. We will employ sophisticated deep learning approaches to train models and select variables that will predict the onset of BSI in hospitalized patients. We will focus on emerging techniques such as deep neural networks with baseline comparisons to conventional methods such as logistic regression, random forests, and support vector machines. We will compare the performance and clinical interpretability of these models to each other to select the best model. We will also develop individualized subgroup-specific models for unique categories of BSI (*i.e.* GP, GN, fungal, and drug resistant infections) and in special clinical populations such as transplant recipients whose clinical and physiological response to BSI may be different than non-transplant recipients due to immunosuppression.

We propose to use deep learning to combine both CRM data and physiological abnormality features in a joint representation for analysis and prediction. First, we will use unsupervised learning to derive a latent representation of CRM using a Stacked Denoising Autoencoder (SDA)[23]. Then, we will combine this latent representation with the physiological features from vital signs and laboratory tests. We plan to detect any

irregularities, noisy, or missing signals in CRM data by employing Recurrent Neural Networks (RNNs) architecture with a multi-layer long short-term memory (LSTM) and a decay effect[26,27]. This deep architecture can recognize patterns in multivariate time series with several million measurements[21,22]. In order to improve the interpretability of the model, we will consider a simple yet effective interpretation measure called feature importance, which indicates the statistical contribution of each feature to the model. This approach is efficient since the physiological abnormality features are generally explainable by a clinician.

Despite the promises of using deep learning in healthcare, there remain several potential challenges which we expect to overcome: **1) Data**: Deep learning relies on the availability of a massive amount of data. The addition of the Pitt data will increase our data ten-fold and allow us to train a comprehensive deep learning model. Additionally, patient data tend to be heterogeneous, ambiguous, noisy, and incomplete[28]. Our deep learning approach is designed to handle data sparsity, redundancy, and missing values; **2) Interpretability**: In healthcare, not only is quantitative algorithmic performance essential, but the reason why the algorithm works is also relevant. Such model interpretability is crucial for convincing medical professionals about the action recommended by the prediction systems[29]. Clinicians are unlikely to adopt a system they cannot understand, so we will build-in clinical interpretability in our deep learning models.

Roles. Dr. Moore: coordination of analyses with Co-PIs, clinical interpretation of data and models with Dr. Alex Zimmet; Dr. Moorman: supervision of Dr. Amanda Zimmet and Mr. Andris in data acquisition and cleaning (UVA and Pitt), and logistic regression analysis; Dr. Nguyen: supervision of Dr. Amanda Zimmet, Mr. Andris, and computer science students in deep learning models; Dr. Clermont: data acquisition (Pitt); consultation on deep learning.

Potential impact. Deep learning algorithms may be able to augment a clinician's bedside perception of the likelihood of BSI in a given critically ill patient by their ability to rapidly integrate and process massive amounts of physiological data. This in turn could lead to earlier diagnostic and therapeutic interventions, and thus, better outcomes. For example, in a study by Co-PI, Dr. Moorman, bedside displays of sepsis risk based on physiological data reduced mortality in very low birth weight infants, even without establishing thresholds or guidelines for interpretation[24]. Successful prediction models created during the course of this funding period could be tested prospectively in a clinical trial to determine their impact on earlier diagnostic and therapeutic interventions and patient outcomes[25]. Furthermore, recognition of patients *unlikely* to have BSI may prevent acquisition of unnecessary blood cultures, which are expensive and often lead to contaminated samples and unnecessary antibiotics along with their potential toxicities. A tangible and crucial impact of this GIDI grant will be the ability to use preliminary data to submit additional grants to continue this line of investigation, and build capacity for academic careers in our trainees (a postdoctoral scientist, a clinical resident, and Computer Science students).

Project milestones. 6-month: 1) Collect, clean, and analyze patient care data from Pitt and an updated UVA dataset, including understanding the available variables and collection standards (*e.g.* how often these variables are collected across units and in what patient populations); 2) validate our UVA derived BSI model; 3) begin deep learning modeling of BSI. 9-month: Validate new deep learning models including SDA, the investigation in the deep representation of the RNN that might contain information for interpretation, and a model explanation which might quantify the contribution of individual physiological features to the BSI prediction. 12-month: We will complete deep learning modeling, write papers, and submit extramural grant proposals.

Extramural support. There are several possible mechanisms of funding to support this work through prospective validation and clinical trials. We anticipate a submission date of Spring/Summer 2021 for a large, collaborative National Institutes of Health cooperative (U01) or program project (P01) grant with Pitt. We also anticipate submitting multiple applications to the Department of Defense (DoD) and Medical Technology Engagement Consortium (MTEC) funding announcements. We will also pursue Biomedical Advanced Research Development Authority (BARDA) opportunities, as well as any relevant program announcements that occur during the project.

References Cited

1. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, Colombara DV, Ikuta KS, Kissoon N, Finfer S, Fleischmann-Struzek C, Machado FR, Reinhart KK, Rowan K, Seymour CW, Watson RS, West TE, Marinho F, Hay SI, Lozano R, Lopez AD, Angus DC, Murray CJL, Naghavi M. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet. 2020 Jan 18;395(10219):200–211. PMID: 31954465
2. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, Van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(0098-7484 (Linking)):801–810. PMID: 26903338
3. Weinstein MP, Reller LB, Murphy JR, Lichtenstein KA. The clinical significance of positive blood cultures: a comprehensive analysis of 500 episodes of bacteremia and fungemia in adults. I. Laboratory and epidemiologic observations. Rev Infect Dis. 1983 Feb;5(1):35–53. PMID: 6828811
4. Weinstein MP, Towns ML, Quartey SM, Mirrett S, Reimer LG, Parmigiani G, Reller LB. The clinical significance of positive blood cultures in the 1990s: a prospective comprehensive evaluation of the microbiology, epidemiology, and outcome of bacteremia and fungemia in adults. Clin Infect Dis. 1997 Apr;24(4):584–602. PMID: 9145732
5. Laupland KB, Church DL. Population-based epidemiology and microbiology of community-onset bloodstream infections. Clin Microbiol Rev. 2014 Oct;27(4):647–664. PMCID: PMC4187633
6. Baron EJ, Miller JM, Weinstein MP, Richter SS, Gilligan PH, Thomson RB, Bourbeau P, Carroll KC, Kehl SC, Dunne WM, Robinson-Dunn B, Schwartzman JD, Chapin KC, Snyder JW, Forbes BA, Patel R, Rosenblatt JE, Pritt BS. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a). Clin Infect Dis. 2013 Aug;57(4):e22–e121. PMCID: PMC3719886
7. Linsenmeyer K, Gupta K, Strymish JM, Dhanani M, Brecher SM, Breu AC. Culture if spikes? Indications and yield of blood cultures in hospitalized medical patients. J Hosp Med. 2016;11(5):336–340. PMID: 26762577
8. Bates DW, Cook EF, Goldman L, Lee TH. Predicting bacteremia in hospitalized patients. A prospectively validated model. Ann Intern Med. 1990 Oct 1;113(7):495–500. PMID: 2393205
9. Bates DW, Goldman L, Lee TH. Contaminant blood cultures and resource utilization. The true consequences of false-positive results. JAMA. 1991 Jan 16;265(3):365–369. PMID: 1984535
10. Hall KK, Lyman JA. Updated review of blood culture contamination. Clin Microbiol Rev. 2006 Oct;19(4):788–802. PMCID: PMC1592696
11. Roth A, Wiklund AE, Pålsson AS, Melander EZ, Wullt M, Cronqvist J, Walder M, Sturegård E. Reducing blood culture contamination by a simple informational intervention. J Clin Microbiol. 2010 Dec;48(12):4552–4558. PMCID: PMC3008442
12. Dawson S. Blood culture contaminants. J Hosp Infect. 2014 May;87(1):1–10. PMID: 24768211
13. Rupp ME, Cavalieri RJ, Marolf C, Lyden E. Reduction in Blood Culture Contamination Through Use of Initial Specimen Diversion Device. Clin Infect Dis. 2017 Jul 15;65(2):201–205. PMCID: PMC5849098
14. Coburn B, Morris AM, Tomlinson G, Detsky AS. Does this adult patient with suspected bacteremia require blood cultures? JAMA. 2012 Aug 1;308(5):502–511. PMID: 22851117
15. Jaimes F, Arango C, Ruiz G, Cuervo J, Botero J, Vélez G, Upegui N, Machado F. Predicting bacteremia at the bedside. Clin Infect Dis. 2004 Feb 1;38(3):357–362. PMID: 14727205
16. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. Who needs a blood culture? A prospectively derived and validated prediction rule. J Emerg Med. 2008 Oct;35(3):255–264. PMID: 18486413
17. Jones GR, Lowes JA. The systemic inflammatory response syndrome as a predictor of bacteraemia and outcome from sepsis. QJM. 1996 Jul;89(7):515–522. PMID: 8759492
18. Moss TJ, Lake DE, Calland JF, Enfield KB, Delos JB, Fairchild KD, Moorman JR. Signatures of Subacute Potentially Catastrophic Illness in the ICU: Model Development and Validation. Crit Care Med. 2016 Sep;44(9):1639–1648. PMCID: PMC4987175
19. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, van der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, Yende S, Angus DC. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019 28;321(20):2003–2017. PMCID: PMC6537818
20. Bailly S, Meyfroidt G, Timsit J-F. What's new in ICU in 2050: big data and machine learning. Intensive Care

Med. 2018 Sep 1;44(9):1524–1527.
21. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. arXiv:160200357 [cs, stat] [Internet]. 2017 Apr 10 [cited 2020 Feb 26]; Available from: http://arxiv.org/abs/1602.00357
22. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to Diagnose with LSTM Recurrent Neural Networks. arXiv:151103677 [cs] [Internet]. 2017 Mar 21 [cited 2020 Feb 26]; Available from: http://arxiv.org/abs/1511.03677
23. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. J Mach Learn Res. 2010 Dec 1;11:3371–3408.
24. Moorman JR, Carlo WA, Kattwinkel J, Schelonka RL, Porcelli PJ, Navarrete CT, Bancalari E, Aschner JL, Walker MW, Perez JA, Palmer C, Stukenborg GJ, Lake DE, O'Shea TM. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. J Pediatr. 2011 Dec;159(6):900-906.e1. PMCID: PMC3215822
25. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, Schenk J, Terwindt LE, Hollmann MW, Vlaar AP, Veelo DP. Effect of a Machine Learning–Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA [Internet]. 2020 Feb 17 [cited 2020 Feb 21]; Available from: https://jamanetwork.com/journals/jama/fullarticle/2761469
26. Kádár A, Chrupała G, Alishahi A. Representation of linguistic form and function in recurrent neural networks. Computational Linguistics. 2017 Dec;43(4):761-80.
27. Karpathy A, Johnson J, Fei-Fei L. Visualizing and understanding recurrent networks. arXiv preprint arXiv:1506.02078. 2015 Jun 5.
28. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics. 2018 Nov;19(6):1236-46.
29. Heinrichs B, Eickhoff SB. Your evidence? Machine learning algorithms for medical diagnosis and prediction. Human Brain Mapping. 2019 Dec 5.