Artificial Intelligence for Anesthesia: What the Practicing Clinician Needs to Know

Michael R. Mathis, MD[1] **[Assistant Professor]**, **Sachin Kheterpal, MD, MBA**[1] **[Associate Professor]**, and **Kayvan Najarian, PhD**[2] **[Associate Professor]**

[1]Department of Anesthesiology, University of Michigan Health System

[2]Department of Computational Medicine and Bioinformatics, and the Department of Emergency Medicine, University of Michigan Health System

Editorial

Machine learning, the quintessential tool currently driving forward the development of artificial intelligence, was discovered and developed decades ago. Nevertheless, only recently has machine learning seen an exponential increase in growth, sophistication, and influence. Recent success stories outside healthcare are numerous, including: in 2014 Facebook unveiled DeepFace, a machine learning technology capable of identifying faces with 97.25% accuracy (compared to human accuracy of 97.53%).[1] In 2016 Google adopted a deep learning approach to language translation, using an algorithm which is fed massive amounts of data to effectively train itself to recognize patterns in speech, with a reduction in translation errors by 87%.[2]

Machine learning techniques like these may be coming soon to an operating room near you: in this issue, we explore three examples of machine learning applied to our field. These include works by Lee *et al.*, using machine learning techniques to predict postoperative

Editorial Corresponding Author: Michael R. Mathis, MD (Assistant Professor) mathism@med.umich.edu, Department of Anesthesiology, University of Michigan, 1H247 UH, SPC 5048, 1500 East Medical Center Drive, Ann Arbor, MI 48109-5048, Phone: 734-936-4280, Fax: 734-936-9091.

mortality from electronic health record data,[3] and works by Kendale *et al.*, and Hatib *et al.*, predicting hypotension through machine learning algorithms leveraging data available during induction of anesthesia[4] and high-fidelity arterial line waveforms,[5] respectively. Previously, in the March 2018 issue of *Anesthesiology*, Lee *et al.* used machine learning to predict bispectral index values produced by target-controlled infusions of propofol and remifentanil.[6] An accompanying editorial provided a valuable summary of the history of artificial intelligence and an introduction to machine learning, the component of artificial intelligence that allows computers to make what humans describe as intelligent choices and predictions.[7] Although disagreement exists whether artificial intelligence, as driven by machine learning algorithms, portends an optimistic or ominous future, it is indisputable that machine learning paradigms have gained widespread traction in every industry.

Within the works featured in this issue, a rich underlying digital health dataset enabled the authors to leverage properties of machine learning to study old problems in new ways. These machine learning properties include an ability to capture numerous variables, better known as machine learning model features, which would otherwise elude human abilities to perceive or simultaneously consider (as is the case for the 2.6 million arterial waveform combinatorial features described by Hatib *et al.*). These also include the ability of machine learning to model complex relationships between model features which otherwise eclipse human understanding (as is the case for the deep neural network model described by Lee *et al.*).

Although some "transparent" machine learning methods provide insight into associations discovered, machine learning predictive models by nature do not require human comprehension in order to work. An ensuing challenge for scientific progress over the next decade will be to create and enforce standards for evaluating these methods, so as not to supersede the ability of authors to explain, or readers to understand. Concurrent with the rise of Big Data has been a rise in the inconsistency and uncertainty of applying machine learning concepts to datasets. If not kept in check, spurious conclusions drawn from methodologically unsound studies threaten the credibility of this science. Answering this call to action, and importantly recognized by all three featured articles, are a set of multidisciplinary guidelines for developing and reporting machine learning predictive models in biomedical research – well worth the read.[8]

Beyond a dire need for reporting standards in machine learning predictive models, it is of equal burden for practitioners to have a basic literacy of machine learning concepts in order to appraise machine learning-based investigations, much in the same way current biomedical literature demands a basic literacy of classical statistics and study design. These machine learning concepts include the use of training, testing, and validation datasets – used respectively to develop, assess internal performance, and externally validate machine learning algorithms (Figure 1). Additionally, just as clinicians are familiar with conventional statistical analyses such as logistic regression (which consequently, happens to be one simple type of algorithm supported by machine learning), it may behoove the perioperative clinician to be familiar with other machine learning techniques, including naïve Bayes, support vector machines, and random forests – to name a few; others are highlighted by Kendale *et al.* in this issue.

As demonstrated by the studies in this issue, the principal advantage of machine learning is the boost in performance it achieves when attempting to predict an observed outcome for which the range of explanatory features is large, or the depth of interactions between features is overwhelmingly complex. To predict hypotension, Hatib *et al.* brilliantly tap into vast arrays of data within the arterial line waveform, extending far beyond simple characteristics such as heart rate and blood pressure (and furthermore, far beyond "complex" characteristics such as pulse pressure variation, systolic pressure rise [dP/dt], and waveform area). When posed with an analytic task in which potential predictive features are in the thousands or millions or of nuanced complexity, the flexibility of machine learning techniques to accommodate inputs simply outmatch any traditional analytic method. In biomedical literature, other fields leveraging machine learning to tackle complex tasks include image processing (e.g. computer vision) of radiographic[9] or whole-slide pathology[10] images, as well as text analysis (natural language processing) of clinical notes.[11,12]

In contrast, for predictive analytic tasks in which features remain countable, or relationships explainable, machine learning may still prove useful, but will likely be of more modest benefit. In the work by Kendale *et al.*, an ensemble of machine learning methods indeed outperformed a classic logistic regression approach for predicting hypotension, but the overall performance of the machine learning model remained far from perfect. In the case of the best-performing algorithm (gradient boosting machines), Kendale *et al.* demonstrate a relatively small improvement compared to a classic logistic regression approach. Similarly, whereas Lee *et al.* successfully demonstrate a deep learning approach to predicting postoperative mortality from intraoperative data, the authors fail to demonstrate improvement compared to logistic regression, a recurring issue in studies promoting the use of deep learning.

As with all methodological approaches, machine learning is not without drawbacks. The most hotly contested is the difficulty of understanding mechanisms driving the prediction models presented. Herein lies the "black magic" of machine learning: although the predictive performance of a machine learning algorithm can be precisely quantified – and sometimes, this performance is staggering – the question of how to interpret and act upon the information generated remains wholly unanswered. In cases where mechanisms are of limited concern, or penalties for incorrect predictions low – such as facial recognition in family photos – machine learning techniques deftly succeed in their purpose. Conversely, in cases where mechanisms are critical, and penalties for error are high – as is often the case in healthcare, and particularly in anesthesiology – a machine learning approach falling anywhere short of nearly perfect remains unviable. Hatib *et al.* importantly note that although prediction of hypotension can be established with high fidelity, it remains entirely unclear as to how a clinician should respond to such an alert. This issue is even more critical, considering the generalizability and reproducibility concerns of such models. In many studies leveraging machine learning, insufficient testing and validation of complex models – particularly those using deep learning – can lead to overfitting of even the largest of datasets.

Despite such limitations, the work in this issue takes courageous shifts in methodologic approaches, and unmistakably establishes that machine learning applications to anesthesiology are not just a fad. The authors should be commended as exemplars for

assertively applying new scientific paradigms to our field. How such machine learning techniques are harnessed in order to improve anesthesia, and more broadly advance health sciences, remains a challenge for decades to come.

Acknowledgments

References

1. Simonite T. [Accessed July 5, 2018] Facebook creates software that matches faces almost as well as you do. Available at: https://www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do

2. Matacic C. [Accessed July 5, 2018] Google's new translation software is powered by brainlike artificial intelligence. Available at: http://www.sciencemag.org/news/2016/09/google-s-new-translation-software-powered-brainlike-artificial-intelligence

3. Lee CK, Hofer I, Gabel E, Baldi P, Cannesson M. Development and Validation of a Deep Neural Network Model for Prediction of Postoperative In-hospital Mortality. Anesthesiology. 2018

4. Kendale S, Kulkarni P, Rosenberg A, Wang J. Supervised Machine Learning Predictive Analytics for Prediction of Post Induction Hypotension. Anesthesiology. 2018

5. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, Rinehart J, Cannesson M. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology. 2018

6. Lee HC, Ryu HG, Chung EJ, Jung CW. Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanil: A Deep Learning Approach. Anesthesiology. 2018; 128:492–501. [PubMed: 28953500]

7. Gambus P, Shafer SL. Artificial Intelligence for Everyone. Anesthesiology. 2018; 128:431–433. [PubMed: 29166324]

8. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res. 2016; 18:e323. [PubMed: 27986644]

9. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017; 284:574–582. [PubMed: 28436741]

10. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak J, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Hass C, Bruni E, Wong Q, Halici U, Oner MU, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuscheit J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvuori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Venancio R. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017; 318:2199–2210. [PubMed: 29234806]

11. Buchan K, Filannino M, Uzuner O. Automatic prediction of coronary artery disease from clinical narratives. J Biomed Inform. 2017; 72:23–32. [PubMed: 28663072]

12. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M, Sundberg P, Yee H, Zhang K, Zhang Y, Flores G, Duggan GE, Irvine J, Le Q, Litsch K, Mossin A,

Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, Butte AJ, Howell MD, Cui C, Corrado GS, Dean J. Scalable and accurate deep learning with electronic health records. npj Digital Medicine. 2018; 1:18.
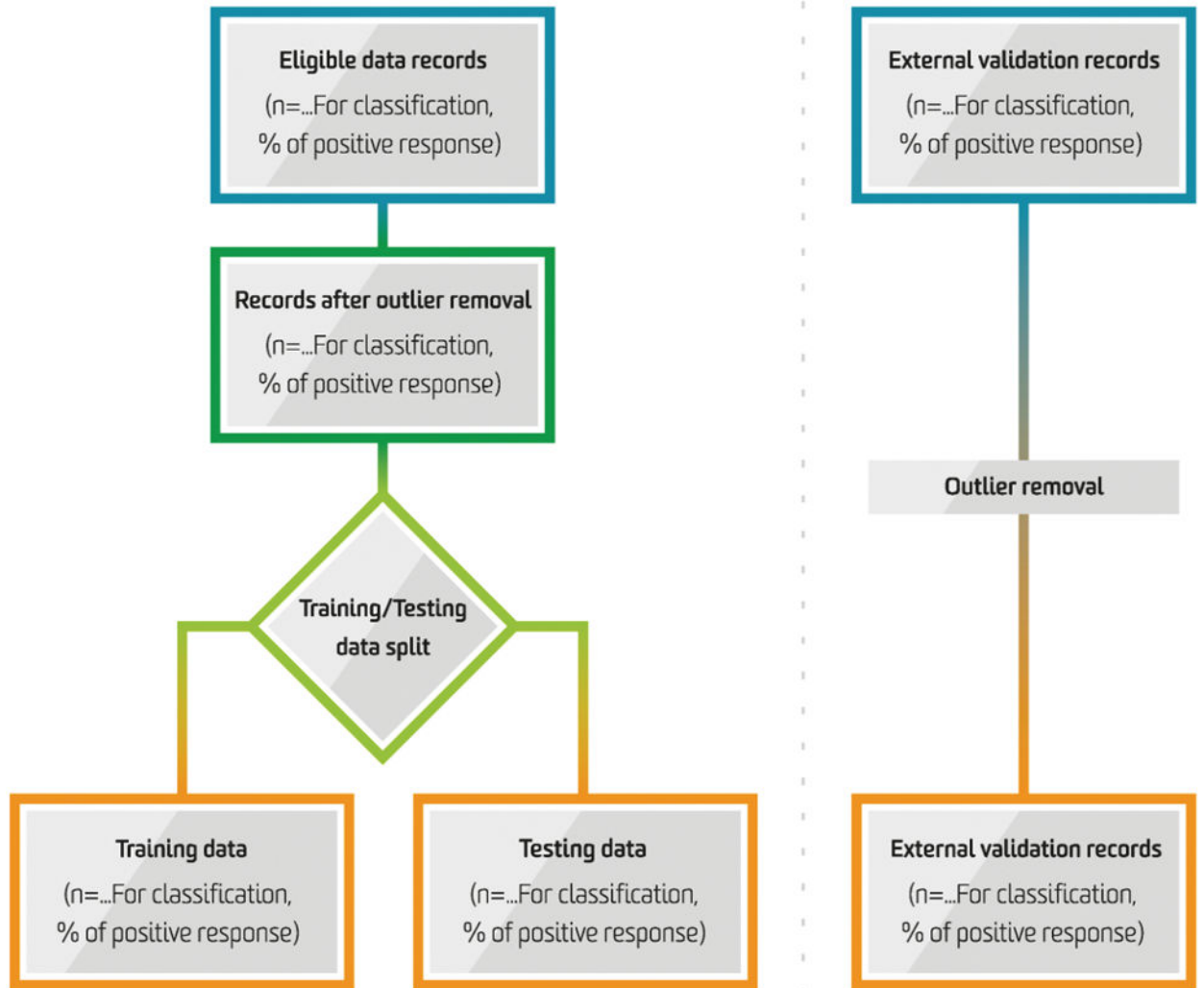
Figure 1.
Information Flow in the Predictive Modelling Process for Machine Learning.
Adapted from Luo W, Phung D, Tran T, Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *Journal of medical Internet research.* 2016;18(12):e323.