

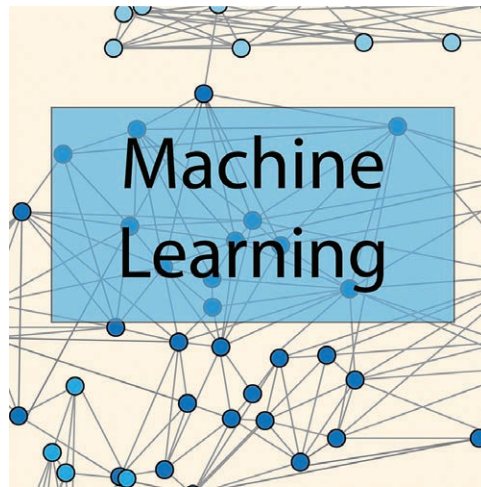
Artificial Intelligence for Anesthesia: What the Practicing Clinician Needs to Know

More than Black Magic for the Art of the Dark

Michael R. Mathis, M.D., Sachin Kheterpal, M.D., M.B.A., Kayvan Najarian, Ph.D.

MACHINE learning, the quintessential tool currently driving forward the development of artificial intelligence, was discovered and developed decades ago. Nevertheless, it is only recently that machine learning has seen an exponential increase in growth, sophistication, and influence. Recent success stories outside of health care are numerous, including Facebook's DeepFace, unveiled in 2014, which is 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 that is fed massive amounts of data to effectively train itself to recognize patterns in speech, with translation errors reduced by 87%.²

Machine-learning techniques like these may be coming soon to an operating room near you; in this issue are three examples of machine learning applied to our field. These include articles by Lee *et al.*,³ where machine-learning techniques are used to predict postoperative mortality from electronic health record data, as well as articles by Kendale *et al.*⁴ and Hatib *et al.*,⁵ in which hypotension is predicted *via* machine-learning algorithms that leverage data available during induction of anesthesia and high-fidelity arterial line waveforms, respectively. Previously, in the March 2018 issue of *ANESTHESIOLOGY*, Lee *et al.*⁶ used machine learning to predict bispectral index values produced by target-controlled infusions of propofol and remifentanyl. An accompanying editorial provided a valuable summary of the history of artificial intelligence and



“...machine-learning properties include an ability to capture... variables...[called] model features... [too numerous for humans] to perceive or simultaneously consider...[and] the ability...to model complex relationships between model features which otherwise eclipse human understanding.”

by Hatib *et al.*⁵). These also include the ability of machine learning to model complex relationships between model features that otherwise eclipse human understanding (as is the case for the deep neural network model described by Lee *et al.*³).

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

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 as to 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 articles 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

Image: J. P. Rathmell.

Corresponding articles on pages 649, 663, and 675.

Accepted for publication July 9, 2018. From the Departments of Anesthesiology (M.R.M., S.K.); Computational Medicine and Bioinformatics (K.N.); and Emergency Medicine (K.N.), University of Michigan Health System, Ann Arbor, Michigan.

Copyright © 2018, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. *Anesthesiology* 2018; 129:619-22

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 to develop, assess internal performance, and externally validate machine-learning algorithms, respectively (fig. 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.*⁴ 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.*⁵ 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

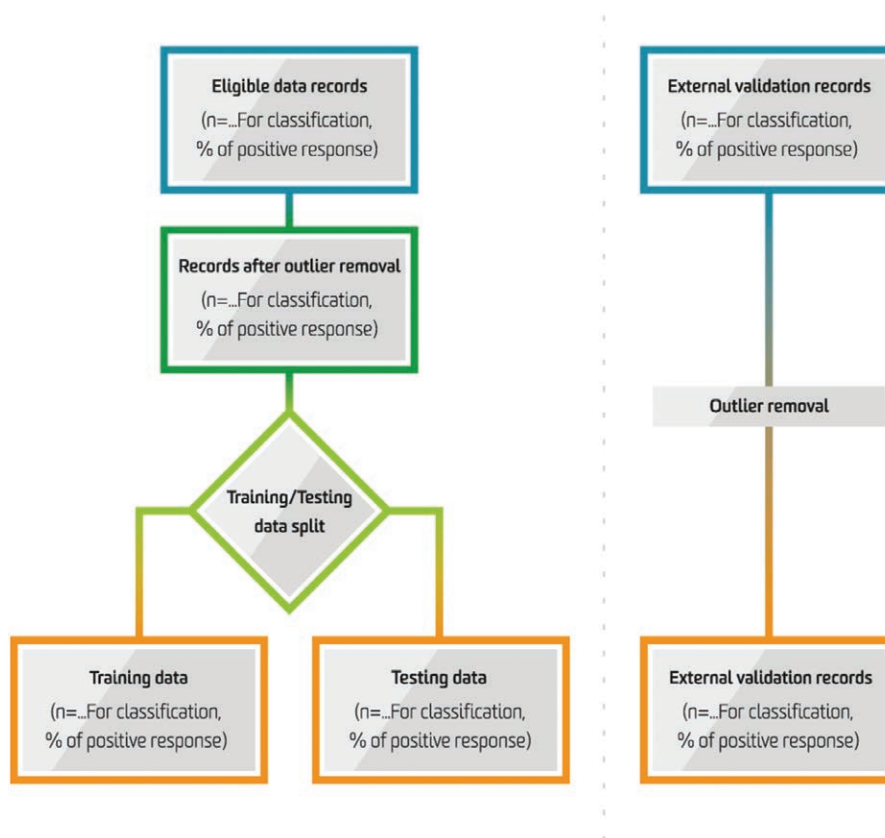


Fig. 1. Information flow in the predictive modeling process for machine learning. Adapted from Luo W, Phung D, Tran T, Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M: Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res* 2016; 18:e323. Available at: <http://www.jmir.org/2016/12/e323>. ©Wei Luo, Dinh Phung, Truyen Tran, Sunil Gupta, Santu Rana, Chandan Karmakar, Alistair Shilton, John Yearwood, Nevenka Dimitrova, Tu Bao Ho, Svetha Venkatesh, Michael Berk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

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 radiographic⁹ or whole-slide pathology¹⁰ 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 article by Kendale *et al.*,⁴ 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.*⁴ demonstrate a relatively small improvement compared to a classic logistic regression approach. Similarly, whereas Lee *et al.*³ 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 methodologic 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 health care, and particularly in anesthesiology—a machine-learning approach falling anywhere short of nearly perfect remains unviable. Hatib *et al.*⁵ 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 articles in this issue^{3–5} take courageous shifts in methodologic approaches, and unmistakably establish 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.

Research Support

All work and partial funding attributed to the Department of Anesthesiology, University of Michigan Medical School (Ann Arbor, Michigan). This editorial review described was supported in part by the National Heart, Lung, and Blood Institute of the National Institutes of Health (Bethesda, Maryland), grant No. 1K01HL141701-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests

Dr. Najarian has developed algorithmic technologies for prediction of cardiac events that have been licensed to Trove Analytics, Inc. (Silicon Valley, California), and New Vital Signs, Inc. (Ann Arbor, Michigan). The other authors are not supported by, nor maintain any financial interest in, any commercial activity that may be associated with the topic of this article.

Correspondence

Address correspondence to Dr. Mathis: mathism@med.umich.edu

References

1. Simonite T: 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>. Accessed July 5, 2018
2. Maticic C: 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>. Accessed July 5, 2018
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; 129:649–62
4. Kendale S, Kulkarni P, Rosenberg AD, Wang J: Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *ANESTHESIOLOGY* 2018; 129:675–88
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; 129:663–74
6. Lee HC, Ryu HG, Chung EJ, Jung CW: Prediction of bispectral index during target-controlled infusion of propofol and remifentanyl: A deep learning approach. *ANESTHESIOLOGY* 2018; 128:492–501
7. Gambus P, Shafer SL: Artificial intelligence for everyone. *ANESTHESIOLOGY* 2018; 128:431–3
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
9. Lakhani P, Sundaram B: Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284:574–82
10. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak

JAWM, 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, Öner MÜ, 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, Ruusuvoori 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, Venâncio R; the CAMELYON16 Consortium: Diagnostic assessment of deep learning algorithms for

detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318:2199–210

11. Buchan K, Filannino M, Uzuner Ö: Automatic prediction of coronary artery disease from clinical narratives. *J Biomed Inform* 2017; 72:23–32
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, Volchenbom 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

ANESTHESIOLOGY REFLECTIONS FROM THE WOOD LIBRARY-MUSEUM

From Horse to Hearse: Tragedy Strikes Laughing-gas Dentist Charles Henry Neall



From the Wood Library-Museum's Ben Z. Swanson Collection, this trade card's red rose conceals the drawn bow of Cupid. The young god or at least his bow seems bent on drawing Philadelphians to the Vine Street dental office of Dr. Charles Henry Neall (1834 to 1890), where laughing gas was administered for anesthesia. Unfortunately, Cupid could not rescue the portly Dr. Neall from dying while vacationing on tour. After his carriage had broken an axle and his horse had bolted, Dr. Neall had fallen, "striking on his head." According to the *Altoona Times*, the dentist's massive girth "made his fall all the more severe" and doomed him to a terminal "concussion of the brain." (Copyright © the American Society of Anesthesiologists' Wood Library-Museum of Anesthesiology.)

George S. Bause, M.D., M.P.H., Honorary Curator and Laureate of the History of Anesthesia, Wood Library-Museum of Anesthesiology, Schaumburg, Illinois, and Clinical Associate Professor, Case Western Reserve University, Cleveland, Ohio. UJYC@aol.com.