



# HHS Public Access

Author manuscript

*Am J Psychiatry*. Author manuscript; available in PMC 2018 February 09.

Published in final edited form as:

*Am J Psychiatry*. 2017 February 01; 174(2): 93–94. doi:10.1176/appi.ajp.2016.16101169.

## Machine learning and electronic health records: A paradigm shift

Daniel E. Adkins, PhD

University of Utah

In the accompanying article, Barack-Corren et al. use machine learning (ML) methods to build a highly predictive model of suicidal behavior using longitudinal electronic health records (EHRs). They do so using a well-established, probability-based ML algorithm, the Naïve Bayes Classifier (NBC), to mine through ~1.7 million patient records, spanning 15 years (1998–2012), from two large Boston hospitals. After training the NBC model on a randomly selected half of the data, the predictive ability of the model was assessed on the second half, yielding accurate (35%–49% sensitivity at 90%–95% specificity) and, critically, early (3 – 4 years in advance on average) prediction of patients’ future suicidal behavior. In this, the authors benefitted from access to a large and high-quality EHR database and choose an appropriate, and powerful, analytical method in NBC. Further, the research has clear clinical applications in the potential for early detection warnings via physician EHR notices. Beyond such specifics, the article has broader significance in its demonstration of how the atheoretical ML approaches popular in Silicon Valley can successfully mine clinical insights from an exponentially growing body of EHR data. It also hints toward a future in which ML of big medical data may become a ubiquitous component of clinical research and practice—a prospect that some are uncomfortable with.

While the pace at which ML applications diffuse into clinical research and practice remains to be seen, methodological development in the ML field continues to accelerate rapidly. And this suggests one primary limitation of the current study. That is, while the NBC is well-suited to the current application, it is an older and remarkably simple method by ML standards. Fundamentally, the NBC is a direct application of Bayes Theorem, simply calculating the product of the prior probability of the outcome of interest (e.g., suicidal behavior) and the probabilities for each predictor in the data conditional on the outcome of interest (Friedman, Geiger, & Goldszmidt, 1997). This analytical simplicity contrasts sharply with more advanced ML techniques including neural nets, deep learning, and ensemble methods, which achieve notable increases in prediction relative to NBC, but are black boxes in terms of estimation, as their models are extremely large, complex, and characterized by “hidden layers” (LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015). So, while there is ample room for improved prediction accuracy in Barack-Corren et al.’s approach, such gains would likely come at the expense of interpretability and inference. Thus, their selection of the NBC method has the further, unintended, merit of providing an unusually lucid, accessible introduction to ML for many researchers and clinicians.

Another limitation, perhaps strategic on Barack-Corren et al.'s part, is the use of a limited set of standard ICD-9 codes and search terms as predictors, versus performing natural language processing (NLP) of the full semi-structured data of the EHR. This analytical decision is a significant limitation as it drastically reduces the analysis feature space (i.e., the number of predictors considered), which generally results in poorer prediction given data of this size (Jurafsky & Martin, 2008; Lin & Dyer, 2010). While the authors do not give a precise number of predictors used in their analysis, we can safely assume it is at least an order of magnitude less than what would be possible using NLP techniques. However, this again raises the issue of model interpretability, as NLP approaches may identify highly predictive features that offer no clear interpretation or clinical significance (Lin & Dyer, 2010). Contrast that opacity with Barack-Corren et al.'s list of the top 100 predictors in their NBC (Supp Mat, Table S-2), which summarizes a wealth of clinical insight, and we again see the precision advantages of more sophisticated approaches counterbalanced by the interpretability of simpler models like Barack-Corren et al.'s NBC. This tradeoff is not specific to the current topic, instead it is a pervasive aspect of ML—a continuum of inference versus prediction that is traversed when moving from simpler approaches, like Barack-Corren et al.'s NBC, to more advanced, opaque approaches including neural nets and deep learning (Breiman, 2001; Kelleher, Namee, & D'Arcy, 2015).

Stepping back from the technical aspects of ML, this study provides an opportunity to reflect on the trend of the field toward increasingly data-driven approaches. Regardless of the promise of ML of EHR, it would be unwise to endorse the approach without first considering the various professional, ethical, and legal issues accompanying the potential improvements in diagnosis and treatment. From the perspective of praxis, it is noteworthy that the approach, carried to its logical conclusion, is fundamentally atheoretical, which marks a stark departure from conventional clinical paradigms built primarily on evidence-based causal models (Greenland, Pearl, & Robins, 1999). Further, for some it may seem like a slippery slope toward ceding power in the clinic to algorithms, and devaluing clinician experience and judgment. But I would note that the majority of a clinician's function would not, and indeed could not, be encroached upon by data-driven analytics. Rather, increasing the role of ML applications to EHRs would just provide additional inputs for the clinician to consider in making diagnostic and treatment decisions. In this way, the emergence of ML EHR prediction may be seen as analogous to the development of imaging, genetic, or any other new source of highly informative medical data. Additionally, there are ethical and legal issues surrounding the mining of EHR, including protecting the patient population from adverse consequences stemming from the analysis of their data. This suggests potentially problematic dynamics if, for instance, EHR data and analytics are accessed by insurance companies, who may use the data to discriminate against patients in the marketplace. This risk is compounded by the possibility of blackbox ML methods inadvertently identifying stratifying criteria that we as a society find unacceptable.

While ethical arguments for the use of participant data often take the form of efforts to limit access to data, as in the well-justified attention paid to patient privacy and nondisclosure, a powerful argument for the opposite exists in regard to enhancing public benefit through the analysis of EHR data. That is, as the data are often collected using some combination of patient permission and government funding, it may be reasonable to consider public benefit

as a goal, or even an obligation, in the collection and analysis of the data. Although this does not argue against private sector activity, it does support a concerted effort to consolidate data and analyses funded by federal research dollars into a public resource—and what a tremendous resource a centralized archive of EHR data staffed with a cadre of ML analysts could be. Currently, this possibility is prevented by data fragmentation as most EHR data is currently proprietary (Hall, 2010; Jensen, Jensen, & Brunak, 2012), but this could change with leadership from federal entities. And we have good precedent from the NIH and VA regarding safeguarding, and maximizing benefit from, comparable archives (e.g., dbGaP).

In sum, as demonstrated by Barack-Corren and colleagues, the application of ML methods to EHRs, and the potential of extending such analyses to other sources of big medical data (e.g., genomics and imaging), could generate enormous – yes, even paradigm-shifting – returns in improved diagnosis and treatment. What remains unclear is the pace at which these benefits will be realized, as well as who the primary beneficiaries will be.

## Acknowledgments

This research was supported by the National Institute of Mental Health (K01MH093731) and the University of Utah, Consortium for Families and Health Research.

## References

- Breiman L. Statistical Modeling: The Two Cultures. *Statistical Science*. 2001; :199–231. DOI: 10.1214/ss/1009213726
- Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine Learning*. 1997; 29(2–3):131–163. DOI: 10.1023/a:1007465528199
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;37–48. [PubMed: 9888278]
- Hall MA. Property, Privacy and the Pursuit of Integrated Electronic Medical Records. *Iowa Law Review*. 2010; 2010:631–663.
- Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012; 13(6):395–405. [PubMed: 22549152]
- Jurafsky, D., Martin, JH. *Speech and language processing*. 2. Prentice Hall; 2008.
- Kelleher, JD., Namee, BM., D’Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics*. The MIT Press; 2015.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553):436–444. DOI: 10.1038/nature14539 [PubMed: 26017442]
- Lin, J., Dyer, C. *Data-Intensive Text Processing with MapReduce*. Morgan and Claypool Publishers; 2010.
- Schmidhuber J. Deep learning in neural networks: An overview. *Neural networks*. 2015; 61:85–117. <http://dx.doi.org/10.1016/j.neunet.2014.09.003>. [PubMed: 25462637]