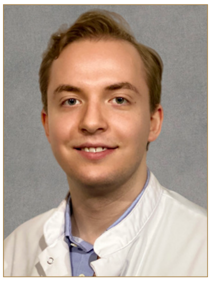
Neurospine 2019;16(4):654-656.
https://doi.org/10.14245/ns.1938434.217

Neurospine
pISSN 2586-6583 eISSN 2586-6591

Editorial

Data Mining in Spine Surgery: Leveraging Electronic Health Records for Machine Learning and Clinical Research



Victor E. Staartjes[1,2], Martin N. Stienen[1]

[1]Department of Neurosurgery, Clinical Neuroscience Center, University Hospital Zurich, Zurich, Switzerland
[2]Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Clinical Neuroscience Center, University of Zurich, Zurich, Switzerland

Corresponding Author
Victor E. Staartjes
E-mail: victoregon.staartjes@usz.ch
https://orcid.org/0000-0003-1039-2098

Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Department of Neurosurgery, Clinical Neuroscience Center, University Hospital Zurich, Zurich, Switzerland

See the article "Applications of Machine Learning Using Electronic Medical Records in Spine Surgery" via https://doi.org/10.14245/ns.1938386.193.

Advances in natural language processing (NLP) and unsupervised learning have recently enabled the long-expected synergy between true "big data" and analytics based on machine learning (ML). The ability to reliably generate structured data from unstructured electronic health records (EHRs) such as free text reports, document scans, or unlabelled medical imaging has on the one hand allowed development of algorithms based on until recently unseen amounts of data, spanning entire hospital or even national populations, and not just databases compiled by human experts. On the other hand, the capacity to primarily generate structured reports from unstructured raw EHRs has proven valuable for hospital analytics, epidemiological studies, and systematic reviews.

In their narrative review, Schwartz et al.[1] report on the utilization of EHRs in spine surgery through ML techniques. The authors are to be commended for their detailed description of data types commonly found in EHRs, learning concepts to generate structured data (such as NLP and machine vision), applications of ML for prognosis and prediction, and finally the challenges inherent to using unstructured data from EHRs in medical practice and research. As the authors show, there is no question that ML is already starting to affect surgical practice relevantly in many aspects. Especially the advent of open source algorithms provided by today's tech giants have largely democratized the development of ML models, and as the authors summarize, this has led to an explosion of publications reporting such algorithms. Still, it is important to preserve the methodological quality of papers utilizing ML techniques, which is often not the case. For example, the authors touch on the issue of ensuring generalizability through robust training structures (i.e., some form of resampling) and external validation, before models are rolled out into clinical practice.

We especially value that the authors discuss the problem of uninterpretable "black box" models.[2] Currently, many groups are applying complex ML algorithms to relatively small patient samples and for relatively simple tasks. While this might lead to slight benefits in model performance, these complex models (such as deep neural networks for nonimaging applications) are often typical "black box" models with a total loss of the ability to explain what factors lead the algorithm to make a certain decision. Explicability is – unfortunately – often traded in for a small and likely irrelevant increase in model accuracy. Especially in today's medicolegal arena, the ability to justify and explain a model's decisions will become more crucial, and therefore model selection should not only be guided by performance measures, but also by algorithm explicability.

Legal and ethical implications are also at the forefront of the discussion on ML in medicine. For example, it is as of yet unclear how federal agencies will regulate the development, distribution, and clinical application of ML algorithms. This is especially true for predictive analytics. The authors make one very interesting point on this topic: Algorithmic outcome prediction may put minorities, or other, specific patients with e.g. less favorably predicted outcome or abnormally high predicted resource utilization at disadvantage. Already now, in the early stages of implementing ML into clinical practice, we should make sure that selective treatment based on predictive analytics is used to identify the optimal healthcare plan for patients, but – on the contrary – that this power is not "misused" by insurance companies to withhold (costly) care for those in dire need of it.

In the context of structuring EHR data, NLP plays a primary role. While in the past, clinical research was based on data manually extracted by medical students, residents, and other medical personnel—often associated with massive inter- and intrarater disagreement[3]—in the future, larger amounts of data can be more accurately and reproducibly extracted using NLP. Aside from the many applications of NLP that the authors mention in their narrative review, Buchlak et al.[4] show that NLP can also help in neurosurgical research through automated classification of search results in systematic reviews. Furthermore, it has been estimated that 18% of cancer trials fail to recruit even half of the required the sample size or are shut down early because of poor recruitment.[5] Even in this situation, NLP has proven valuable in identifying patients from EHRs who are both eligible and likely to participate in a given trial, as e.g., the investigators of a large breast cancer trial did to increase recruitment speed.[6] In the field of spine care, Huhdanpaa et al.[7] show that NLP can be used to identify all patients with Modic type 1 endplate changes for clinical research. Lastly, it has to be mentioned that inclusion of population- or hospital-level big data potentially reduces the inequalities commonly seen in medical research: Around 90% of trial participants worldwide are white, and it has been suggested that NLP-powered patient matching could lead to more diverse trial cohorts, democratizing access to state-of-the-art clinical trials.[8]

The authors mention that, according to an IBM report,[9] 90% of medical big data consist of imaging files. While it may certainly be true that the majority of big data measured in terms of file size is imaging data, a large part of the radiological rating tasks to which ML have been applied with correct evaluation methodology have actually not demonstrated a significant performance increase compared to human experts, especially because often clinically irrelevant findings are produced.[10,11] In our opinion, it can currently be said that the most interesting applications of machine vision to medical imaging do not necessarily lie within diagnostics or other tasks that human experts can perform as well, but rather within tasks that humans cannot ordinarily perform. Examples for this are the extraction of radiomic features such as genomic alterations from magnetic resonance imaging (MRI),[12] conversion of musculoskeletal MRI to computed tomography,[13] or to reduce the amount of Gadolinium contrast agent for MRI scans.[14] Furthermore, machine vision can help to prevent wrong-level spine surgery.[15] These innovative tasks are often those at which algorithms can demonstrably outperform human experts, and we expect the clinically relevant novelties to be within this realm. We also appreciate the author's attempts to also explain the terms artificial intelligence (AI) and ML, and would like to add that generally it can be said that when the term AI is used, learning methods that acquire general capabilities are meant, whereas ML refers to learning techniques for very specific tasks (such as prediction of proximal junctional kyphosis after spinal fusion). We thus recommend to refrain from the term "AI," which is often very liberally used, when actually speaking about ML, under which most published applications of learning techniques in spine surgery currently would fall, taking the definitions strictly. Also, the article is not a systematic review, and as such, a number of contributions to the literature with relevance to the current discussion may not have been included. Nonetheless, the overview provided by Schwartz et al.[1] enable a thorough examination of the current trends in ML applications to EHR utilization.

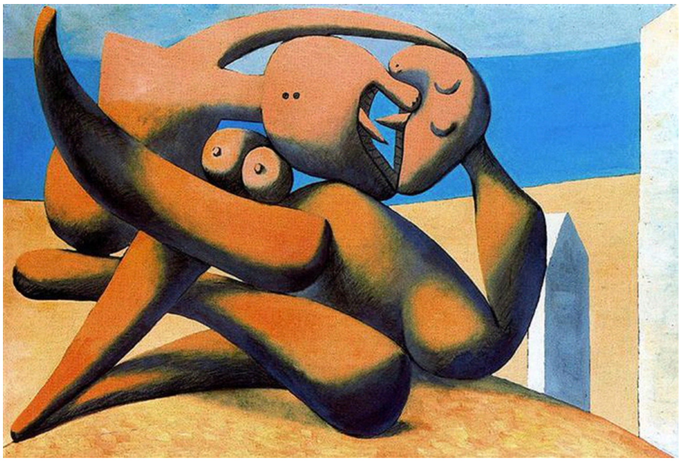
In the near future, we expect 2 main rationales for the use of ML in conjunction with EHRs: First, to automatically leverage structured big datasets from unstructured raw EHRs using learning techniques such as NLP, which then allows for adequate training of diagnostic, prognostic, or predictive ML algorithms. Here, generation of these ML algorithms—which require structured data—from unstructured EHRs is the primary goal. And second, to generate structured reports from unstructured raw EHRs for evaluation, research, and assistance in clinical practice. Both applications would be impossible without structuring previously unstructured data, and both are exciting and encouraging examples of how new technologies can be translated from pure in-silico research applications to real-world benefits in daily clinical practice.

CONFLICT OF INTEREST

The authors have nothing to disclose.

REFERENCES

1. Schwartz JT, Gao M, Geng EA, et al. Applications of machine learning using electronic medical records in spine surgery. Neurospine 2019;16:643-53.
2. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206-15.
3. Mooney MA, Hardesty DA, Sheehy JP, et al. Interrater and intrarater reliability of the Knosp scale for pituitary adenoma grading. J Neurosurg 2017;126:1714-9.
4. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev 2019 Aug 17 [Epub]. doi: 10.1007/s10143-019-01163-8.
5. Bennette CS, Ramsey SD, McDermott CL, et al. Predicting low accrual in the National Cancer Institute's Cooperative Group Clinical Trials. J Natl Cancer Inst 2015 Dec 29;108(2). pii: djv324. https://doi.org/10.1093/jnci/djv324.
6. Hughes KS, Schnaper LA, Bellon JR, et al. Lumpectomy plus tamoxifen with or without irradiation in women age 70 years or older with early breast cancer: long-term follow-up of CALGB 9343. J Clin Oncol 2013;31:2382-7.
7. Huhdanpaa HT, Tan WK, Rundell SD, et al. Using natural language processing of free-text radiology reports to identify type 1 modic endplate changes. J Digit Imaging 2018;31:84-90.
8. Knepper TC, McLeod HL. When will clinical trials finally reflect diversity? Nature 2018;557:157-9.
9. Papp L, Spielvogel CP, Rausch I, et al. Personalizing medicine through hybrid imaging and medical big data analysis. Front Phys 2018;6:Article 51. https://doi.org/10.3389/fphy.2018.00051
10. Senders JT, Arnaout O, Karhade AV, et al. Natural and artificial intelligence in neurosurgery: a systematic review. Neurosurgery 2018;83:181-92.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
12. Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. Clin Cancer Res 2018;24:1073-81.
13. Florkow MC, Zijlstra F, Willemsen K, et al. Deep learning-based MR-to-CT synthesis: The influence of varying gradient echo-based MR images as input channels. Magn Reson Med 2019 Oct 8 [Epub]. https://doi.org/10.1002/mrm.28008.
14. Gong E, Pauly JM, Wintermark M, et al. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 2018;48:330-40.
15. Zagzoog N, Yang VX. Novel extended vertebral registration for wrong level spinal surgery (NEVER Wrong) [abstract]. In: Congress of Neurological Surgeons 2018 Annual Meeting; 2018 Oct 6-10. Houston (TX), USA. Available from: https://www.cns.org/annual-meeting-2018/awards.



Title: Figures at the seaside
Artist: Pablo Piccaso
Year: 1931
A series of bizarre erotic beach scenes, including The Kiss, was painted in the summer of 1931 at Picasso's French Riviera vacation resort, Juan-les-Pins. Said to be inspired by the 50-year-old painter's liaison with 19-year-old model, Marie-Therese Walter, the grotesque nature of the depicted forms reduces this moment of intimate contact to a level of crudity, probably more representative of his deteriorating relationship with his wife, Olga. The praying mantis-like head of the two figures was a popular image with the Surrealists because the perverse concept of the female insect eating her mate after intercourse provided another visual metaphor of the 'life and death' paradox. Here, the heads incorporate Picasso's obscene vagina dentatta teeth imagery, as well as penile tongues
More information: https://www.pablopicasso.org/figures-at-the-seaside.jsp