Check for updates

Editorial: Machine Learning in Action: Stroke Diagnosis and Outcome Prediction

Vida Abedi [1]*, Yuki Kawamura [2], Jiang Li [3], Thanh G. Phan[4,5] and Ramin Zand [6]

[1]Department of Public Health Sciences, College of Medicine, The Pennsylvania State University, Hershey, PA, United States, [2]Department of Medicine, University of Cambridge, Cambridge, United Kingdom, [3]Department of Molecular and Functional Genomics, Weis Center for Research, Geisinger Health System, Danville, PA, United States, [4]Stroke and Aging Research Group, Clinical Trials, Imaging and Informatics Division, School of Clinical Sciences at Monash Health, Melbourne, VIC, Australia, [5]Department of Neurology, Monash Health, Melbourne, VIC, Australia, [6]Department of Neurology, College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Editorial on the Research Topic
Machine Learning in Action: Stroke Diagnosis and Outcome Prediction

Machine learning—the ability of computers to "learn" to perform a task rather than being explicitly programmed for the purpose—has seen significant developments in recent years. Biomedical research is no exception to its far-reaching impact and has seen more than a ten-fold increase in the number of publications related to machine learning in the last decade (1). In this Research Topic, we present recent advances in developing machine learning algorithms in the context of cerebrovascular diseases to highlight promising approaches that represent various areas of potential clinical utility in stroke care. The focus is on applications with high clinical value and a solid technical foundation.

Deployment of machine learning algorithms in the clinic principally involves four stages of the care workflow: primary prevention, acute-phase treatment, post-diagnosis prediction, and secondary prevention (2). Primary prevention includes personalized or stratified patient risk prediction and identification of gaps in care, whereas integration into acute phase treatment aims to aid physician diagnosis and referrals. Machine learning algorithms for post-diagnosis and secondary prediction can provide predicted outcomes that allow the identification of patients who would be responsive to treatment or require careful monitoring due to a higher risk of recurrent disease. Together, machine learning algorithms can aid clinical decision-making in each step by providing recommendations and pointing to possible missed cases for critical conditions. As suggested by Mainali et al., machine learning algorithms can have particular utility in alleviating two of the clinical challenges of stroke: the time-sensitive nature of the acute-phase treatment and the difficulty of predicting outcomes, especially in

the acute phase. Given these potential benefits, calibrating the algorithms to prevent excessive alerts and supporting physician autonomy through careful assessment of human-computer interaction is key to maximizing adoption (3).

Electronic health records (EHR) are one of the principal sources of standardized clinical information on a patient and can serve as a valuable starting point for algorithm development. The results of Rana et al. are encouraging, demonstrating that models trained on EHR data outperformed models trained on a limited number of features clinically associated with stroke, confirming the benefits of additional information obtained by data extraction from EHR. Using EHR data, Darabi et al. compared the performance of multiple machine learning models in predicting 30-day hospital readmission. Their models improved upon previous predictive models based on logistic regression and provided promising results that could direct targeted intervention for high-risk patients. Notably, features that their best predictive model indicated as being key predictors of 30-day readmission agree with results from independent studies (4, 5) and clinical intuition, underscoring the interpretability of their model.

Complementation of EHR data with additional modalities of clinical investigations holds promise in further improving prediction accuracy. Herein, Lineback et al. employ Natural Language Processing (NLP) to glean freeform textual data. In contrast, Rajashekar et al. combines MRI and CT imaging data to improve prediction models trained solely on EHR data. Multimodal approaches can require more sophisticated models to extract information from various data types but more closely approximate decision-making by physicians and better integrate multifaceted information collected *via* clinical investigations and examinations.

Imaging is a rich source of information. Imaging has critical clinical relevance in neurology and a high affinity for sophisticated deep learning models, such as convolutional neural networks. Indeed, many of the recent advances in machine learning in healthcare have centered on image analysis, including the use of retinal images for cardiometabolic disease prediction (6–9) and analysis of histopathological slides (10–15). Models focus on cerebrovascular disease, however, have been comparatively scant. McLouth et al. validate the performance of a commercially available deep learning software in assessing intracranial hemorrhage and large vessel occlusion using CT images. Implementing analysis software within the imaging workflow can provide venues where machine learning algorithms can seamlessly integrate into clinical decision-making. Furthermore, incorporating features from MRI scans, such as in the study by Xiao et al. predicting hypoperfusion in ischemic stroke patients, could define a clinically relevant threshold that directs decision-making in a facile manner. Integration of images in machine learning algorithms provides several benefits, including higher accuracy of diagnosis and improved objectivity compared to physical examinations. Given

that imaging is routinely performed for stroke patients and is uniquely capable of providing functionally relevant anatomical information, image analysis models are promising candidates for clinical deployment in stroke care.

Machine learning can be an invaluable asset, especially in cases where diagnosis requires extensive examination or training or when the diagnosis is based on subtle features and are thus inherently prone to misdiagnosis. The algorithms described by Kim et al. to identify acute central dizziness and by Lin et al. to identify mild stroke patients at risk of disability exemplify the possibilities—supporting physicians in making challenging clinical decisions. Both models closely approximate or outperform existing risk scores without requiring extensive neurological examinations, allowing more patients to be screened and thus reducing the chances of a deteriorating patient escaping notice.

While machine learning holds promises, several challenges persist in implementing these technologies in healthcare. First, technical limitations can stem from the type and quality of the datasets available. EHR data can often be poorly standardized and sparse, posing problems in model generalizability. Investigators such as Rana et al. and Darabi et al. have only used administrative data from EHR with additional clinical variables such as NIHSS. By contrast, mining the free text in the patient chart (such as provider note, triage notes, discharge note, etc.) pose significant challenges. The free text is written by multiple clinicians, often with successive clinicians copying and pasting the written comments by the previous clinicians (16) in addition to auto-generated text that populate the patient chart. In addition, the tabular nature of clinical data extracted from EHR can often pose a difficulty even for advanced deep learning modalities, which often fail to surpass performances on simpler tree-based architectures (17). However, performance can be improved by extensive regularization (18). Sophisticated machine learning algorithms have had better success when applied to image datasets; however, even these complex deep learning algorithms can suffer from confounding factors, partially due to variation amongst institutions. Indeed, a recent study demonstrated that deep neural nets trained to predict SARS-CoV2 infection from X-ray images tend to select confounding "shortcuts" over signals in generating predictions (19). Attributes of datasets can limit the accuracy and generalizability of models, especially for external cohorts with different demographics and dataset characteristics. The development of standardized data protocols can aid the implementation of machine learning models that are more accurate and generalizable across multiple institutions. In addition to curating better datasets, models can also be adjusted for better generalizability; fine-tuning of pre-trained algorithms *via* transfer learning using site-specific data achieved superior results for external cohorts (20), and continuous domain adaptation has been explored to tackle temporal drifts in data (21, 22). It is essential to take all possible precautions

to ensure that the machine learning algorithms provide reliable, relevant, and interpretable results free from systemic biases. To achieve that, care must be taken to minimize confounding variations in the datasets that might affect generalizability and ensure fine-tuning approaches are integrated to allow the models to more closely approximate results for the underlying patient distribution.

Secondly, more complicated machine learning models can often be challenging to interpret, hindering the translation from prognosis to patient management. High-performing "black box" models lacking interpretability are of limited use in the clinic as they do little to inform physicians of actionable points. In particular, identifying modifiable risk factors is essential in the primary and secondary prevention of cerebrovascular events. To this end, Cui et al. used feature importance metrics to rank specific features mainly associated with predictive capability in each machine learning model. Analyses of feature importance could prove helpful in guiding intervention, especially if a factor is consistently listed as important across multiple models. For image analysis models, localization maps generated by methods such as Grad-CAM (23) could provide a limited level of interpretability. Separating interpretation from the prediction modeling to provide more flexibility is a strategy that has been getting more traction in recent years. Still, the usefulness of the algorithms can be diminished by confounding "shortcuts," as mentioned earlier. Since model depth is generally associated with better predictive capability, efforts must be made to create models that predict and inform. Desirable models should also consider workflow disruption or the possibility of causing "alert fatigue" before planning for implementation. Designing and training models so that interpretable features can be gleaned from model parameters and incorporating feedback from healthcare providers can improve the interpretability of models. In this respect, theoretical advances in model architecture and interpretation, combined with enhancing training data robustness, could prove fruitful.

Finally, ethical considerations must not be ignored. Model predictions can often be influenced by the socioeconomic, racial, and gender composition of the training datasets, the awareness of which is necessary to mitigate potential biases in models. For example, machine learning models were found to consistently underdiagnose patients in disadvantaged populations across three large chest X-ray datasets, especially where a patient was a member of more than one underserved group (24). The precedent of undertreatment in disadvantaged populations can further exacerbate biases by making it less likely for the algorithm to recommend treatment for members of the underprivileged sub-group of the population if similar patients were not provided treatment in the past. The performance of machine learning models must thus be thoroughly evaluated in different cohorts to assess the presence of systematic bias, which must be rectified before deployment. Further, while it is often possible to impute information that a patient declined to provide (e.g., smoking, HIV status, etc.), doing so can have ethical implications (25). Implementing machine learning algorithms in the clinic should proceed with special care to avoid unwittingly perpetuating health care inequalities in the training cohort. Finally, it is essential to reflect that algorithms are and will continue to be part of our medical system, including our medical education system. Thus, as a two-way street, we have to consider how such recommendations influence physicians' decisions and how this decision-making process potentially shifts with continued interaction.

In conclusion, recent developments in machine learning present ample opportunities for automated models that guide clinical decision-making and improve patient outcomes. The studies included herein represent selections of advances employing machine learning in various contexts in stroke care in our collective efforts to promote improved patient health through effective prevention, diagnosis, and intervention.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor J-CB declared a shared affiliation with the author YK at the time of review.

Publisher's note

References

1. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digital Med.* (2020) 3:126. doi: 10.1038/s41746-020-00333-z

2. Abedi V, Razavi S-M, Khan A, Avula V, Tompe A, Poursoroush A, et al. Artificial intelligence: a shifting paradigm in cardio-cerebrovascular medicine. *J Clin Med.* (2021) 10:5710. doi: 10.3390/jcm10235710

3. Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, et al. Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Ther Adv Neurol Disor.* (2020) 13:1756286420938962. doi: 10.1177/1756286420938962

4. Qiu X, Xue X, Xu R, Wang J, Zhang L, Zhang L, et al. Predictors, causes and outcome of 30-day readmission among acute ischemic stroke. *Neurol Res.* (2021) 43:9–14. doi: 10.1080/01616412.2020.1815954

5. Lichtman JH, Leifheit-Limson EC, Jones SB, Watanabe E, Bernheim SM, Phipps MS, et al. Predictors of hospital readmission after stroke. *Stroke.* (2010) 41:2525–33. doi: 10.1161/STROKEAHA.110.599159

6. Cheung CY, Xu D, Cheng C-Y, Sabanayagam C, Tham Y-C, Yu M, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng.* (2021) 5:498–508. doi: 10.1038/s41551-020-00626-4

7. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

8. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0

9. Wolf RM, Channa R, Abramoff MD, Lehmann HP. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol.* (2020) 138:1063–9. doi: 10.1001/jamaophthalmol.2020.3190

10. Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat Commun.* (2021) 12:1613. doi: 10.1038/s41467-021-21896-9

11. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* (2019) 25:1301–9. doi: 10.1038/s41591-019-0508-1

12. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer.* (2020) 1:800–10. doi: 10.1038/s43018-020-0085-8

13. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. *Nature.* (2020) 578:615–20. doi: 10.1038/s41586-019-1876-x

14. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y

15. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun.* (2020) 11:3877. doi: 10.1038/s41467-020-17678-4

16. Markel A. Copy and paste of electronic health records: a modern medical illness. *Am J Med.* (2010) 123:e9. doi: 10.1016/j.amjmed.2009.10.012

17. Shwartz-Ziv R, Armon A. *Tabular Data: Deep Learning is Not All You Need* (2021) [arXiv:2106.03253 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210603253S (accessed June 1, 2021).

18. Kadra A, Lindauer M, Hutter F, Grabocka J. *Well-tuned Simple Nets Excel on Tabular Datasets* (2021) [arXiv:2106.11189 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210611189K (accessed June 1, 2021).

19. DeGrave AJ, Janizek JD, Lee S-I. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nat Mach Intel.* (2021) 3:610–9. doi: 10.1038/s42256-021-00338-7

20. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Dig Med.* (2022) 5:69. doi: 10.1038/s41746-022-00614-9

21. Lao Q, Jiang X, Havaei M, Bengio Y. *Continuous Domain Adaptation with Variational Domain-Agnostic Feature Replay* (2020) [arXiv:2003.04382 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200304382L (accessed March 1, 2020).

22. Wang H, He H, Katabi D. *Continuously Indexed Domain Adaptation* (2020) [arXiv:2007.01807 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200701807W (accessed July 1, 2020).

23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization* (2016) [arXiv:1610.02391 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2016arXiv161002391S (accessed October 1, 2016).

24. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med.* (2021) 27:2176–82. doi: 10.1038/s41591-021-01595-0

25. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* (2019) 25:1337–40. doi: 10.1038/s41591-019-0548-6