Taylor & Francis
Taylor & Francis Group

Neurological Research

A Journal of Progress in Neurosurgery, Neurology and Neurosciences

ISSN: 0161-6412 (Print) 1743-1328 (Online) Journal homepage: https://www.tandfonline.com/loi/yner20

Predictive analytics and machine learning in stroke and neurovascular medicine

Hamidreza Saber, Melek Somai, Gary B. Rajah, Fabien Scalzo & David S. Liebeskind

To cite this article: Hamidreza Saber, Melek Somai, Gary B. Rajah, Fabien Scalzo & David S. Liebeskind (2019): Predictive analytics and machine learning in stroke and neurovascular medicine, Neurological Research, DOI: 10.1080/01616412.2019.1609159

To link to this article: https://doi.org/10.1080/01616412.2019.1609159

Published online: 30 Apr 2019.

NEUROLOGICAL RESEARCH
https://doi.org/10.1080/01616412.2019.1609159

Taylor & Francis
Taylor & Francis Group

Predictive analytics and machine learning in stroke and neurovascular medicine

Hamidreza Saber[a], Melek Somai[b], Gary B. Rajah[c], Fabien Scalzo[d] and David S. Liebeskind[d]

[a]Wayne State Department of Neurology, Wayne State University, Detroit, MI, USA; [b]Neuro-Epidemiology and Ageing Research Unit, School of Public Health, Imperial College London, London, UK; [c]Wayne State Department of Neurosurgery, Wayne State University, Detroit, MI, USA; [d]Departement of Neurology, University of California Los Angeles, Los Angeles, CA, USA

ABSTRACT
Advances in predictive analytics and machine learning supported by an ever-increasing wealth of data and processing power are transforming almost every industry. Accuracy and precision of predictive analytics have significantly increased over the past few years and are evolving at an exponential pace. There have been significant breakthroughs in using Predictive Analytics in healthcare where it is held as the foundation of precision medicine. Yet, although the research in the field is expanding with the profuse volume of papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care. This lack of impact stands in stark contrast to the enormous relevance of machine learning to many other industries. Regardless of the status of its current contribution, the field of predictive analytics is expected to fundamentally change the way we diagnose and treat diseases, as well as the conduct of biomedical science research. In this review, we describe the main tools and techniques in predictive analytics and will analyze the trends in application of these techniques over the recent years. We will also provide examples of its application in medicine and more specifically in stroke and neurovascular research and outline current limitations.

ARTICLE HISTORY
Received 4 January 2019
Accepted 12 April 2019

KEYWORDS
Machine learning; deep learning; predictive analysis; neurovascular

Introduction

We are now witnessing a disruption by Artificial Intelligence (AI) and Predictive Analytics in almost all industries, healthcare is no exception. The main features of this unprecedented transformation are the magnitude and the complexity at which information is being collected, digested, and applied to almost every single aspect of our human interactions and activities. This is enabled by a remarkable increase in the velocity and volume of data that is captured digitally, leveraged by a striking increase in computational power, and novel robust statistical algorithms and tools. The field of AI and Predictive Analytics is not new; it first appeared in the mid-twentiethx century followed by multiple cycles of hopes and hypes. However, the recent striking performance of certain algorithms (e.g. deep learning) in areas such as image recognition, speech recognition, and online marketing has reached and for some cases surpassed the human-level performance, which has renewed the interest and hope in this field globally.

In this era of Big Data, conventional statistical methods are not optimal because of several limitations [1]: (1) the algorithms require extensive cleansing process which necessitates human intervention; (2) limited aptitude to incorporate high-dimensional data as well as unstructured and large datasets; (3) very sensitive to the common characteristics of big datasets such as data sampling heterogeneity, missing data, and data complexity; and (4) complex dependencies between clinical events, treatment protocols, and the disease diagnosis and progress that might be very difficult to digest and analyze with conventional methods. On the other hand, modern predictive analytics start with a large amount of data and rely on various computational methods to build the most accurate predictive algorithms.

Today, predictive analytics and AI are becoming widely used in diverse industries such as finance, marketing, finance, retailing, and healthcare. In finance, High-Frequency Trading (HFT) uses predictive analytics algorithms to instantaneously analyze big data for optimal monetary gain. In the 'data economy' industry, companies such as Google, Apple, Amazon, Uber, and Airbnb have become global leaders by embedding predictive analytics into their 'DNA' and in almost every single facet of their services and operations [2]. Ironically, despite the great potentials given the sheer volume and variety of the medical information and the analytical nature of healthcare decision-making, the healthcare industry has not yet fully grasped the opportunities offered by Predictive Analytics. A major prospect of predictive analytic systems is to pave the way to the new era of precision-medicine, in which the

CONTACT Melek Somai, MD MPH msomai@mcw.edu Department of Medicine, Medical College of Wisconsin, Wisconsin.

diagnosis of diseases and the discovery of treatments are based on the knowledge extracted from the deep-phenotypic profile of everyone [3]. In that regard, we use the term Predictive Analytics to represent the application of Machine Learning, AI, and other big data tools to forecast future outcomes, events and to devise a decision plan [4].

In this review, we describe the main tools and techniques in predictive analytics. We explore the trend of their application in medicine. Finally, we discuss examples in neurovascular research, namely stroke, aneurysms, and vascular malformations, and outline the current limitations of their applications in healthcare.

Predictive analytics tools

The boundaries of the domain of Predictive Analytics and Artificial Intelligence (AI) are very difficult to outline. AI is a broad term and applies to techniques that enable machines and computers to mimic human intelligence using various algorithms. Predictive Analytics, born at the intersection of statistics and computer science, is the discipline that binds efforts to learn relationships from data with the efficient computing algorithms and processing power. Predictive Analytics offers the prospect to gain insights to not only understand and make sense of the data at hand but to utilize information that happened in the past to predict a best assessment of the unknown in the future and devise a data-driven course of action (Table 1).

In our review, we focus our description of 'Predictive Analytics' to the 'use of a variety of algorithms and statistical techniques, such as machine learning and neural networks to recognize patterns and identify the likelihood of future outcomes based on available data'.

Machine learning

Machine learning (ML) focuses on how computers learn from data using experimentation and is a major domain within predictive analytics [5]. ML encompasses several analytical algorithms which could be classified as supervised, unsupervised, or semi-supervised learning depending on whether the outcome of interest is incorporated into the model and used to 'train' the algorithm to reach a high predictive power.

Supervised machine learning

Supervised learning means that the algorithm is provided with both the outcome of interest and the set of input parameters. The algorithm is then trained to find some relationships between the outcome and the input values such as the patient traits (Figure 1(a)). Supervised learning considers the subjects' outcomes together with their traits and goes through a certain automated and iterative training process to determine the best outputs associated with the inputs. For instance, integrating imaging test results from patients with recent ruptured cerebral aneurysm, the supervised machine learning algorithm can be used to analyze the images and other relevant information of patients with unruptured cerebral aneurysm and be optimized to identify patterns that tend to indicate a higher risk of aneurysm rupture. Once trained, the model can then be integrated in Electronic Health Records and used in decision support systems to flag future images from patients who have not yet received a prognosis.

Unsupervised machine learning

In unsupervised learning, the outcome of interest is undefined. Instead, the algorithm tries to isolate naturally occurring patterns or groupings within the data (Figure 1(b)). It is blind to outcomes, hence the term unsupervised. Unsupervised machine learning is often used to extract useful features when building classifiers or other predictors and to identify patterns and dependencies of unstructured and structured multi-dimensional 'big-data'.

The two major unsupervised learning methods are *clustering* and *principal component analysis* (PCA). Clustering groups subject with similar traits together into clusters, without any outcome information. Popular clustering algorithms include *k-means clustering, hierarchical clustering* and *Gaussian mixture clustering*

Table 1. Comparison of traditional and modern predictive modeling.

	Traditional Models	Modern Predictive Analytics
Model	Static	Dynamic adaptive
Human intervention	Requires experts to handle data cleansing, data analysis, and interpretation	Could be fully automated
Scalability	Not scalable	Optimized for scalability and Big Data
Data Volume	Low	High
Data Dimensions	Low dimentionality	High dimentionality
Data Structure	Structured	Unstructed
Quality of data	Requires very good data. Very senstitive to missing data (specifically if not missing at random)	More robust in devising accurate prediction using noisy datasets with missing data
Algorithm	Regression, Clustering, General linear models	Deep learning, Neural Networks
Hardware	Traditional computer hardware	
Software	Generally proprietary (STATA, SAS) also Open Source (R)	Mainly open source community (TensorFlow)
Understandibility	Understandable association between the outcome and the predcition variables	Black box

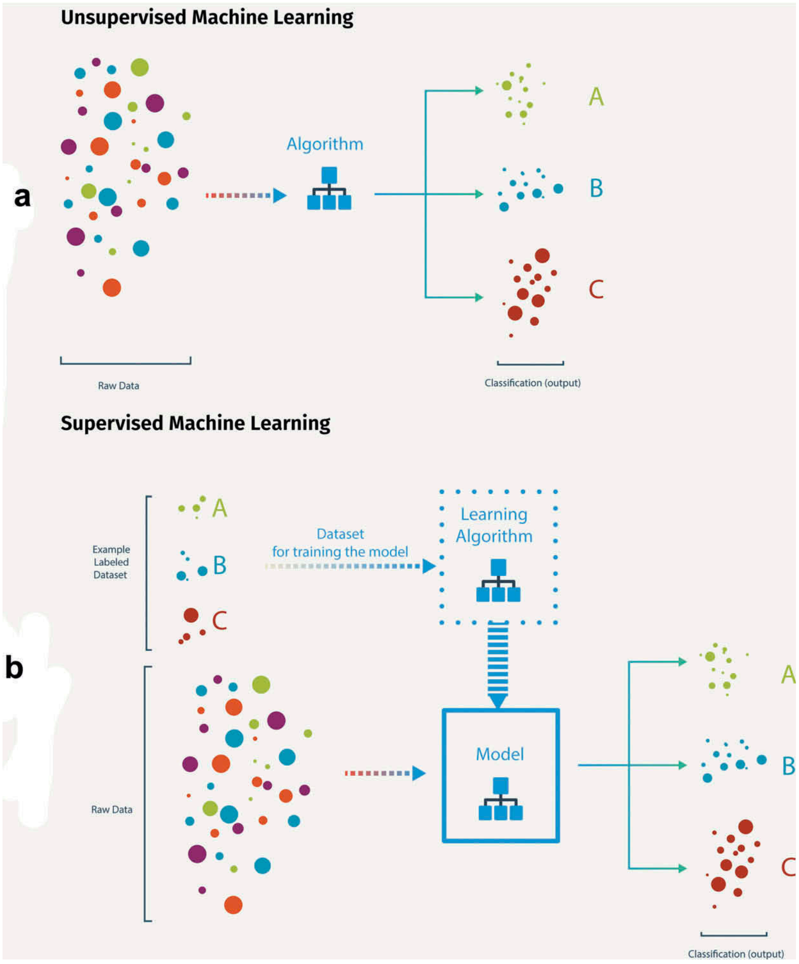

Figure 1. Supervised and unsupervised machine learning.

[6]. These algorithms are especially useful when the trait of interest is extracted from multi-dimensionsional datasets, such as the number of genes in a genome-wide association study. For example, unsupervised machine learning has been applied to extract embedded and latent features from EHR data [7]. In addition, unsupervised learning can be applied to explore new applications such as disease characterization for heterogeneous disorders (where the etiology is devised according to the various pathophysiologic mechanisms), which could in turn provide new paths to therapy.

Semi-supervised machine learning

Semi-supervised ML leverages both supervised and unsupervised ML (i.e. combination of both labeled and unlabeled data is fed into the algorithm). Labeled data is used to help machines identify that there are specific groups of features or outcomes present in the data. The algorithm is then trained on unlabeled data using unsupervised methods to identify new types of clusters that were not specified in the previous step. This is useful for processing massive amounts of data such as genetic sequencing where including both labeled and unlabeled data during the training process would improve the accuracy of the algorithm and reduce the time and the cost [8].

Reinforcement learning

Reinforcement learning algorithms are another way that machines can learn. These algorithms learn from interaction with the environment to achieve a goal and try to explore the best ways to earn the greatest reward. Goals or rewards could be winning a game or finding optimal treatment policies. These methods can develop optimal treatment strategies for problems with large state and overcome the computational complexity of the strategy-iterations. This approach has been shown to be capable of developing effective treatment strategies for sequential medical decision problems [9].

Neural networks and deep learning

A neural network can be considered as an extension of ML that captures complex non-linear relationships between input variables and an outcome and can address some of the limitations of traditional analytics. In neural network, the associations between the outcome and the input variables are computed using one or multiple hidden layers, each with a series of algorithms (nodes) that take the data of the previous layers and output new data that feeds the subsequent layers. Using large annotated datasets, the nodes and the hidden layers of the neural network are trained iteratively (thousands to millions of iterations) to adjust the characteristics of the algorithms (e.g. hyper-parameters) in a way to reach an optimized prediction model overall.

Deep neural networks resemble the learning process of the brain; hence its name 'neural'. For instance, a child points to an object and names it as 'dog'. This is followed by the parents' response of yes vs no. With time, the child learns more about the specifics of a dog as he or she points to various animals and builds a concept by using the knowledge acquired from prior layer of the hierarchy to clarify the complex concept of a dog. Deep learning algorithms use the same hierarchy. Initially, the system is provided with training data, e.g. a set of images labeled as "ischemic stroke" or "not ischemic stroke" by a human, aka "meta-tag". A deep learning algorithm uses this information to build a feature set for "ischemic stroke" by creating a predictive model. Using patterns of pixels in the digital data and with each iteration, the predictive model becomes more accurate and complex (Figure 2).

Predictive analytics in clinical medicine

There has been an increasing trend in the adoption of predictive analytics in medicine (Figure 3(a)). It is widely used in biomedical research, including the '-omics' fields (e.g. genomics, epigenomics, proteomics, lipidomics, and metabolomics) [10,11]. Other examples include ICU real-time risk prediction [12,13], accurate prediction of sepsis or septic shock [14–16], pathological labeling and pattern recognition for various types of pathologies including diagnosis and prognosis of tumors [17–21], and efficient processing of imaging biomarkers required for early detection and progression of various disorders [22]. A study using machine learning algorithm used positron emission tomography (PET) scan imaging data to train a model to look for differences between elderly patients with expected, age-related stable MCI and those with progressive MCI based on the regional information from amyloid [22–25]. Other growing applications are in the field of population health for management interventions and for tailoring cost-effective health-care interventions [26,27]. The role of artificial intelligence has been continuously growing in all aspects of neuroscience, with the development of international big data neuroscience initiatives, e.g. the Human Brain Project, the BRAIN initiative, the Human Connectome Project, and the National Institute of Mental Health's Research Domain Criteria initiative. Overall, these techniques facilitate predicting clinical events and promote the identification, stratification, and management of patients who are at highest risk of poor health outcomes or who will benefit most from specific therapeutics or interventions.

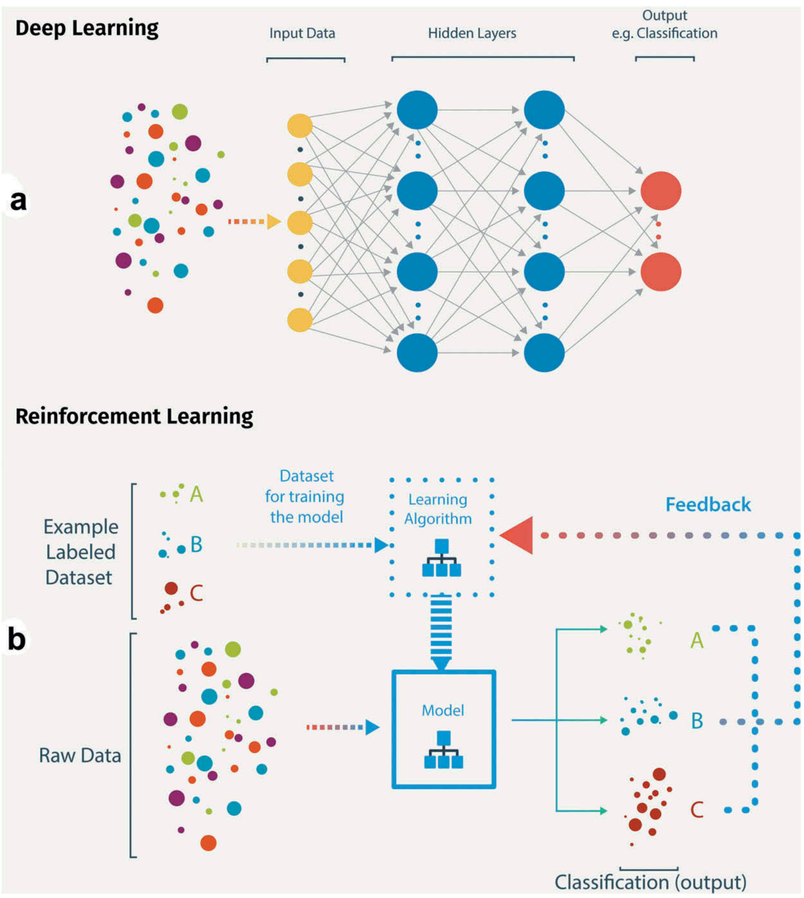

Figure 2. Deep learning hierarchy.

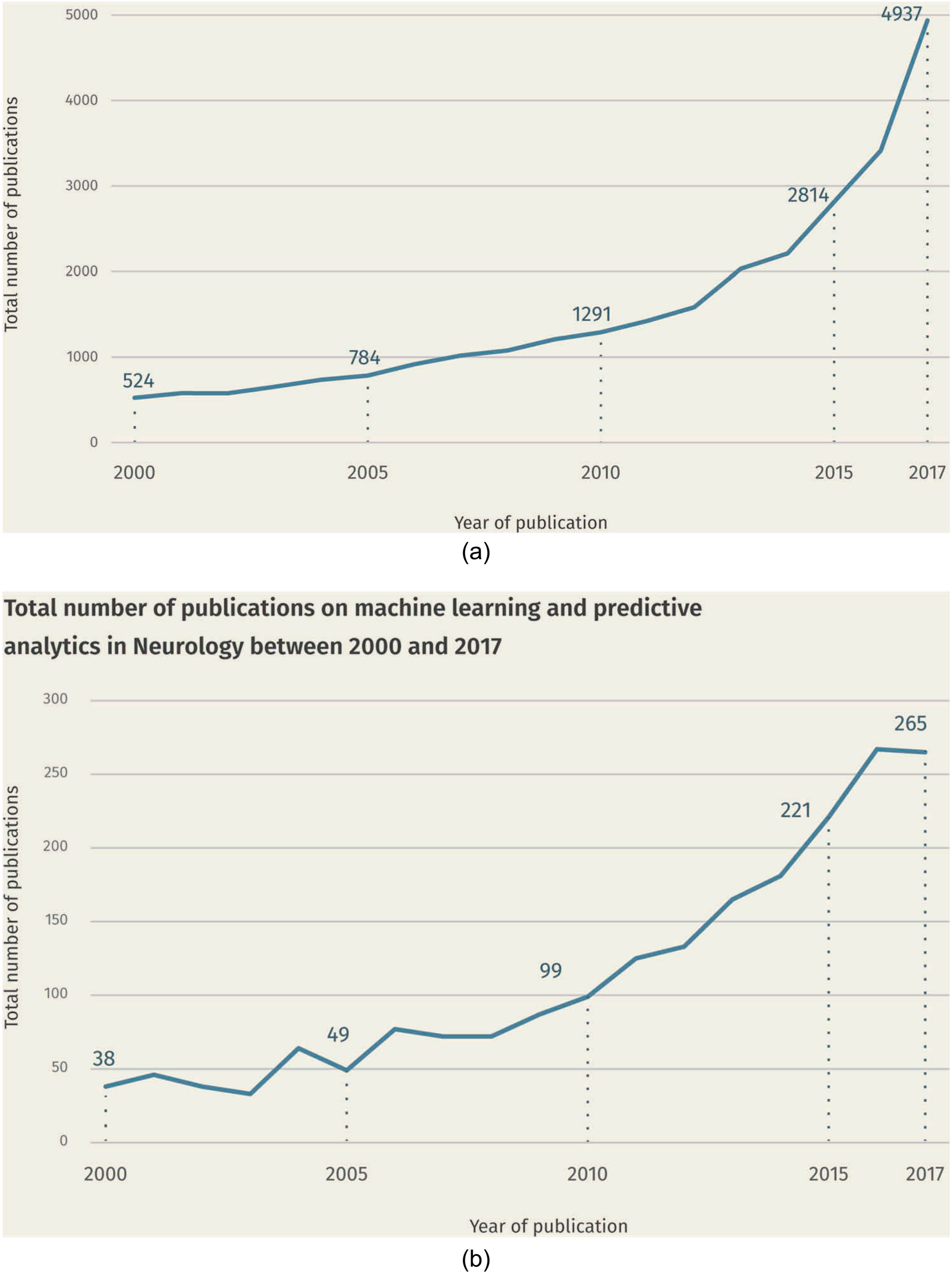

Figure 3. Predictive analytics trend in medicine and neurovascular disorders research.

Predictive analytics in neurovascular disorders

Predictive analytics is increasingly used in neurological research with a growing number of researches published in the last few years (Figure 3(b)). We discuss below some of these applications with focus on neurovascular conditions.

Intracranial vascular malformations

Cerebral aneurysms

Machine learning techniques have already started a new era to investigate the combination of factors that may affect the risk of cerebral aneurysm rupture. Ongoing research efforts are in progress to predict the risk of cerebral aneurysm rupture using machine learning and image processing techniques based on structured and non-structured data from big data [28]. Nakao et al. detected 94.2% (98/104) of aneurysms from non-contrast-enhanced time-of-flight images using a Convolutional Neural Network, a specific deep learning algorithm [29]. Millan et al. used a machine learning method to characterize the aneurysm rupture risk. This model demonstrated an 80% success rate in predicting whether an aneurysm is ruptured or not in 53 patients [30]. In another study, Bisbal et al. used a data mining approach in a database of 157 cases and used 294 features including basic demographics and clinical information, morphological characteristics computed from the patient's medical images, and individualized blood flow simulations to achieve 95% classification accuracy [31]. Lo et al. proposed a sophisticated Deep Learning approach, e.g.

Bayesian Fuzzy Logic Neural Networks, to create a prognosticating program for outcomes in aSAH [32]. Predictive Analytics is also bridging the gap between the biomedical informatics and medical research. For instance, multi-omics and high throughput DNA sequencing with whole exsome sequencing is being applied to aneurysm formation and rupture risk modeling [33].

Similarly, artificial neural network was compared to standard logistic regression models for the prediction of symptomatic vasospasm. A simple artificial neural network model was significantly more specific and sensitive than traditional regression for detection of vasospasm following SAH [34]. In another study of 25 patients with aneurysmal SAH, a neural network model achieved a sensitivity of 100% and specificity of 84% for the prediction of symptomatic cerebral vasospasm [35].

Arteriovenous malformations (AVMs)

There has been a great interest for developing outcome prediction tools for cerebral AVMs due to the young age of patients at diagnosis and the catastrophic neurological consequences following AVM rupture [36]. Today, we have the opportunity to build more accurate models in the age of machine learning. Recently, in a study of patients with brain AVMs, a machine learning model using neural networks showed superior accuracy (97.5%) for predicting fatal outcome after endovascular treatment over a mean follow-up of 5 years, while a conventional regression model demonstrated an accuracy of 43% [37]. Oermann et al. proposed a novel machine learning algorithm for predicting outcomes following AVM radiosurgery [38]. Using three prospective databases, the algorithm achieved a success rate measured as the average area under the curve (AUC) of 0.71 compared to existing clinical systems of 0.63, providing the best possible predictions of AVM radiosurgery outcomes to date. The approach also devised the 3D surface dose (defined as *AVM surface area* multiplied by *margin dose of delivered radiation)* as a novel radiobiological feature of prognostic outcome.

Predictive analytics in stroke

Predictive analytics has been applied to several aspects of stroke care, including accurate diagnosis, subtype classification, management strategies and prognosis of these patients [39].

Stroke diagnosis

In one study, researchers showed a high accuracy for lesion detection and segmentation by utilizing convoluted neural networks (CNN) to extract predictive features from diffusion-weighted images (DWI) [40]. Boldsen et al. proposed an unsupervised multimodal algorithm, ATLAS (Automatic Tree Learning Anomaly Segmentation) combining the DWI and apparent diffusion coefficient (ADC) data to detect acute ischemic lesions [41]. The performance of ATLAS was significantly higher than the current COMBAT score [42].

Other studies have focused on applying Predictive Analytics to interpret CT images. In one study, the Alberta Stroke Program Early Computed Tomography Score (ASPECTS) used a machine learning image classifier for the interpretation of brain CT and showed a non-inferior performance compared to stroke experts [43]. Chen et al. validated an ML algorithm that can detect and quantify cerebral small vessel disease on CT images, as accurately as expert consensus using FLAIR MRI images [44]. In another study, machine learning showed 97.5% sensitivity to detect the hyperdense middle cerebral artery (MCA) sign in CT images [45].

More advanced applications are integrating decision support systems with predictive analytics. The biotech company Viz.ai© has developed an FDA approved artificial intelligence platform that uses deep learning algorithms, similarly to the one used for detecting faces in an iPhone, to detect large vessel occlusion strokes from CTA and CTP images. When the algorithm detects large vessel occlusion, it instantly alerts a vascular neurologist and interventionalist through a smartphone application, thus saving significant amount of time by alerting the specialists more rapidly. Infervision© is another platform that uses Predictive Analytics technology to accurately estimate the volume, the type (ischemic vs hemorrhagic), and the location of the stroke. It then provides useful decision-making information for a surgical vs conservative management. On the other bay of the Atlantic, PRECISE4Q, a new EU collaborative project funded via Europe's 'Horizon 2020 Framework Programme for Research and Innovation' aims to develop personalized prevention, treatment and rehabilitation strategies using Predictive Analytics and AI [46].

Stroke prognosis and outcome

Ischemic stroke is a highly heterogenous condition with varying rates of progression and outcomes for different individuals. Findings in both functional and structural MRI have been used as input features in machine learning models to predict motor and cognitive outcomes after stroke [47–51]. ML algorithms have been shown efficient for prediction and quantification of the evolution of cerebral edema in patients with malignant hemispheric infarct [52]. Similarly, Bentley et al. reported significant accuracy (area under the ROC 0.74) for prediction of the symptomatic intracranial hemorrhage in acute ischemic stroke patients treated with IV thrombolysis that was superior to prior established prognostication tools [53]. Asadi et al. reported considerable accuracy using supervised machine learning to predict outcomes following endovascular vs medical management of stroke [54]. In another study, Xu et al. used machine learning to identify how individual stroke patients might respond differently to stroke treatments based on their unique lesion fingerprint [55].

Limitations

Several limitations and concerns must be resolved before we see these novel predictive analytical algorithms becoming mainstream in clinical practice: 1) *Data*: To work efficiently, predictive models need to be continuously trained with health-care data. Current medical scientific environment does not promote data sharing efficiently. Efforts are in progress to reform access and regulation of information exchange in health care [56,57]. Other obstacles include the issues involving the data governance and to the lack of standardization and interoperability of the major health-care platforms. 2) *Understandability*: One of the challenges is that most novel machine learning and deep learning algorithms are considered 'Black Boxes'. In traditional statistical methods, i.e. logistic regression, there is a clear and simple relationship between the input variables and the output variable (a regression formula). In Deep Learning, because of the complexity of the network and the number of parameters and hyperparameters, it is impossible to *explain* the relationship between the input parameters and the output. In addition, most novel deep learning is autonomous and adjust to novel data without the need of human intervention. 3) *Clinical Efficacy*: The ultimate measure of the impact of Predictive Analytics should be not limited to the accuracy of these methods, but to the extent to which these tools could improve clinical practice. This clinical efficacy is not only affected by the quality of the predictive model but is also related to how clinical practice will adapt and properly integrate these tools into the clinical decision process securely, efficiently, and ultimately for the benefit of the patient. Current research in the field of Predictive Analytics is still in its infancy regarding the matter and is still limited to controlled research environments. A wider adoption in the next decades will prove whether Predictive Analytics will translate into better clinical practice. 4) Safety: There are challenges and difficulty to verify the models for safety and reliability in clinical practice. 5) *Regulations*: There is currently no optimal regulatory forces in the area of machine learning driven applications, and current FDA standards are outdated for machine learning. 6) *Education and Training*: There is a large gap between the advancements in Predictive Analytics and its adoption by health-care professional. The prospect of incorporating Predictive Analytics into everyday practice of medicine raises the question whether clinicians and health-care professionals should be trained to analyze, incorporate and devising clinical decision supported by AI. It is becoming clear that the evolution of Medicine in the foreseeable future would require training practitioners about the techniques, methods, and the ethics of Data Science and Predictive Analytics.

Ethics

The extent of power of Predictive Analytics and Machine Learning in shaping Healthcare is challenging the current ethical frameworks and open a range of unanswered and complex ethical dilemmas. Some of these ethical challenges are discernible and require a more careful attention. This is the case for instance about the ethical implications of human biases that exist in the data which is then used to train machine learning to make clinical decision at scale. In other domains, the ethical challenges are more difficult and require the implication of a new set of inquiries in the realm of bioethics [56]. The fiasco with Facebook and Cambridge Analytica serves as an important reminder. These devices are nowadays capable of harvesting almost every datum of our human behavior. These data points could therefore be used to create a unique digital fingerprint of each individual and then used to predict behavioral outcomes. More importantly, it could be used to incessantly 'nudge' the individual towards prescribed behaviors. This raises several ethical questions that surpasses the current realm of hypocrite's oath and positions Predictive Analytics as a 'third-participant' in the relationship between a patient and health-care professionals. Certain advocates are calling researchers and institutions to adopt a new code of Ethics for AI.

Conclusion

Future predictive modeling will likely combine the use of a wide range of data points such as imaging, biomarkers, and real-time continuous monitoring using wearables. This deluge of data would feed advanced deep learning and neural network algorithms to accurately predict the risk of health conditions such as aneurysm rupture or stroke treatment. Moreover, Predictive Analytics steers towards the development of clinical pathways that are adaptive and continuously updated, and in which healthcare decision-making espouse the machine sophisticated algorithms to provide the best course of action effectively and safely. This also parlays into an advantage of being distinctly translatable into a practice-specific tool. The result is maximally accurate for clinical decision support in the management of individual patients rather than broadly generalizable knowledge about a disease state. Ultimately, machine learning techniques will provide a better insight into the molecular pathways and underlying mechanisms of disease processes by connecting pathological features with outcome data. The potential for predictive analytics to revolutionize many aspects of healthcare is clear in the horizon.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research received no specific grant from anyfunding agency in the public, commercial or not-for-profit sectors.

Notes on contributors

Hamidreza Saber is a physician and neurology resident in training at the Wayne State University School of Medicine, Detroit, MI. His focus is in the field of neurointerventional surgery. He is particularly interested in development of innovative methods for interventional management of neurovascular conditions.

Melek Somai is an Assistant Professor of Medicine at Medical College of Wisconsin, Milwaukee, WI. He is the co-director of the Clinical Informatics and Data Science Lab at Inception Research. He is also Principal in Digital Health at Inception Health. Dr. Melek's research focuses on data and engineering to redesign healthcare delivery.

Gary B. Rajah is the chief neurosurgical resident at the Wayne State University School of Medicine, Detroit, MI. with particular interest in minimally-invasive neuroendovascular surgical approaches for the treatment of cerebral vascular abnormalities and stroke.

Fabien Scalzo is the director of the AI in Imaging and Neuroscience Lab at the University of California, Los Angeles (UCLA). His works include developing Machine Learning and Computer Vision algorithms to improve our understanding of neurological disorders. Dr. Scalzo's research is directed to Stroke and traumatic brain injuries (TBI) using brain mapping of imaging and biosignals.

David S. Liebeskind is Professor of Neurology at the University of California, Los Angeles (UCLA) where he is Director of Outpatient Stroke and Neurovascular Programs. He is Director of the Neurovascular Imaging Research Core, leading global efforts to advance data science and precision medicine of stroke imaging for prevention, acute therapies and recovery after stroke.

ORCID

Melek Somai http://orcid.org/0000-0002-2692-9681

Contributorship Statement

HS and MS participated in the conception and design of the study. HS and MS analyzed and interpreted the data. GR, DL, FS, and HS revised the draft paper for intellectual content.

References

[1] Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. IEEE J Biomed Heal Informatics. 2017;21:4–21.
[2] Chen X-W, Lin X. Big data deep learning: challenges and perspectives. IEEE Access. 2014;2:514–525.
[3] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372:793–795.
[4] Cohen IG, Amarasingham R, Shah A, et al. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff. 2014;33:1139–1147.
[5] Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc. 2016;3:205395171562251.
[6] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv. 1999;31:264–323.
[7] Miotto R, Li L, Kidd BA, et al. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6:26094.
[8] Peikari M, Salama S, Nofech-Mozes S, et al. A cluster-then-label semi-supervised learning approach for pathology image classification. Sci Rep. 2018;8:7193.
[9] Poolla R A reinforcement learning approach to obtain treatment strategies in sequential medical decision problems. [Grad Theses Diss]. [updated 2003 Jun 1; cited 2018 Jun 19]. Available from: http://scholarcommons.usf.edu/etd/1450
[10] Chaudhary K, Poirion OB, Lu L, et al. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res. 2018;24:1248–1259.
[11] Zhu B, Song N, Shen R, et al. Integrating clinical and multiple omics data for prognostic assessment across human cancers. Sci Rep. 2017;7:16954.
[12] Suresh H, Hunt N, Johnson A, et al. Clinical intervention prediction and understanding using deep networks. [updated 2017 May 23; cited 2018 May 26]. Available from: http://arxiv.org/abs/1705.08498
[13] Pirracchio R, Petersen ML, Carone M, et al. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med. 2015;3:42–52.
[14] Nemati S, Holder A, Razmi F, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2018;46:547–553.
[15] Shimabukuro DW, Barton CW, Feldman MD, et al. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4:e000234.
[16] Desautels T, Calvert J, Hoffman J, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inf. 2016;4:e28.
[17] Christiansen EM, Yang SJ, Ando DM, et al. In silico labeling: predicting fluorescent labels in unlabeled images. Cell. 2018;173:792–803.e19.
[18] Yu K-H, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun. 2016;7:12474.
[19] Macyszyn L, Akbari H, Pisapia JM, et al. Imaging patterns predict patient survival and molecular

subtype in glioblastoma via machine learning techniques. Neuro Oncol. 2016;18:417–425.
[20] Liu Y, Gadepalli K, Norouzi M, et al. Detecting cancer metastases on gigapixel pathology images. [updated 2017 Mar 3; cited 2018 May 26]. Available from: http://arxiv.org/abs/1703.02442
[21] Cruz-Roa A, Gilmore H, Basavanhally A, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Sci Rep. 2017;7:46450.
[22] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
[23] Mathotaarachchi S, Pascoal TA, Shin M, *et al.* Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiol Aging. 2017;59:80–90.
[24] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–118.
[25] Narula S, Shameer K, Salem Omar AM, et al. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. J Am Coll Cardiol. 2016;68:2287–2295.
[26] Hernandez I, Zhang Y. Pharmaceutical use and spending trend in medicare beneficiaries with dementia, from 2006 to 2012. Gerontol Geriatr Med. 2017;3:2333721417704946.
[27] Thompson S, Sasinowski M, Zedler B, et al. Predictive analytics can help reduce prescription opioid overdoses and health care costs; [cited 2018 May 26]. Published Online First: 2017. Available from: http://venebio.com/wp-content/uploads/2018/01/Venebio-Opioid-Advisor-HEOR.pdf
[28] Chabert S, Mardones T, Riveros R, et al. Applying machine learning and image feature extraction techniques to the problem of cerebral aneurysm rupture. Res Ideas Outcomes. 2017;3:e11731.
[29] Nakao T, Hanaoka S, Nomura Y, et al. Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. J Magn Reson Imaging. 2018;47:948–953.
[30] Millan RD, Dempere-Marco L, Pozo JM, *et al.* Morphological characterization of intracranial aneurysms using 3-D moment invariants. IEEE Trans Med Imaging. 2007;26:1270–1282.
[31] Bisbal J, Engelbrecht G, Villa-Uriol M-C, et al. Prediction of cerebral aneurysm rupture using hemodynamic, morphologic and clinical features: a data mining approach. Berlin, Heidelberg:Springer; 2011. 59–73. DOI: 10.1007/978-3-642-23091-2_6.
[32] Lo BWY, Macdonald RL, Baker A, et al. Clinical outcome prediction in aneurysmal subarachnoid hemorrhage using Bayesian neural networks with fuzzy logic inferences. Comput Math Methods Med. 2013;2013:904860.
[33] Farlow JL, Lin H, Sauerbeck L, et al. Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. PLoS One. 2015;10:e0121104.
[34] Dumont TM, Rughani AI, Tranmer BI. Prediction of symptomatic cerebral vasospasm after aneurysmal subarachnoid hemorrhage with an artificial neural network: feasibility and comparison with logistic regression models. World Neurosurg. 2011;75:57–63.
[35] Dumont TM. Prospective assessment of a symptomatic cerebral vasospasm predictive neural network model. World Neurosurg. 2016;94:126–130.
[36] Friedman WA, Blatt DL, Bova FJ, et al. The risk of hemorrhage after radiosurgery for arteriovenous malformations. J Neurosurg. 1996;84:912–919.
[37] Asadi H, Kok HK, Looby S, et al. Outcomes and complications after endovascular treatment of brain arteriovenous malformations: a prognostication attempt using artificial intelligence. World Neurosurg. 2016;96:562–569.e1.
[38] Oermann EK, Rubinsteyn A, Ding D, et al. Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. Sci Rep. 2016;6:21161.
[39] Feng R, Badgeley M, Mocco J, et al. Deep learning guided stroke management: a review of clinical applications. J Neurointerv Surg. 2018;10:358–362.
[40] Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. NeuroImage Clin. 2017;15:633–643.
[41] Boldsen JK, Engedal TS, Pedraza S, et al. Better diffusion segmentation in acute ischemic stroke through automatic tree learning anomaly segmentation. Front Neuroinform. 2018;12:21.
[42] Valdes-Sosa PA, Popa-Wagner A, Llibre-Guerra JJ, et al. Better diffusion segmentation in acute ischemic stroke through automatic tree learning anomaly segmentation. Front Neuroinform. 2015;12. DOI:10.3389/fninf.2018.00021.
[43] Nagel S, Sinha D, Day D, et al. e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. Int J Stroke. 2017;12:615–622.
[44] Chen L, Carlton Jones AL, Mair G, et al. Rapid automated quantification of cerebral leukoaraiosis on CT images: a multicenter validation study. Radiology. 2018;171567. DOI:10.1148/radiol.2018171567.
[45] Takahashi N, Lee Y, Tsai D-Y, et al. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. Radiol Phys Technol. 2014;7:79–88.
[46] Optimizing personalized treatment strategies for stroke patients: Charité – Universitätsmedizin Berlin; [cited 2018 Jun 10]. Available from: https://www.charite.de/en/service/press_reports/artikel/detail/personalisierte_und_optimierte_versorgung_von_schlaganfallpatienten/
[47] Rondina JM, Filippone M, Girolami M, et al. Decoding post-stroke motor function from structural brain imaging. NeuroImage Clin. 2016;12:372–380.
[48] Siegel JS, Ramsey LE, Snyder AZ, et al. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. Proc Natl Acad Sci. 2016;113:E4367–76.
[49] Heo J, Yoon J, Park HJ, et al. Abstract 194: machine learning-based model can predict stroke outcome. Stroke. 2018 accessed 2018 Jun 7;49: http://stroke.ahajournals.org/content/49/Suppl_1/A194
[50] Bochniewicz EM, Emmer G, McLeod A, et al. Measuring functional arm movement after stroke using a single wrist-worn sensor and machine learning. J Stroke Cerebrovasc Dis. 2017;26:2880–2887.

[51] Rehme AK, Volz LJ, Feis D-L, et al. Identifying neuroimaging markers of motor disability in acute stroke by machine learning techniques. Cereb Cortex. 2015;25:3046–3056.
[52] Chen Y, Dhar R, Heitsch L, et al. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. NeuroImage Clin. 2016;12:673–680.
[53] Bentley P, Ganesalingam J, Carlton Jones AL, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. NeuroImage Clin. 2014;4:635–640.
[54] Asadi H, Dowling R, Yan B, et al. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One. 2014;9:e88225.
[55] Xu T, Rolf Jäger H, Husain M, et al. High-dimensional therapeutic inference in the focally damaged human brain. Brain. 2018;141:48–54.
[56] Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. N Engl J Med. 2018;378:981–983.
[57] Bauchner H, Golub RM, Fontanarosa PB. Data sharing. an ethical and scientific imperative. JAMA. 2016;315:1238–1240.