Narrative Review Article

Are Thinking Machines Breaking New Frontiers in Neuro-Oncology? A Narrative Review on the Emerging Role of Machine Learning in Neuro-Oncological Practice

Abstract

Medical science in general and oncology in particular are dynamic, rapidly evolving subjects. Brain and spine tumors, whether primary or secondary, constitute a significant number of cases in any oncological practice. With the rapid influx of data in all aspects of neuro-oncological care, it is almost impossible for practicing clinicians to remain abreast with the current trends, or to synthesize the available data for it to be maximally beneficial for their patients. Machine-learning (ML) tools are fast gaining acceptance as an alternative to conventional reliance on online data. ML uses artificial intelligence to provide a computer algorithm-based information to clinicians. Different ML models have been proposed in the literature with a variable degree of precision and database requirements. ML can potentially solve the aforementioned problems for practicing clinicians by not just extracting and analyzing useful data, by minimizing or eliminating certain potential areas of human error, by creating patient-specific treatment plans, and also by predicting outcomes with reasonable accuracy. Current information on ML in neuro-oncology is scattered, and this literature review is an attempt to consolidate it and provide recent updates.

Keywords: *Artificial intelligence, machine learning, neuro-oncology*

Mustafa Mushtaq Hussain[1], Ainsia Shabbir[2], Saqib Kamran Bakhshi[1], Muhammad Shahzad Shamim[1]

[1]Department of Neurosurgery, Aga Khan University Hospital, [2]Department of Computer and Information Systems Engineering, NED University of Engineering and Technology, Karachi, Pakistan

Introduction

Neuro-oncological practice routinely involves confrontation with questions regarding risks, benefits, and outcomes of the surgical interventions for neoplastic lesions. Decision-making is often influenced by surgeons' experience, in addition to evidence from the literature. It is not uncommon for surgeons to get dismayed by a lack of consolidated data to provide evidence-based recommendations to individual patients, especially when dealing with unusual pathologies, or if confronted with a combination of complex diseases. Neurosurgical procedures are prone to the risk of worsening neurological status and to allow learning from each other and to minimize adverse outcomes; a vast amount of biomedical data are published each year. To extract meaningful information using conventional statistics from this "huge data" is overwhelming for a practicing surgeon.[1] Statistical methods employed in medical research make assumptions in determining the level of significance (e.g., setting a *P* value) by estimating the correlation between variables and draw population inference from the sample. Statistical inferences become less precise when the number of input variables and possible associations among them increase.[2]

Machine learning (ML) is a field of computer science that studies algorithms and techniques for automating solutions to complex problems. It differs from traditional statistical methods in that it learns from a set of labelled data, and the larger the dataset, the more robust it becomes.[3] ML builds complex computational models that can process information from raw data and generate the outcome of interest. Neuro-oncological practice is encompassed by a myriad of diagnostic and therapeutic challenges, with a growing need to tailor therapy to the individual patient to achieve the best possible outcomes. ML models have proven as the new armamentarium for clinical experts with widespread utility in neuro-imaging, histopathological grading,

Address for correspondence:
*Dr. Muhammad Shahzad Shamim,
Aga Khan University Hospital,
Stadium Road, P.O. Box 3500,
Karachi 74800, Pakistan.
E-mail: shahzad.shamim@
aku.edu*

designating the best treatment options, and as outcome predictors.

The current review aims to provide a brief overview of the conceptual background behind ML and provide insight into its practical application in neuro-oncological care and outcome prediction.

Machine Learning Overview

ML can be broadly categorized into supervised learning (SL), semi-supervised, unsupervised, and reinforcement learning.

In SL algorithms, machine is registered with a set of datasets with right answers, i.e., a "*labelled dataset*" to a question pertaining to the data points. The model utilizes the key characteristic features of each data point and predicts the outcome; if any unseen data are entered, the algorithm predicts the outcome. The simple utility of the SL model could be seen in brain tumor detection in brain magnetic resonance imaging (MRI), where information about lesion's shape, length, consistency, and vascularity is used to classify lesions into normal and abnormal, with abnormal being subclassified into benign and malignant tumors.[4] If the model is contained with too many features relative to the number of cases, it may incorporate random error or noise as a signal, also referred to as *overfitting*. This results in reduced generalizability to unseen data and an increase in error. To overcome this problem, the SL model should be tested on data not involved in the learning process, also referred to as the *validation set*. Three most common supervised ML algorithms are [Figure 1]:
1. Decision tree: Algorithm that makes a group of items based on their values. Each tree consists of nodes and branches. Nodes represent questions about the data and branches denote possible answers
2. Naïve bayes: Based on Bayes' theorem, it creates trees based on their probability of occurrence. Mainly used for clustering purposes[5]
3. Support vector machine (SVM): SVM works principally by identifying some pattern in data points and draws a margin between the data groups called hyperplane, to separate into two classes based on pattern difference. SVM model is good for nonlinear relationships but is sensitive to outliers.[6]

In unsupervised learning, the machine is provided with a dataset and no right answer is provided. i.e., "*un-labelled data.*" The machine will determine the trend of similarity among items and generate the clusters. Here, the aim is to predict patterns in the data rather than an outcome. With the ability to find hidden relationships within data, unsupervised learning algorithms have applications in association and clustering tasks. For example, to identify patterns in genomic data for brain tumor patients.[7] The two main algorithms for clustering are given below:
1. K-means clustering: It automatically creates clusters, and items with similar features are placed in the same cluster. The mean value of a particular cluster lies in the center of that cluster
2. Principal component analysis (PCA): PCA reduces the dimensionality of data using orthogonal transformation, and by doing that reduces the use of a large amount of computational power.

Semi-SL lies between supervised and unsupervised learning, in which few data points are labelled. The algorithm will run clustering techniques to locate groups, and will identify a few labelled data points to provide labels to other data points in the group. It spares time and effort in labelling all the data points.

Reinforcement learning involves learning the ideal behavior within specific circumstances based on reward feedback mechanism. The algorithm aims to maximize the total amount of reward. For example,the Q learning agent, a basic form of reinforcement learning model, that interacts with virtual glioblastoma multiform (GBM) to learn and identify tumor parameters to get the best response with Temozolomide therapy and thus providing an appropriate mathematical framework for the optimal chemotherapy regimen in GBM patients.[8]

Neural Network learning (or artificial neural network [ANN]) is based on the biological concept of neurons. The input layer receives input (like dendrites), hidden layer processes the input (like soma), and the output layer sends the calculated output (like axonal terminals).[9] ANNs are universal predictors that can be applied to a wide variety of data, better represent complex biological processes that have nonlinear nature. Use of ANN in clinical decision-making for example involves symptom recognition, imaging analysis, and clinical diagnosis interpretation, etc.

Deep learning (DL) is a subset of ML and is widely based on ANN. The term deep signifies the number of hidden layers that increases in DL compared to a regular ANN. DL algorithm can work on diverse, unstructured, and inter-connected data without need of any manual feature extraction like that needed in ANN. Some most common DL algorithms are deep neural networks, deep belief networks, recurrent neural networks, and convolutional neural networks (CNNs).[10]

CNN is one of the most sought after deep neural network algorithm, working mainly on images and videos. CNN following the basic model of DL consists of multiple hidden layers along with the input and output layers. A convolutional layer extracts features from the input image using small matrices of input data while conserving relationship between pixels. A pooling layer reduces the number of parameters needed to learn the input to reduce dimensionality and finally a fully connected layer that flattens image into a column vector and forward it to the regular neural network that finally classify the given input.[11]
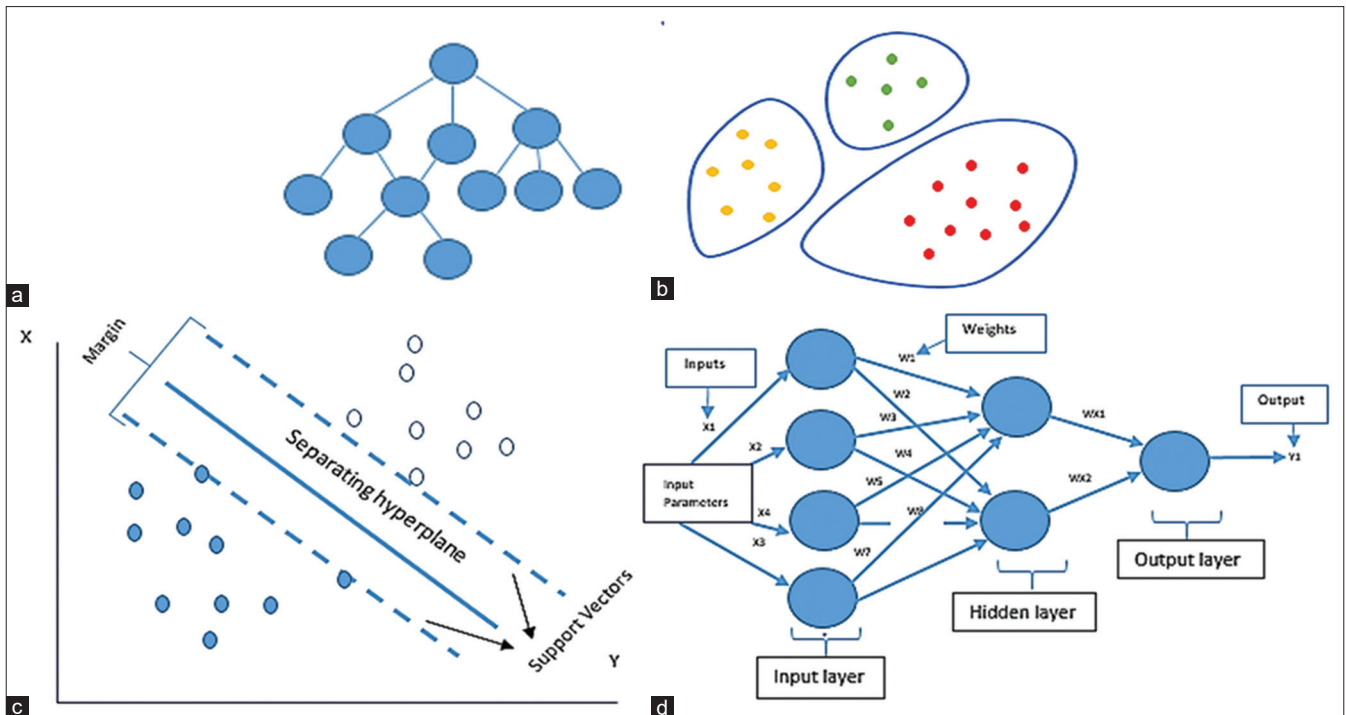
Figure 1: (a) Decision tree algorithm: A supervised learning algorithm that models a decision tree having nodes and edges using the data sets, answers any query using a series of questions with answers usually consisting of binary value. It classifies data and predicts any new dataset based on the modeled tree. (b) K-mean algorithm: An unsupervised learning algorithm that clusters the input data based on a similarity value. It has moderate to high efficiency and is used for problems where the data are not highly dimensional. (c) Support vector machine: Classify data points by selecting the "separating hyperplane" to separate the data into two classes based on pattern difference. (d) Artificial neural networks: Simulate the behavior of a biological neuron and are organized in layers of interconnected nodes, with nodes in the input layer receiving input features and hidden layers process the input to relay through the output layer

Methods

A literature search was performed using PubMed. The primary aim was to review all indexed publications in English language medical journals. The search syntax included a combination of Mesh keywords ("machine learning, brain neoplasms, diagnostic imaging, pathology, therapy, surgery, radiotherapy, survival outcome, and prognosis") entered in PubMed search builder without any publication time limits. All studies that evaluated ML models application in neuroimaging, diagnosis, therapy, histopathological grading, and prognostication in neuro-oncological practice were included. We excluded animal-based studies, conference abstracts, case reports, ongoing clinical trials, book chapters, editorials, letters to the editor, articles without full text, and non-English language publications. Search terms yielded 27 research articles, out of which nine articles were included for a brief discussion. Nineteen articles were excluded, as they were not relevant to the review question after titles and abstract screening. The narrative approach was used to summarize the key findings of each study included.

Discussion

Glial tumors grading

Much of the research in neuro-oncology is focused on diffuse gliomas. World Health Organization (WHO) has graded gliomas into lower-grade (WHO Grades I and II) and higher grade (Grade III and glioblastoma or Grade IV). Conventional MRI sequences are good at delineating tumor morphology but the delineation of infiltration of adjacent brain parenchyma on the T2-weighted image or fluid attenuated inversion recovery (FLAIR) sequence is nearly impossible. Diffusion tensor imaging (DTI) and diffusion kurtosis imaging (DKI) are advanced MRI sequences and have been investigated for preoperative prediction of glioma grade. DTI uses a Gaussian distribution model to image the diffusion behavior of water molecules[12] while DKI assumes non-Gaussian diffusion of water molecules.[13]

In the study conducted by Takahashi *et al.*, ML models were used to review MRI sequences of glioma patients and to preoperatively distinguish glioblastoma from lower-grade gliomas (Grades 2 and 3). ML model was created using six specific features extracted from apparent diffusion coefficient (ADC) and mean kurtosis (MK) - a type of diffusion kurtosis imaging. and generated 504 differentiating features, both semantic (e.g., location, shape) and agnostic (e.g., individual voxels) with significant differences (false discovery rate <0.05) between high and low-grade glioma. The SVM successfully predicted the preoperative glioma grades with area under the curve (AUC) values of 0.93 ± 0.03 and 0.91.[14]

Outcome Prediction

Peeken *et al*. in their retrospective study used radiomic

models (the science of extraction of quantitative data from medical images using algorithms) and combined imaging and treatment features to elucidate prognostic factors of GBM. One hundred and eighty-nine patients with GBM, who had received adjuvant chemo-radiation were included. MRI features based on Visually Accessible Rembrandt Images set, which is a system created to enable consistent description of gliomas, were employed. Multiple random survival forest prediction models were generated based on the patient training set, and internal validation was performed. These models combined clinical, pathological, and radiological features with treatment. MRI-based model had the highest prediction performance for overall survival (C-index: 0.61 [95% confidence interval (CI): 0.51–0.72]) and progression-free survival (C-index: 0.61 [0.50–0.72]). A combination of all the factors including treatment-related information further increased prognostic performance up to C-indices of 0.73 (0.62–0.84) for overall survival.[15]

Papp *et al.* had included in their study seventy patients with treatment-naïve glioma that was L-S-methyl-^{11}C-methionine (^{11}C-MET) positron emission tomography (PET)-positive (*in vivo* features), and histopathological grading and isocitrate dehydrogenase 1 R132H mutational status was known (*ex vivo* features). Using ML three predictive models were created to predict 36 months survival. One model was based on a combination of *in vivo*, *ex vivo*, and patient information ($M36_{IEP}$); second was based on *in vivo* and patient information only ($M36_{IP}$), and a third was based on *in vivo* information only ($M36_I$). $M36_{IEP}$ model after cross-validation was noted to have the highest AUC value of 0.9. It demonstrated that patients' younger age (<45 years), IDH-R132H positive status, smaller tumor volume, and lesser tumor-to-background ratio on ^{11}C-MET PET scan were more likely to have achieved 36 months survival. Apart from patients' clinical characteristics and histopathological grading, these validated ML models in this study quantified tumor shape features (such as spherical dice coefficient and volume) on imaging, and showed improved predictability, thus signifying the vital role of ML models application in brain tumors survival prognostication.[16]

In higher-grade gliomas, DTI features can also help in predicting survival differences by providing information about white matter integrity.[17] Functional MRI can also reflect angiogenesis around the tumor field which is a key feature of malignancy.[18] Dong *et al.*, reported the adoption of three-dimensional (3D) CNNs to automatically extract features from preoperative brain images. Sixty-nine patients with high-grade gliomas were divided into two groups: those who had survived more than 22 months (35 subjects) and those who had survival less than 22 months (34 subjects). 3D CNNs were trained to learn features from MRI related to survival time prediction and final output of extracted features were fed into the SVM for survival prediction model with an accuracy of 89.9%.

This study highlights another important functional role of ML in neuro-oncology.[19]

In a retrospective analysis of 400 patients who had trans-sphenoidal resection of pituitary adenoma, multivariate odds ratio analysis revealed that age <40 years was associated with 2.86 greater odds of postoperative diabetes insipidus, and patients with body mass index of <30 were more likely to develop postoperative hyponatremia. After model training, a logistic regression model with elastic net was able to predict similar early postoperative outcomes after pituitary adenoma surgery with an overall accuracy of 87%, (AUC value of 82.7).[20]

Brain Metastases

The response of brain metastases (BM) to stereotactic radiosurgery (SRS) has been demonstrated by the use of CNN-based ensemble radiomic models, which interpret computer tomography (CT) images. CNN-based ML models were taught pairs of tumor images and responses to SRS and then were used to predict SRS responses for unlearned images. Out of 110 tumor images, 57 images were classified as responders to SRS and 53 images as nonresponders to SRS. Tumors diameters and total dose of radiation between the two groups did not significantly differ. The greatest number of tumors in the responder group was mainly of breast (40%), followed by lung (35%), while in the nonresponder group, the most frequent site was lung (30%) followed by breast (25%). Trained ensemble neural models which comprised of 10 individual neural networks had better predictive performance than the individual neural network with AUC values ranged from 0.761 (95% CI = 55.2%–97.1%) to 0.856 (95% CI = 68.2%–100%). After learning from planning CT images, CNN-based radiomic models were highly accurate in predicting the BM response to SRS from unlearned images.[21]

Takada *et al.* created ML models using an alternating decision tree algorithm, wherein the predictions of multiple decision trees were integrated in a process called ensemble methods to predict the chances of disease-free survival (DFS) and BM within 5 years after neoadjuvant chemotherapy plus trastuzumab in postoperative breast cancer patients with human epidermal growth factor receptor 2-positive status. The DFS and BM models had a high accuracy in predicting prognosis with the AUC values were 0.785 (95% CI = 0.740–0.831, P<0.001) for the DFS model and 0.871 (95% CI = 0.830–0.912, P<0.001) for the BM model.[22] These models can optimize future surveillance methods in breast cancer patients, which is the second only to lung cancer for the development of BM.[23]

Gauging Clinical Response

Follow up of high-grade brain tumors heavily relies on the Response Assessment in Neuro-oncology criteria (RANO

criteria) that utilizes the measurement of enhancing and nonenhancing tumor components to assess disease progression or complete, partial, or no response to primary therapy. Blumenthal *et al*. evaluated 140 MRI scans of 32 high-grade gliomas and six patients with BM. All patients with high-grade lesions had a recurrence and had been treated with standard chemoradiation. SVM classifier system was trained to classify lesions based on four components: enhancing and nonenhancing, tumor, and nontumor, based on T1-weighted, FLAIR, and dynamic-contrast-enhancing MRI sequences. SVM classifier results were cross-validated. One hundred percent sensitivity and specificity was noted in detecting enhancing and nonenhancing areas in lesions. In 27 patients with high-grade lesions consistent results were attained by SVM classifier between changes in the volume of the lesion, and radiologist's review on follow up scans. However, in 5 (16%) patients increase in the volume of the nonenhancing tumor component was detected prior to the diagnosis made by radiologist (on RANO criteria) by several months. This proposed automatic RANO criteria system might help in future in improving therapy response assessment and progression monitoring.[24]

Limitations of machine learning

ML models have also been phrased as "black boxes."[25] There are debates about problems looming around its regulation, or whether artificial intelligence technology will remain in the hands of the few. One of the major limitations of ML models is that intent and causation relations are difficult to prove.[26] These ML algorithms are capable of internalizing massive data and can use it to make decisions like humans, without ever being able to communicate their reasons. The recent development of methods such as saliency maps could unravel the black-box nature of these models by cross-examining internal algorithm feature vectors.[27] Another possible challenge is to get the availability of large heterogeneous data to further improve the generalizability of results across the population.[28] Sharing of data among hospitals could help mitigate this data gap. ML models will never replace human expertise but can help strengthen clinical decision-making process in neuro-oncological patient care, and can bring efficiency and consistency in delivering precision medicine.[29]

Conclusion

ML models are robust and reasonably accurate predictive algorithms, with the ability to apprehend all previous institutional experiences and creating an individualized patient care plan. The use of these models in neuro-oncological practice can help physicians in effective communication with patients and their families regarding disease and its outcomes. ML models in neuro-oncology are likely to play an important role in achieving evidence-based and efficient, individualized patient care.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

References

1. Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett 2016;382:110-7.
2. Bzdok D, Altman N, Krzywinski M. Points of Significance: Statistics versus Machine Learning. Nature Publishing Group; 2018.
3. Rebala G, Ravi A, Churiwala S. Machine Learning Definition and Basics. InAn Introduction to Machine Learning 2019 (pp. 1-17). Springer, Cham.
4. George DN, Jehlol HB, Oleiwi AS. Brain tumor detection using shape features and machine learning algorithms. Int J Adv Res Comp Sci Software Eng 2015;5:454-9.
5. Lowd D, Domingos P, editors. Naive Bayes Models for Probability Estimation. Proceedings of the 22nd International Conference on Machine Learning; 2005.
6. Land WH, Schaffer JD. The Support Vector Machine. In The Art and Science of Machine Intelligence 2020 (pp. 45-76). Springer, Cham.
7. Senders JT, Zaki MM, Karhade AV, Chang B, Gormley WB, Broekman ML, *et al*. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien) 2018;160:29-38.
8. Ebrahimi Zade A, Shahabi Haghighi S, Soltani M. Reinforcement learning for optimal scheduling of Glioblastoma treatment with Temozolomide. Comput Methods Programs Biomed 2020;193:105443.
9. Sharma V, Rai S, Dev A. A comprehensive study of artificial neural networks. International Journal of Advanced research in computer science and software engineering. 2012 Oct;2(10).
10. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep learning for health informatics. IEEE journal of biomedical and health informatics. 2016 Dec 29;21:4-21.
11. Sakib S, Ahmed N, Kabir AJ, Ahmed H. An Overview of Convolutional Neural Network: Its Architecture and Applications; 2019.
12. Caulo M, Panara V, Tortora D, Mattei PA, Briganti C, Pravatà E, *et al*. Data-driven grading of brain gliomas: A multiparametric MR imaging study. Radiology 2014;272:494-503.
13. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magnetic resonance in medicine: An Off J Int Soc Magnetic Res Med 2005;53:1432-40.
14. Takahashi S, Takahashi W, Tanaka S, Haga A, Nakamoto T, Suzuki Y, *et al*. Radiomics analysis for glioma malignancy evaluation using diffusion kurtosis and tensor imaging. Int J Radiat Oncol Biol Phys 2019;105:784-91.
15. Peeken JC, Goldberg T, Pyka T, Bernhofer M, Wiestler B, Kessel KA, *et al*. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. Cancer Med 2019;8:128-36.
16. Papp L, Pötsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M, *et al*. Glioma survival prediction with combined analysis of *in vivo* 11 C-MET PET features, *ex vivo* features, and patient features by supervised machine learning. J Nucl Med 2018;59:892-9.
17. Zacharaki EI, Morita N, Bhatt P, O'Rourke DM, Melhem ER, Davatzikos C. Survival analysis of patients with high-grade gliomas based on data mining of imaging variables. AJNR Am J

Neuroradiol 2012;33:1065-71.

18. Pillai JJ, Zacá D. Clinical utility of cerebrovascular reactivity mapping in patients with low grade gliomas. World J Clin Oncol 2011;2:397-403.

19. Nie D, Zhang H, Adeli E, Liu L, Shen D, editors. 3D Deep Learning for Multi-Modal Imaging-Guided Survival time Prediction of Brain Tumor Patients. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016.

20. Hollon TC, Parikh A, Pandian B, Tarpeh J, Orringer DA, Barkan AL, *et al*. A machine learning approach to predict early outcomes after pituitary adenoma surgery. Neurosurg Focus 2018;45:E8.

21. Cha YJ, Jang WI, Kim MS, Yoo HJ, Paik EK, Jeong HK, *et al*. Prediction of response to stereotactic radiosurgery for brain metastases using convolutional neural networks. Anticancer Res 2018;38:5437-45.

22. Takada M, Sugimoto M, Masuda N, Iwata H, Kuroi K, Yamashiro H, *et al*. Prediction of postoperative disease-free survival and brain metastasis for HER2-positive breast cancer patients treated with neoadjuvant chemotherapy plus trastuzumab using a machine learning algorithm. Breast Cancer Res Treat 2018;172:611-8.

23. Witzel I, Oliveira-Ferrer L, Pantel K, Müller V, Wikman H. Breast cancer brain metastases: Biology and new clinical perspectives. Breast Cancer Res 2016;18:8.

24. Blumenthal DT, Artzi M, Liberman G, Bokstein F, Aizenstein O, Ben Bashat D. Classification of high-grade glioma into tumor and Nontumor components using support vector machine. AJNR Am J Neuroradiol 2017;38:908-14.

25. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Mach Intelligence 2019;1:206-15.

26. Bathaee Y. The artificial intelligence black box and the failure of intent and causation. Harv JL Tech 2017;31:889.

27. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. InEuropean conference on computer vision 2014 Sep 6 (pp. 818-833). Springer, Cham.

28. AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. Med Phys 2018;45:1150-8.

29. Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. The British journal of radiology. 2019 Aug;92(1100):20190001.