WEP Warwick Evans Publishing
Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Prediction of Ischemic Stroke in MRI Images based on Machine Learning: A Systematic Review

Ziwei He*

Vanke School of Public Health, Tsinghua University, Beijing, China

*Corresponding author: hezw21@mails.tsinghua.edu.cn

Abstract. Ischemic stroke is a dangerous disease that endangers human health and is one of the important causes of disability and death. Early detection and prediction of ischemic stroke are essential for timely intervention and prevention of serious consequences. Fortunately, deep learning-based medical image research has made great progress, which can assist doctors in diagnosing conditions. This paper analyzes the overall effectiveness of existing machine learning (ML) methods in predicting final infarction from baseline imaging. First, this paper searched 11 relevant studies and analyzed the impact of these methods on patients with acute ischemic stroke (AIS). Then, the performance, effect and characteristics of the model are analyzed. Although these methods are generally good at predicting infarction, they show great potential. Based on machine learning methods, reliance on abundant data is essential for effective learning. Insufficient data can lead to significant errors in model predictions. Therefore, future research should focus on collecting ample clinical data for model training to enhance both accuracy and robustness.

Keywords: Ischemic Stroke; deep learning; MRI.

1. Introduction

Despite being the third most common cause of death and disability globally, stroke's global disease burden has significantly declined in recent years due to advances in acute stroke care [1]. The only therapy that can effectively restore blood flow and stop brain function decline is recanalization of the blocked arteries. In terms of global deaths in 2019, stroke remains the second leading cause of death and the third leading cause of disability [2]. Similar risk factors exist for coronary heart disease and other vascular disorders as well as for stroke. Targeting the three main modifiable factors—diabetes, high cholesterol, and hypertension—is an effective preventative strategy. It is also possible to address the risks associated with certain lifestyle factors, such as smoking, inactivity, poor food, and abdominal obesity [3].

Understanding the composition, location, and amount of thrombus in acute ischemic stroke has been a central area of research for the past 30 years as therapeutics have been developed [4]. This field of investigation is motivated by the idea that if management is geared toward particular thrombi features, improved patient selection, and increased treatment success can be attained [5]. Research and interest in this area have increased again due to advances in imaging, novel therapies, and the importance of treatment time. Gaining a better understanding of thrombus composition can help identify therapy targets and gain insight into the pathophysiology and causation of ischemic stroke [6]. These scoring models were divided into three types of assessment outcomes, including low, medium and high risk, based on the severity and etiology of the patient. This has led to the development of secondary prevention options for stroke, such as single antiplatelet therapy. Patients who have had an ischemic stroke are far less likely to experience another one when using these medications. The prompt identification of modifiable risk variables is necessary for effective secondary prevention. However, every scoring tool has limitations due to methodology; incorporating too many risk indicators might make them difficult to utilize in clinical settings. However, combining too few can lead to the loss of essential data and reduce the reliability of risk assessment. In risk rating models, this sparks debate.

Machine learning, a subfield of artificial intelligence, aims to find intricate patterns in data that can be utilized for sophisticated exploratory data analysis, advanced prediction, or classification of previously undiscovered data [7]. In radiology, photographs are essentially large amounts of data that can be mined, making it a perfect topic for supervised AI methods. AI has drawn much attention from the medical profession as a field that uses techniques to identify trends by labeling "ground truth" data [8]. These encouraging findings encouraged researchers to develop new model approaches that they applied in various clinical settings, combined multi-modality input parameters, and improved algorithm architectures. Because machine learning can identify extremely complicated patterns with relatively few assumptions about the data, it is particularly well-suited to analyze such multi-dimensional and multi-parametric data. Machine learning can benefit the medical field, especially in precision medicine. Precision medicine uses deep phenotype and genotype data and lifestyle and environmental data to create personalized health models intended to improve disease prevention, diagnosis, and treatment despite significant advances made in the last few decades. Consequently, tailored treatment plans must be created based on each patient's risk characteristics and etiological classification to achieve more successful secondary prevention when mining and screening critical risk characteristics. Machine learning (ML) algorithms have notably improved over conventional techniques in recent years. Complex nonlinear data associations and enormous volumes of clinical data are no problem for machine learning algorithms. This technique addresses the drawbacks of conventional linear score prediction models. With thrombolytic therapy and vascular revascularization, it can precisely identify patients who are at risk of functional impairments and recurring occurrences. This allows for more precise secondary prevention strategy creation and tailored risk factor analysis.

While the number of studies on this subject is increasing, few examine the general uses of the most recent ML-based techniques for ischemia core estimate. This paper conducted a systematic review to assess the overall effectiveness of current approaches, offer recommendations for future research that could help in the management of acute ischemic stroke (AIS). In addition, the potential benefits and unresolved issues of ML-based modeling approaches for predicting final infarct lesions from acute stroke imaging are outlined.

2. Literature Retrieval Method

From their creation until December 31, 2023, this paper thoroughly searched the following databases: PubMed and EMBASE. This paper used the following keywords in our search: "machine learning," "deep learning," "stroke," and "magnetic resonance imaging." Among them was research that developed machine learning methods to use baseline acute stroke imaging to predict the final infarct lesion. Inclusion criteria for the meta-analysis study: Firstly, patients with ischemic stroke were included in the study population, and MRI imaging was used to diagnose ischemic stroke and identify identifiable lesions. Secondly, using machine learning to predict CT or MRI imaging in patients with ischemic stroke. Then, the included articles need a clearly defined training and testing set, typically CT or MRI images of patients with ischemic stroke. Finally, the articles included in the analysis should be in English, and articles with better algorithm performance should be selected.

Machine learning-based stroke prediction technique. For example, a core model, a penumbra model, and a perfusion model were the three machine learning models developed in one paper. Using a two-stage training method, two machine learning models were developed to predict penumbra and core, respectively.

A forest of decision trees is constructed using the Random Forests method of medical image segmentation, and each tree is trained using a random sample of the image's voxels. Practically, the neuroradiologist and stroke neurologist visually analyze imaging data and make therapy decisions. The "penumbra," or essential surrounding tissue that may be salvaged is identified by magnetic resonance imaging (MRI) using diffusion-weighted imaging and hypo-perfused images to determine the location of the infarct core. The goal of acute stroke treatment options is to revascularize blocked vessels as soon as possible and sustainably, ideally with intravenous thrombolysis or thrombectomy performed in specialized facilities. Making quick conclusions based on more thorough information that visual analysis would have overlooked is made possible by the use of automatic techniques for medical image analysis.

They were using a random forest classifier to forecast each voxel's risk of myocardial infarction. The RF classifier connects many CT characteristics and temporal factors of individual voxels to a statistical infarction probability model at the voxel level. Precisely, the cerebral blood volume and CT average maps were smoothed using median filtering, and the Z-score method was applied to normalize the outcomes. Additionally, voxel-level CT values and patient-specific temporal parameters were used, such as the time delay between the beginning of stroke symptoms and imaging. A voxel-wise machine learning model was trained using the random forest technique, employing the CT parameters for each voxel for that particular patient as features. The follow-up infarct's hand segmentations were utilized as responses (1: infarct, 0: normal tissue). In the random forest model, 80 trees were used, bootstrapping was used, and the standard and infarct tissue class weights were assigned automatically based on the actual sample distributions. The trees' depth was fixed at ten. The trained machine learning model can generate infarct probabilities for each voxel in a new subject image and classify each voxel as infarct or not (with a cutoff of 0.5). The Pearson correlation coefficient investigated the association between the follow-up infarct volume and the estimated infarct volume from the ML model. In addition, sensitivity analyses regarding volumetric differences were performed by classifying patients into three groups based on the interval between CT imaging and reperfusion.

3. Results

Ten papers were included in the meta-analysis. All three training, validation, and test image sets were distinctly characterized in three investigations [9-11]. Model performance was assessed using an external test set in just one study. In terms of data sources, one study obtained image data from four companies [12], five studies used data from a single company, and the other studies did not provide any information. Each study described the pre-processing step for the data, used appropriate reference standards to train their models, and showed the prediction performance as determined by various performance indicators. Partially public algorithms from two papers were made accessible on the GitHub website.

The DL-based models performed better than the ML-based models, with the most outstanding results from combining DL methods and CT data pre-processing. Current machine learning techniques demonstrated good performance for ischemia tissue outcome prediction from baseline imaging despite the substantial heterogeneity across studies. Because infarct lesions vary widely in shape, location, and development, predicting eventual disease progression from baseline imaging can be challenging. When it appears that traditional thresholding techniques are insufficient, machine learning applications aim to extract as much predictive power as possible from the given multi-modality imaging data. A few studies demonstrated a considerable increase in measurement findings when employing ML-based methods verifying their suggested approaches against traditional thresholds for core infarcted tissue delineation. These tactics use machine learning algorithms' capacity to process data to deliver quick and accurate assessments, showing promise for assisting clinical management.

One of its most vital points is the independence of the created machine learning technique from deconvolution algorithms. While deconvolution techniques can accurately simulate perfusion status, their computing complexity is significantly increased when physiological differences in contrast supply via the arteries, collateral flow effects, and venous outflow components of cerebral perfusion are included. Two standard machine learning algorithms may find it difficult to process such massive data points. For the kind of picture analysis being tried here, convolutional neural networks (CNNs), which are deep learning approaches, are also potential tools. However, there are specific difficulties in implementing CNNs. These include constraints when working with short training datasets that lack diversity or imbalanced samples (small objects [infarct region] vs. backdrop, for example).

Despite promising to predict eventual infarct lesions, machine learning-based methods are still not widely adopted and used in clinical practice. Due to the difficulty of medical image acquisition, these methods are trained on fewer samples and therefore have large errors. The k-fold cross-validation approach was widely used in validation studies to offer an objective assessment with small sample sizes. However, without third-party data to test, the actual predicted performance is overstated.

4. Conclusion

This paper used machine learning prediction and performed a meta-analysis and systematic review of recent publications. To classify and predict outcomes for people and communities, multi-omics data must be analyzed and processed using machine learning techniques, which include the recently developed deep learning models. Despite the differences in the study, subsequent investigations are advised to focus on machine learning training. Any clinically relevant information, such as the topography of the stroke, its severity, the vascular supply of the hypo-perfused area, and other prognostic markers, should be recognized and included as modeling parameters.

References

[1] Brinjikji, W., Duffy, S., Burrows, A., Hacke, W., Liebeskind, D., Majoie, C. B. L. M., Dippel, D. W. J., Siddiqui, A. H., Khatri, P., Baxter, B., Nogueira, R., Gounis, M., Jovin, T., & Kallmes, D. F. Correlation of imaging and histopathology of thrombi in acute ischemic stroke with etiology and outcome: a systematic review. Journal of neurointerventional surgery, 2017, 9(6), 529–534.

[2] Maegerlein, C., Friedrich, B., Berndt, M., Lucia, K. E., Schirmer, L., Poppert, H., Zimmer, C., Pelisek, J., Boeckh-Behrens, T., & Kaesmacher, J. Impact of histological thrombus composition on preinterventional thrombus migration in patients with acute occlusions of the middle cerebral artery. Interventional neuroradiology : journal of peritherapeutic neuroradiology, surgical procedures and related neurosciences, 2018, 24(1), 70–75.

[3] Simons, N., Mitchell, P., Dowling, R., Gonzales, M., & Yan, B. Thrombus composition in acute ischemic stroke: a histopathological study of thrombus extracted by endovascular retrieval. Journal of neuroradiology = Journal de neuroradiologie, 2015, 42(2), 86–92.

[4] Seymour, C. W., Kennedy, J. N., Wang, S., Chang, C. H., Elliott, C. F., Xu, Z., Berry, S., Clermont, G., Cooper, G., Gomez, H., Huang, D. T., Kellum, J. A., Mi, Q., Opal, S. M., Talisa, V., van der Poll, T., Visweswaran, S., Vodovotz, Y., Weiss, J. C., Yealy, D. M., … Angus, D. C. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis, 2019, 321(20), 2003–2017.

[5] Debs N, Decroocq M, Cho T H, et al. Evaluation of the realism of an MRI simulator for stroke lesion prediction using convolutional neural network, Simulation and Synthesis in Medical Imaging: 4th International Workshop, SASHIMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 4. Springer International Publishing, 2019: 151-160.

[6] Zhou, J., Lee, S., Liu, Y., Chan, J. S. K., Li, G., Wong, W. T., Jeevaratnam, K., Cheng, S. H., Liu, T., Tse, G., & Zhang, Q. Predicting Stroke and Mortality in Mitral Regurgitation: A Machine Learning Approach. Current problems in cardiology, 2023, 48(2), 101464.

[7] Kumar, S., Marchina, S., Massaro, J., Feng, W., Lahoti, S., Selim, M., & Herzig, S. J. ACDD4 score: A simple tool for assessing risk of pneumonia after stroke. Journal of the neurological sciences, 2017, 372, 399–402.

[8] Heo, T., Kim, Y.S., Choi, J.M., Jeong, Y.S., Seo, S.Y., Lee, J.H., Jeon, J.P., & Kim, C. Prediction of Stroke Outcome Using Natural Language Processing-Based Machine Learning of Radiology Report of Brain MRI. Journal of Personalized Medicine, 2020, 10.

[9] Li, M., Lang, M., Deng, F., Chang, K., Buch, K.A., Rincon, S., Mehan, W.A., Leslie-Mazwi, T.M., & Kalpathy-Cramer, J. Analysis of Stroke Detection during the COVID-19 Pandemic Using Natural Language Processing of Radiology Reports. American Journal of Neuroradiology, 2020, 42, 429 - 434.

[10] Ong, C.J., Orfanoudaki, A., Zhang, R., Caprasse, F.P., Hutch, M.R., Ma, L., Fard, D., Balogun, O., Miller, M.I., Minnig, M.A., Saglam, H., Prescott, B., Greer, D.M., Smirnakis, S.M., & Bertsimas, D. Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports. PLoS ONE, 2020, 15.

[11] Kim, C., Zhu, V.J., Obeid, J.S., & Lenert, L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. PLoS ONE, 2019, 14.

[12] Garg, R., Oh, E., Naidech, A.M., Kording, K.P., & Prabhakaran, S. Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association, 2019, 28(7), 2045-2051.