A survey on common biostatistics tools in neuroscience: Machine learning and Bayesian modeling

Ziyi Xue
University of Illinois at Urbana- Champaign, Illinois,61820,USA.

Abstract. Machine learning was characterized by building models and finding correlations between data features, while logistic regression, decision trees, support vector machines (SVM), random forest (RF) and neural networks were recognized as common ML approaches. Bayesian modeling model uncertainty, which can estimate the features from the dataset directly instead of from sampling distribution. Their roles were extremely useful for the detection and progression for diseases in neuroscience. This review summarize different approaches in various diseases, hoping to introduce the potential roles of biostatistics tools in neuroscience.

Keywords: Neuroscience, Biostatistics, Bayesian Modelingy.

1. Introduction

Within the years pasted, the methodologies of biostatistics tools have been used to improve the detection and outcomes of neuroscience. Machine learning (ML) is characterized by using the tools from computer sciences, building models from large datasets, and finding correlations between data features. Based on input-output labels, the most common ML approach, supervised machine learning methods can be used for making decisions or predictions. Classification models are evaluated by F1-score, accuracy, precision, and recall, while regression models are evaluated by mean squared error (MSE) and R-squared. Logistic regression, decision trees, support vector machines (SVM), random forest (RF) and neural networks are the common approaches for the classification models, while linear regression, Lasso regression, ridge regression, decision trees (for regression), and neural networks (for regression) are characterized as useful ways for the regression model.

Bayesian modelling (BM) has been considered as an approach to model uncertainty, which can estimate the features from the dataset directly instead of from sampling distribution. Describing probability distributions, BM treat the features as random variables, which can represent the dependency relationships between variables and handle the uncertainty and randomness between different variables.

The tools of ML and BM have been used in the field of neuroscience, including Alzheimer's disease (AD)(1), Parkinson's disease(PD)(2), cerebrovascular diseases(3), and epilepsy(4). The application of BM can also provide useful approach to the diagnosis and progression of diseases in neuroscience. Thus, the effectiveness of biostatistics tools, ML and BM for instance, to improve the accuracy of detection for diseases in neuroscience will be examined, as well as the optimal choices for clinical decision making and patient outcome.

2. AD

Mild cognitive impairment (MCI) has characterized as the prodromal stage of AD dementia. The deposition of amyloid β (Aβ) is regarded as the highly risk of conversion of MCI to AD dementia, however, the deposition of pathological biomarker has mainly based on the positron emission tomography (PET) and cerebrospinal fluid (CSF), which is low availability due to high costs, safety issues, and invasive approaches. Therefore, by integrating and analyzing large amounts of data from different resources, including clinical features, and neuroimaging,

ML can help improve the detection of MCI and the progression of MCI to AD. Recent study found that model(including age, gender, education, diabetes, hypertension, and apolipoprotein E genotype, and neuropsychological test) showed good accuracy(area under operating characteristic cure[AUC]

0.837) in cross validation, and fair accuracy(AUC 0.765) in external validation using RF approach for the detection of amyloid positivity in MCI (5). Logistic regression was useful to predict the existence of pathological biomarkers from clinical risk factors and MRI variables, while the good accuracy was shown from amyloid (AUC 0.79), tau(AUC 0.73) and neurodegeneration(AUC 0.86)(6).

ML also exhibit good accuracy for the prediction of the progression of cognitive impairment. Logistic regression also found that model(including age, instrumental activities of daily living, marital status, and baseline cognitive function) showed fair accuracy (AUC 0.814) to predict cognitive impairment for 3-year by logistic regression(7).Another study also focused on other AD-specific antibody and out-performed accuracy was identified by SVM(AUC 0.94)(8).

Voice signal was also identified as an automated approach for the detection and progression of MCI and AD dementia, while ML was helpful to confirm their potential diagnostic value. The change of prosody was useful for the detection of MCI while principal component analysis was also used to combine the features with the similar character. The discrimination accuracy was 0.90 for training data and 0.65 for validation data measuring by logistic regression(9). The acoustic variables were also able to predict the amyloid status by various ML approaches (AUC 0.79), outperformed the model based on conventional neuropsychological test (AUC 0.66)(10). The joint fusion of acoustic and linguistic parameters were able to achieve AUC with 92.53 for training dataset and 93.89 for test dataset, indicating that speech-processing with able to act as an screening tools for patients with cognitive impairment(11).

BM was also recognized as potential useful approaches for the dection and progression of MCI and AD dementia. Recent study also found that participants with high blood pressure are more likely to exhibit high-risk of pathological biomarkers(12). Another study also found that glial activation of glial activation measuring by baseline PET imaging is a good way to predict the progression of AD(13). The effects of memory and executive performance in amyloid negative participants were also identified by BM(14).

3. PD

ML and BM were also useful for the detection and prediction of PD. Various studies focused on the potential diagnostic value of voice signal using ML approaches. Neural network showed that speech data was helpful to differentiate PD patients from other neurodegenerative diseases with the AUC of 0.93(15). Both ML and deep learning were useful for the dection of PD patients from voice signal(AUC 0.74 and 0.80 respectively)(16). The variables from voice and facial signals showed the good accuracy of the detection for PD using muliple ML approaches, while the RF and logic regression approaches were recognized as the best classifiers(AUC 0.85 and 0.84)(17).

The wearable device was also able to help the dection of frezzing-of-gait from PD patients, while the ML showed a good acurracy measuring by SVM(sensitivity 83%,specificity 80%)(18). Another study also showed that the kinematic data from wearable device was able to detect the subtype from PD with the F1 score of 79.6%(19). RF and decision trees were also helpful from the prediction of the contraction pattern of tremor from PD patients with the AUC of 0.92(20).MLwas trained using accelerometry data to achieve best test performance to distinuguish clinically-diagnosed PD, while the results was outperformed compared with genetics, lifestyle and blood biochemistry(21).The fall risk can be predicted by gait parameters deviced from the results in the real-word(21).

BM was also recognized as a useful approach for the ML approaches in the detection and prediction of PD. Degeneration of posterior basal forebrain was associated with cognitive impairment of PD measuring by BM(22). Another study also focused on the roles of locus coeruleus by 7T MRI and the results showed the the vatiablity of response inhibition significantly associated with the variability of locus coerulus integrity measruing by BM(23).BM approaches also used to predict the progression of clinical symptom, while higher systolic blood pressure variablity was recognized as a predictor for the progression of cognitive impairment(24). Measuring by BM, another study also found that the change of executive function was correlated with the change of walking performance and the effect of early motor rehabiliation was also limited by the dysfunction of execution and attention(25).

4. Cerebrovascular disease

Multiple ML approaches were used to evuluate the prediction of the risk of post-stroke depression and anxiety. The results showed that several factors, stroke incidence and history of antidepressants for instance, were able to predict the risk of post-stroke depression and anxiety. Another study also found that clinical risk factor, including diabetes, hypertension, and hyperlipidemia, are able to predict large-artery atherosclerosis with AUC value of 0.93 measuring by logic regression(26). There also exist some studies focusing on the prediction of post-stroke cognitive impairment. Cortical infarcts, atrophy of medial lobe, the severity of initial stroke and the history of stroke were recognized as the predictors for cognitive impairment after stroke with the AUC value of 0.792 by neural network(27). Multiple ML approaches were also helpful to identify the onset time of wake-up stroke from MRI, while the best AUC was 0.895 from SVM(28).

ML was also used to predict the progression of intracerebral hemorrhage after intravenous thrombolysis for stroke. The clinical factors and history of treatment were used as the variables to develop models while the highest AUC was 0.87 for logistic regression(29). The level of consciousness, vital sign, sudden headache, and speech abnormalities were the useful approach to accurately predict the need for surgery with the AUC of 0.802 by extreme Gradient Boosting(30). Combined with clinical and laboratory factors, ten features were able to predict the possibility of death after hemorrhagic transformation with the AUC of 0.85(31). Other study also focused on the potential variables in CT to predict the expansion of perihematomal edema expansion for patients with intracranial hemorrhage with the combined AUC of 0.840(32).

BM was also useful in the prediction and detection of cerebrovascular disease. Glycemic gap, measuring the acute glycemic exxcursion, was helpful to predict the risk of stroke recurrence by bayesian logistic regression(33). BM was also helpful to estimate the prevalence of other symptom. The prevalence of apraxia of speech was evuluate for chronic aphasia after stroke by BM model(34). For the prediction of the adverse events from drug usage, BM was also recognized as a helpful approach. Several clinical risk factors, sex, anaemia included, were recognized as a model to predict the major bleeding event from the useage of novel oral atrial anticoagulants with atrial fibrillation participants(35).

5. epilepsy

ML approaches were also used for the detection and progression of epilepsy. Deep learning was able to diagnose temporal lobe epilepsy with high accurate based on T2-weighted imaging(T2WI) and fluid-attenuated inversion recovery(FLAIR)(36). Other study also found that centrotemporal spike-waves was able to detect epilepsy with a extremely high level of sensivity(99.8%) and specificity(98.4%) by neural network(37).Other study also focuses on the potential roles of using the electroencephalography to predict impaired consciousness and the best classifier achieved 100% predictive value(38).Neural network was also used to identify focal cortical dysplasias with the sensivity of 85%(39).

6. Conclusion

Recently, various studies focused on the potential roles of ML, BM in particular, for the detection and progression of neurodegenerative diseases. ML was able to identify various model to predict the progression of MCI to AD dementia, and to detect the stage of MCI using other automated device, voice signal for instance. Wearable device and voice signal were also able for the detection of PD and their subtype. For cerebrovascular diseases, ML and BM were helpful to construct model to predict the risk of progression. ML were also able to develop model to detect special type of epilepsy and predict impaired consciousness.

References

[1] Seixas FL, Zadrozny B, Laks J, Conci A, Muchaluat Saade DC. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Comput Biol Med. 2014;51:140-58.

[2] Landolfi A, Ricciardi C, Donisi L, Cesarelli G, Troisi J, Vitale C, et al. Machine Learning Approaches in Parkinson's Disease. Curr Med Chem. 2021;28(32):6548-68.

[3] Boyd C, Brown G, Kleinig T, Dawson J, McDonnell MD, Jenkinson M, et al. Machine Learning Quantitation of Cardiovascular and Cerebrovascular Disease: A Systematic Review of Clinical Applications. Diagnostics (Basel). 2021;11(3).

[4] Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. Epilepsia. 2019;60(10):2037-47.

[5] Kang SH, Cheon BK, Kim JS, Jang H, Kim HJ, Park KW, et al. Machine Learning for the Prediction of Amyloid Positivity in Amnestic Mild Cognitive Impairment. J Alzheimers Dis. 2021;80(1):143-57.

[6] Lew CO, Zhou L, Mazurowski MA, Doraiswamy PM, Petrella JR, Alzheimer's Disease Neuroimaging I. MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum. Radiology. 2023;309(1):e222441.

[7] Hu M, Shu X, Yu G, Wu X, Valimaki M, Feng H. A Risk Prediction Model Based on Machine Learning for Cognitive Impairment Among Chinese Community-Dwelling Elderly People With Normal Cognition: Development and Validation Study. J Med Internet Res. 2021;23(2):e20298.

[8] Fang L, Jiao B, Liu X, Wang Z, Yuan P, Zhou H, et al. Specific serum autoantibodies predict the development and progression of Alzheimer's disease with high accuracy. Brain Behav Immun. 2023;115:543-54.

[9] Higuchi M, Nakamura M, Omiya Y, Tokuno S. Discrimination of mild cognitive impairment based on involuntary changes caused in voice elements. Front Neurol. 2023;14:1197840.

[10] Garcia-Gutierrez F, Marquie M, Munoz N, Alegret M, Cano A, de Rojas I, et al. Harnessing acoustic speech parameters to decipher amyloid status in individuals with mild cognitive impairment. Front Neurosci. 2023;17:1221401.

[11] Zolnoori M, Zolnour A, Topaz M. ADscreen: A speech processing-based screening system for automatic identification of patients with Alzheimer's disease and related dementia. Artif Intell Med. 2023;143:102624.

[12] Sible IJ, Nation DA, Alzheimer's Disease Neuroimaging I. Visit-to-Visit Blood Pressure Variability and CSF Alzheimer Disease Biomarkers in Cognitively Unimpaired and Mildly Impaired Older Adults. Neurology. 2022;98(24):e2446-e53.

[13] Yasuno F, Kimura Y, Ogata A, Ikenuma H, Abe J, Minami H, et al. Neuroimaging biomarkers of glial activation for predicting the annual cognitive function decline in patients with Alzheimer's disease. Brain Behav Immun. 2023;114:214-20.

[14] Teipel SJ, Dyrba M, Levin F, Altenstein S, Berger M, Beyle A, et al. Cognitive Trajectories in Preclinical and Prodromal Alzheimer's Disease Related to Amyloid Status and Brain Atrophy: A Bayesian Approach. J Alzheimers Dis Rep. 2023;7(1):1055-76.

[15] Eguchi K, Yaguchi H, Kudo I, Kimura I, Nabekura T, Kumagai R, et al. Differentiation of speech in Parkinson's disease and spinocerebellar degeneration using deep neural networks. J Neurol. 2023.

[16] Costantini G, Cesarini V, Di Leo P, Amato F, Suppa A, Asci F, et al. Artificial Intelligence-Based Voice Assessment of Patients with Parkinson's Disease Off and On Treatment: Machine vs. Deep-Learning Comparison. Sensors (Basel). 2023;23(4).

[17] Lim WS, Chiu SI, Wu MC, Tsai SF, Wang PH, Lin KP, et al. An integrated biometric voice and facial features for early detection of Parkinson's disease. NPJ Parkinsons Dis. 2022;8(1):145.

[18] Li D, Hallack A, Gwilym S, Li D, Hu MT, Cantley J. Investigating gait-responsive somatosensory cueing from a wearable device to improve walking in Parkinson's disease. Biomed Eng Online. 2023;22(1):108.

[19] Gong NJ, Clifford GD, Esper CD, Factor SA, McKay JL, Kwon H. Classifying Tremor Dominant and Postural Instability and Gait Difficulty Subtypes of Parkinson's Disease from Full-Body Kinematics. Sensors (Basel). 2023;23(19).

[20] Vescio B, De Maria M, Crasa M, Nistico R, Calomino C, Aracri F, et al. Development of a New Wearable Device for the Characterization of Hand Tremor. Bioengineering (Basel). 2023;10(9).

[21] Schalkamp AK, Peall KJ, Harrison NA, Sandor C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. Nat Med. 2023;29(8):2048-56.

[22] Schumacher J, Kanel P, Dyrba M, Storch A, Bohnen NI, Teipel S, et al. Structural and molecular cholinergic imaging markers of cognitive decline in Parkinson's disease. Brain. 2023.

[23] Ye R, Hezemans FH, O'Callaghan C, Tsvetanov KA, Rua C, Jones PS, et al. Locus Coeruleus Integrity Is Linked to Response Inhibition Deficits in Parkinson's Disease and Progressive Supranuclear Palsy. J Neurosci. 2023;43(42):7028-40.

[24] Xiao Y, Yang T, Zhang L, Wei Q, Ou R, Hou Y, et al. Association between the blood pressure variability and cognitive decline in Parkinson's disease. Brain Behav. 2023:e3319.

[25] Geritz J, Welzel J, Hansen C, Maetzler C, Hobert MA, Elshehabi M, et al. Cognitive parameters can predict change of walking performance in advanced Parkinson's disease - Chances and limits of early rehabilitation. Front Aging Neurosci. 2022;14:1070093.

[26] Sun TH, Wang CC, Wu YL, Hsu KC, Lee TH. Machine learning approaches for biomarker discovery to predict large-artery atherosclerosis. Sci Rep. 2023;13(1):15139.

[27] Lee M, Yeo NY, Ahn HJ, Lim JS, Kim Y, Lee SH, et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. Alzheimers Res Ther. 2023;15(1):147.

[28] Jiang L, Wang S, Ai Z, Shen T, Zhang H, Duan S, et al. Development and external validation of a stability machine learning model to identify wake-up stroke onset time from MRI. Eur Radiol. 2022;32(6):3661-9.

[29] Wen R, Wang M, Bian W, Zhu H, Xiao Y, He Q, et al. Machine learning-based prediction of symptomatic intracerebral hemorrhage after intravenous thrombolysis for stroke: a large multicenter study. Front Neurol. 2023;14:1247492.

[30] Yoshida Y, Hayashi Y, Shimada T, Hattori N, Tomita K, Miura RE, et al. Prehospital stroke-scale machine-learning model predicts the need for surgical intervention. Sci Rep. 2023;13(1):9135.

[31] Li X, Xu C, Shang C, Wang Y, Xu J, Zhou Q. Machine learning predicts the risk of hemorrhagic transformation of acute cerebral infarction and in-hospital death. Comput Methods Programs Biomed. 2023;237:107582.

[32] Chen Y, Qin C, Chang J, Lyu Y, Zhang Q, Ye Z, et al. A machine learning approach for predicting perihematomal edema expansion in patients with intracerebral hemorrhage. Eur Radiol. 2023;33(6):4052-62.

[33] Yuan K, Xie M, Hou H, Chen J, Zhu X, Wang H, et al. Association of glycemic gap with stroke recurrence in patients with ischemic stroke. J Diabetes. 2023;15(9):714-23.

[34] Ziegler W, Aichert I, Staiger A, Willmes K, Baumgaertner A, Grewe T, et al. The prevalence of apraxia of speech in chronic aphasia after stroke: A bayesian hierarchical analysis. Cortex. 2022;151:15-29.

[35] Barnett-Griness O, Stein N, Kotler A, Saliba W, Gronich N. Novel bleeding prediction model in atrial fibrillation patients on new oral anticoagulants. Heart. 2022;108(4):266-73.

[36] Sakashita K, Akiyama Y, Hirano T, Sasagawa A, Arihara M, Kuribara T, et al. Deep learning for the diagnosis of mesial temporal lobe epilepsy. PLoS One. 2023;18(2):e0282082.

[37] Jeon Y, Chung YG, Joo T, Kim H, Hwang H, Kim KJ. Deep Learning-Based Detection of Epileptiform Discharges for Self-Limited Epilepsy With Centrotemporal Spikes. IEEE Trans Neural Syst Rehabil Eng. 2022;30:2939-49.

[38] Springer M, Khalaf A, Vincent P, Ryu JH, Abukhadra Y, Beniczky S, et al. A machine-learning approach for predicting impaired consciousness in absence epilepsy. Ann Clin Transl Neurol. 2022;9(10):1538-50.

[39] Spitzer H, Ripart M, Whitaker K, D'Arco F, Mankad K, Chen AA, et al. Interpretable surface-based detection of focal cortical dysplasias: a Multi-centre Epilepsy Lesion Detection study. Brain. 2022;145(11):3859-71.