International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Beyond Chronological Age: A Comprehensive Assessment of Machine Learning Predictors in Neuroimaging

Dr T BHARATH KRISHNA[1], M VINOD KUMAR[2], J KAVITHA[3] Y NANCHARI[4]

[1]Associate Professor, Department of Computer Science and Engineering, Anu Bose Institute of Technology for women's, KSP Road, New Paloncha, Bhadradri Kothagudem District, Telangana (TS), 507115

[2, 3, 4]Assistant Professor, Department of Computer Science and Engineering, Anu Bose Institute of Technology for women's, KSP Road, New Paloncha, Bhadradri Kothagudem District, Telangana (TS), 507115

ABSTRACT

Machine learning (ML) algorithms are a fundamental component of brain age estimation frameworks, yet the extent to which regression algorithm choice affects prediction accuracy has not been fully investigated. This study assessed the performance of different regression algorithms for brain age estimation. We developed a framework trained on a large cohort of cognitively healthy (CH) individuals (N = 788) and evaluated 22 regression algorithms. Each algorithm was tested on independent samples comprising 88 CH individuals, 70 patients with mild cognitive impairment, and 30 patients with Alzheimer's disease. In the independent CH test set, mean absolute error (MAE) ranged from 4:63 to 7:14 years, while R2 ranged from 0:76 to 0:88. The Quadratic Support Vector Regression algorithm achieved the highest accuracy (MAE = 4:63 years, R2 = 0:88; 95% CI = [□1:26; 1:42]), whereas the Binary Decision Tree algorithm had the lowest (MAE = 7:14 years, R2 = 0:76; 95% CI = [□1:50; 2:62]). These findings indicate that regression algorithm selection affects prediction accuracy in brain age estimation and suggest that advanced ML methods may improve brain age predictions in clinical settings.

1. INTRODUCTION

Brain age-delta is an inheritable measure used to monitor cognitive health across aging and support the diagnosis of neurological disorders and comorbidities. It reflects the difference between a person's chronological age and the age estimated by machine-learning models trained on brain imaging data. Aging brings structural changes to the brain at scales ranging from molecular to morphological. A brain age-delta of zero is consistent with a healthy aging trajectory, whereas a marked delta may indicate accelerated cognitive aging and a greater risk of age-related neurological disease or atypical brain changes. Neurological conditions such as Alzheimer's disease (AD), Parkinson's disease, epilepsy, and schizophrenia have been investigated using brain-age estimation methods. These studies draw on imaging modalities including anatomical MRI, functional MRI, fluorodeoxyglucose positron emission tomography, and diffusion tensor imaging. Following preprocessing, the images are used to extract features. Anatomical MRI is especially popular because it is widely available and provides high spatial resolution and strong tissue contrast. If the number of extracted features exceeds the sample size, methods such as principal component analysis (PCA) can help address dimensionality challenges. In the prediction stage, supervised

learning methods, including regression algorithms, estimate brain age from the input data. Choosing an appropriate regression algorithm is essential for achieving accurate predictions in clinical settings. Common choices include Gaussian process regression and support vector regression. Several factors should be considered when selecting an algorithm for brain-age estimation: Accuracy and sensitivity to individual data points: Models are generally assessed using the Mean Absolute Error (MAE) between predicted and chronological age. Capacity to capture naturally occurring variation, including variation associated with genetic factors: Genetic influences play a significant

1.1 Purpose

This study aims to evaluate how effectively machine learning algorithms can predict brain age. By examining the difference between a person's chronological age and their predicted brain age, researchers seek to assess the brain's aging process. This investigation is intended to deepen understanding of neurological aging patterns and may help identify early signs of cognitive decline or neurological disorders.

1.2 Scope

The scope of this research encompasses a comprehensive evaluation of various machine learning algorithms applied to neuroimaging data. Techniques such as structural MRI, functional MRI, and diffusion tensor imaging (DTI) may be utilized to capture different aspects of brain structure and function. The study may explore the impact of different preprocessing methods, feature extraction techniques, and model architectures on the accuracy of brain age prediction.

1.3 Need For System

Advancing Neuroscience Research: Understanding brain aging is crucial for advancing neuroscience research and developing interventions for age-related cognitive decline and neurological disorders. By accurately predicting brain age, researchers can identify deviations from typical aging trajectories and explore underlying biological mechanisms.

Clinical Applications: Accurate prediction of brain age has significant clinical implications. It can assist healthcare professionals in diagnosing neurological disorders such as Alzheimer's disease, dementia, and Parkinson's disease at early stages when interventions may be more effective. Additionally, it can help in personalized medicine by guiding treatment strategies tailored to an individual's brain age profile.

Personalized Healthcare: Predicting brain age using machine learning algorithms enables personalized healthcare approaches. By assessing an individual's brain age relative to their chronological age, clinicians can identify individuals at higher risk of cognitive decline or neurological disorders and provide targeted interventions or lifestyle modifications to promote brain health and delay the onset of age-related cognitive decline.

Improving Quality of Life: Early detection of age-related neurological changes can lead to interventions that improve the quality of life for individuals as they age. By identifying risk factors and implementing preventive measures, such as cognitive training, physical exercise, and dietary interventions, individuals can maintain cognitive function and independence for longer periods, enhancing their overall well-being and quality of life.

Public Health Implications: Understanding the factors influencing brain aging and accurately predicting brain age can have broader public health implications.

It can inform public health policies aimed at promoting brain health across populations, including educational initiatives, community-based interventions, and healthcare resource allocation strategies tailored to address the needs of aging populations.

In conclusion, the proposed study on predicting brain age using machine learning algorithms addresses a critical need in neuroscience research and clinical practice. By evaluating the accuracy of prediction models and exploring their potential applications, this

research can contribute to advancements in understanding brain aging, improving clinical outcomes, and promoting brain health across populations.

1.3.1 Existing System

Recent studies have indicated that employing supervised regression methods on MRI brain imaging can accurately predict an individual's brain age, which can be useful for detecting diseases associated with abnormal brain aging. In an existing system, a convolutional neural network, inspired by the VGG architecture, has been developed to predict brain age with precision. This network is trained on healthy grey-matter segmented images and then applied to clinical T1-weighted MRIs.

The model is trained using publicly available healthy datasets and then tested on a clinical dataset comprising patients with Schizophrenia, Parkinson's Disease, and Post-Traumatic Stress Disorder. Initially, bias in brain age prediction is observed, but corrective measures are applied to enhance the reliability of the results. After correcting the bias, the BrainAge model achieves a mean absolute error (MAE) of 4.03 years and an R2 value of 0.96 on the healthy dataset.

To extend its applicability, transfer learning is employed to apply the BrainAge model to the clinical dataset. However, upon comparing the brain age delta (predicted age – chronological age) for each condition, the results do not indicate statistically significant differences ($p <= 0.5$). This suggests that the brain age delta does not reliably signify abnormal brain aging in this particular scenario.

Disadvantages

- The system is not implemented Regression Algorithms in which the user can get less accuracy on datasets.
- The prediction accuracy level in the brain age estimation frameworks is not associated with different items such as feature extraction methods, data reduction strategies, bias correction methods, and regression algorithms.

1.4 Proposed System

- In selecting a regression algorithm for brain age estimation within a prediction model, several crucial considerations should be made:
- Accuracy and Sensitivity: The chosen algorithm must demonstrate high accuracy and sensitivity to various data points within the training dataset. Typically, model performance is evaluated based on metrics such as Mean Absolute Error (MAE), which measures the disparity between predicted age and chronological age.
- Ability to Capture Natural Variations: It's essential that the algorithm can capture the inherent variations present in brain aging, including those influenced by genetic factors. As many aspects of brain aging and susceptibility to age-related brain diseases are believed to be genetically influenced, the model should be adept at learning and accommodating these variations.
- Generalizability: The algorithm should be capable of producing reliable results across diverse datasets and patient groups. This ensures that the model's predictions are robust and applicable beyond the specific dataset on which it was trained, enabling its utility in clinical settings with varying demographics and conditions.
- In summary, the regression algorithm chosen for brain age estimation should prioritize accuracy, sensitivity to data variations, and generalizability across different datasets and patient cohorts to ensure its effectiveness and applicability in clinical applications.

Advance

The k-Nearest Neighbors algorithm is essentially non-parametric classification method, which was later expanded for regression. Under this algorithm, the closest 'k' samples from the dataset are taken with respect to the object under consideration.

Ridge regression is a model tuning method that is used to analyze the data that suffers from multi-collinearity

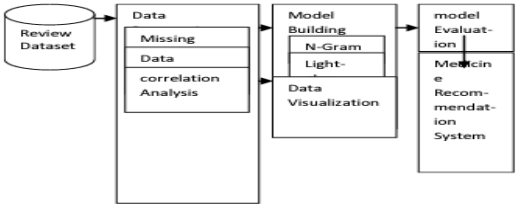


Fig 1: System Architecture

2.ALGORITHMS

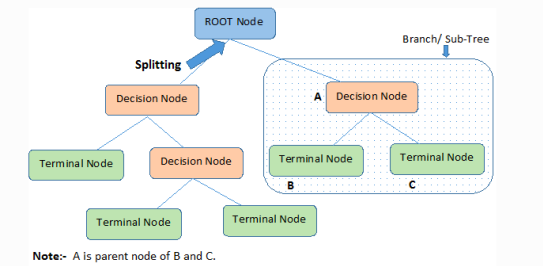
Decision tree classifiers

Decision tree classifiers are widely applied across various domains due to their ability to capture descriptive decision-making knowledge from provided data. One of their most notable features is their capability to be generated from training sets. The process for generating decision trees based on a set of objects (S), each belonging to one of the classes C1, C2, …, Ck, unfolds as follows:

Step 1: If all objects in S belong to the same class (e.g., Ci), the decision tree for S comprises a leaf labeled with this class.

Step 2: Otherwise, select a test (T) with potential outcomes (O1, O2,…, On). Each object in S has a specific outcome for test T, thus partitioning S into subsets (S1, S2,…, Sn), where each object in Si corresponds to outcome Oi for test T. Test T becomes the root of the decision tree, and for each outcome Oi, a subsidiary decision tree is constructed by recursively invoking the same procedure on set Si.

Block Diagram for Decision Tree Algorithm:



Gradient boosting

Gradient boosting is a versatile machine learning technique employed in regression and classification tasks, among others. It constructs a prediction model in the form of an ensemble of weak prediction models, typically decision trees. When using decision trees as the weak learner, the resulting algorithm is referred to as gradient-boosted trees, often surpassing the performance of random forests. The construction of a gradient-boosted trees model occurs in a stage-wise manner, similar to other boosting methods, but it stands out by allowing the optimization of an arbitrary differentiable loss function.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a straightforward yet highly effective classification algorithm that operates based on a similarity measure. It is non-parametric and employs lazy learning, meaning it does not "learn" until presented with a test example. Whenever there is a new data point to classify, KNN identifies its K-nearest neighbors from the training data and determines its classification based on their majority vote or weighted vote.

Logistic regression

Logistic regression analysis explores the relationship between a categorical dependent variable and a set of independent variables. The term "logistic regression" is applied when the dependent variable has only two values, such as 0 and 1, or Yes and No. On the other hand, "multinomial logistic regression" is used when the dependent variable has three or more unique values, like Married, Single, Divorced, or Widowed. While the nature of data for the dependent variable differs from that of multiple regressions, the practical application of the procedure remains similar.

Logistic regression serves as a competitor to discriminant analysis in analyzing categorical-response variables. Many statisticians favor logistic regression due to its versatility and suitability for modeling various situations compared to discriminant analysis. This preference arises from logistic regression's ability to not assume that the independent variables follow a normal distribution, unlike discriminant analysis.

NAIVE BAYES

The naive Bayes approach is a supervised learning method founded on a simple assumption: it presumes that the presence or absence of one feature of a class is

independent of the presence or absence of any other feature.

Despite its simplicity, it demonstrates robustness and efficiency comparable to other supervised learning techniques. One explanation often highlighted in the literature is based on representation bias.

The naive Bayes classifier operates as a linear classifier, akin to linear discriminant analysis, logistic regression, or linear support vector machines (SVMs). However, the distinction lies in the method used to estimate the classifier's parameters, known as the learning bias. Although the naive Bayes classifier finds extensive use in the research community due to its ease of programming, parameter estimation simplicity, rapid learning even with large datasets, and reasonably good accuracy compared to other methods, it remains less popular among practitioners seeking practical results. Researchers appreciate its simplicity and efficacy. However, practitioners often struggle with its interpretability and deployment, as they may not grasp its relevance or utility.

Random forests

Random forests, also known as random decision forests, represent an ensemble learning technique used for classification, regression, and other tasks. They function by constructing numerous decision trees during training. For classification tasks, the output of the random forest is determined by the class selected by the majority of trees. Conversely, for regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests aim to mitigate the issue of decision trees overfitting to their training set.

In general, random forests tend to outperform individual decision trees, although they may have lower accuracy compared to gradient boosted trees. Nonetheless, the performance of random forests can be influenced by the characteristics of the data.

The concept of random decision forests was first introduced in 1995 by Tin Kam Ho, who utilized the random subspace method. This method, as formulated by Ho, serves as an implementation of the "stochastic discrimination" approach to classification initially proposed by Eugene Kleinberg.

Support Vector Machine (SVM)

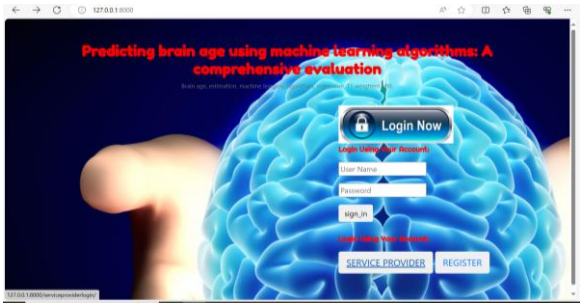
Support Vector Machine (SVM) represents a discriminant machine learning technique commonly used in classification tasks.

It aims to find a discriminant function, based on an independently and identically distributed training dataset, that accurately predicts labels for newly acquired instances. Unlike generative machine learning approaches, which necessitate computations of conditional probability distributions, a discriminant classification function assigns a data point x to one of the classes involved in the classification task.

Compared to generative approaches, discriminant methods may be less powerful, particularly in outlier detection scenarios. However, they require fewer computational resources and less training data, especially in multidimensional feature spaces and when only posterior probabilities are necessary. Geometrically, learning a classifier equates to identifying the equation for a multidimensional surface that optimally separates the different classes in the feature space.

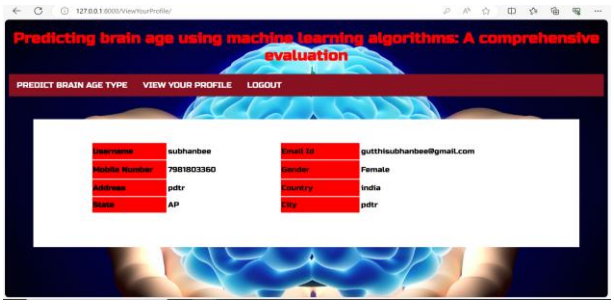
SVM is a discriminant technique that solves convex optimization problems analytically, consistently yielding the same optimal hyperplane parameters. In contrast, genetic algorithms (GAs) and perceptrons, both widely used for classification in machine learning, may produce solutions highly dependent on initialization and termination criteria. With a specific kernel transforming data from the input space to the feature space, SVM training returns uniquely defined model parameters for a given training set, whereas perceptron and GA classifier models vary with each training iteration.

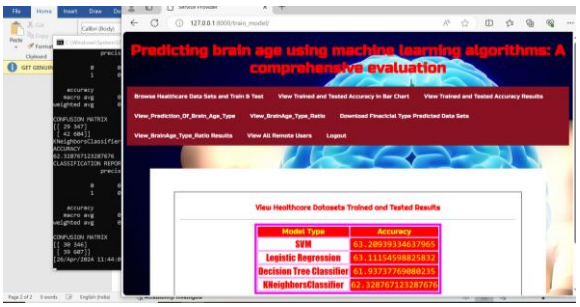
3.SCREEN SHOTS

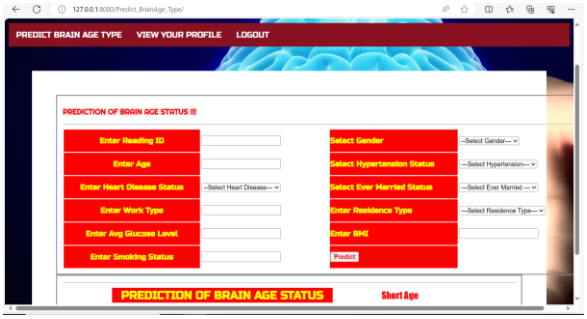


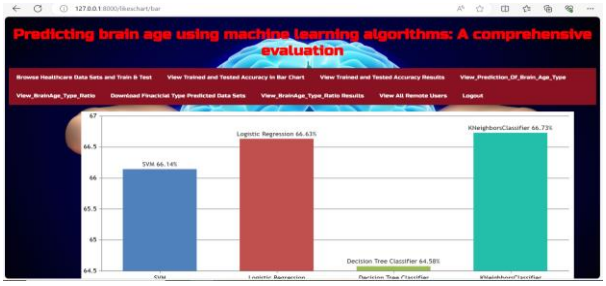


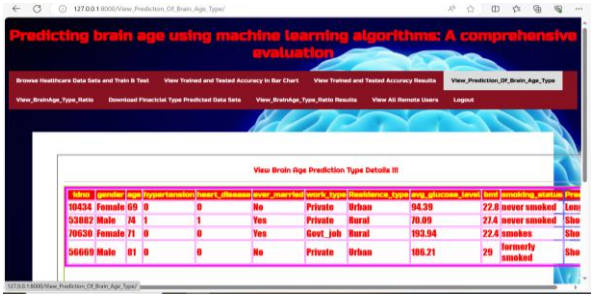












4.Conclusion

This study aimed to comprehensively evaluate various regression models for estimating Brain Age not only on CH individuals but also in clinical population. We assessed 22 different regression models on a dataset comprising CH individuals as a training set. We then quantified each regression model on independent test sets composed of CH individuals, MCI subjects, and AD patients. Our comprehensive evaluation suggests that the type of regression algorithm affects downstream comparisons between groups, and caution should be taken to select the regression model in clinical settings.

REFERENCES

[1]. T. Vos, A. D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, S. Abdalla, V. Aboyans et al., "Years lived with disability (ylds) for 1160 sequelae of 289 diseases and injuries 1990–

2010: a systematic analysis for the global burden of disease study 2010," The Lancet, vol.380, no. 9859, pp. 2163–2196, 2012.

[2]. K. Franke and C. Gaser, "Ten years of BrainAGE as a neuroimaging biomarker of brain aging: what insights have we gained?" Frontiers in Neurology, p. 789, 2019.

[3]. M. Cearns, T. Hahn, and B. T. Baune, "Recommendations and future directions for supervised machine learning in psychiatry," Translational Psychiatry, vol. 9, no. 1, pp. 1–12, 2019.

[4]. E. Luders, N. Cherbuin, and C. Gaser, "Estimating brain age using high-resolution pattern recognition: Younger brains in longterm meditation practitioners," NeuroImage, vol. 134, pp. 508–513, 2016.

[5]. S. Mishra, I. Beheshti, and P. Khanna, "A review of neuroimaging-driven brain age estimation for identification of brain disorders and health conditions," IEEE Reviews in Biomedical Engineering,2021.

[6]. D. Sone, I. Beheshti, S. Shinagawa, H. Niimura, N. Kobayashi, H. Kida, R. Shikimoto, Y. Noda,S. Nakajima, and S. Bun, "Neuroimaging-derived brain age is associated with life satisfaction in cog nitively unimpaired elderly: A communitybased study," Translational Psychiatry, vol. 12, no. 1, pp.1–6, 2022.

[7]. L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on deep learning for neuroimagingbased brain disorder analysis," Frontiers in Neuroscience, p. 779, 2020.

[8]. H. Sajedi and N. Pardakhti, "Age prediction based on brain MRI image: a survey," Journal of Medical Systems, vol. 43, no. 8, pp. 1–30, 2019.

[9]. L. Baecker, R. Garcia-Dias, S. Vieira, C. Scarpazza, and A. Mechelli, "Machine learning for brain age prediction: Introduction to methods and clinical applications," EBioMedicine, vol. 72, p. 103600, 2021.

[10]. M. Tanveer, B. Richhariya, R. Khan, A. Rashid, P. Khanna, M. Prasad, and C. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 16, no. 1s, pp. 1–35, 2020.

[11]. M. Tanveer, A. H. Rashid, R. Kumar, and R. Balasubramanian, "Parkinson's disease diagnosis us ing neural networks: Survey and comprehensive evaluation," Information Processing & Management, vol. 59, no. 3, p. 102909, 2022.

[12]. K. Muhammad, S. Khan, J. Del Ser, and V. H. C. De Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 507–522, 2020.

[13]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," ArXiv Preprint ArXiv:1409.1556, 2014.

[14]. F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An introductory review of deep learning for prediction models with big data," Frontiers in Artificial Intelligence, vol. 3, p. 4, 2020.

[15]. M. d. Condorcet, "Essay on the application of analysis to the probability of majority decisions," Paris: Imprimerie Royale, 1785.