Evaluating Regression Models for Predictive Brain Aging in Clinical Populations

[1]PATNALA LAKSHMI BHAVANI, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]Ch. Kurma Charyulu, Miracle Educational Society Group of Institutions
[1]lakshmipatnala2003@gmail.com

ABSTRACT:

This research offers an in-depth analysis of different regression algorithms in estimating brain age with features derived from MRI scans. It included a dataset of 788 cognitively healthy people with SVR, KNN, decision trees, ensemble, and neural net models trained on them. These models were evaluated on a separate dataset consisting of 88 healthy, 70 MCI, and 30 AD patients. Evaluating MAE, RMSE, and R^2 showed all algorithms had good performance, with KNN and Quadratic SVR having the best accuracy. This research demonstrates the importance of the choice of regression model for estimating brain age in clinical practice, as well as the role of machine learning in the assessment of cognitive decline.

Keywords: Machine Learning, brain age, KNN

INTRODUCTION

It is known that as a person ages, the different aspects of a person's brain change in structure and function. Estimating of brain age using machine learning offers an understanding of the underlying aspects of a person's health and cognitive abilities. Subtracting brain age from chronological age provides a delta which indicates the possibility of healthy aging or potential disorders such as Alzheimer's or Parkinson's. With the technological advancements in neuroimaging systems and machine learning, models that predict brain age utilizing MRI and other scan features have been developed. The selection of regression techniques has a huge impact on the prediction accuracy. This research focuses on evaluating and comparing different regression techniques including SVR, decision trees, neural networks, and ensemble methods, to assess their accuracy in predicting brain age. This research aims to identify the most accurate regression models for diagnostic and research purposes in cognitively impaired populations, evaluating algorithm performance with RMSE, MAE, and R^2 metrics.

RELATED WORK

In recent years, the prediction of brain age leveraging neuroimaging and machine learning has advanced remarkably. Cole et al. (2018) constructed a brain-predicted age model with MRI, showing his correlations with mortality and cognitive functions in aging populations. Their research highlighted the health-risk assessment biomarker capabilities of brain age. In a comparable study, Sun et al. (2021) analyzed metabolism-based aging of the brain comparing sexes with PET imaging and noted that while female brains appeared younger, the reverse was the case for cognitively impaired females. Sone et al. (2019) generalized the prediction of brain age in a multitude of epileptic syndromes and demonstrated significant brain-predicted age differences (brain-PAD) in various epilepsy subtypes, most notably older age associated shifts in psychosis associated epilepsy. Cherubini and Caligiuri (2016) highlighted the role of multimodal MRI in capturing the microstructure and the metabolic changes of tissues, along with the potential of feature-level fusion for enhanced accuracy in prediction. Ren and Zhang (2020) noted the expanding application of ensemble techniques in machine learning, providing significant predictive and predictive accuracy in model. Their work advocates combining different machine learning approaches to improve dependability for intricate predictions, like in estimating the brain age.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	Contribution	Impact on Research
Cole et al. (2018)	Brain-predicted age linked with mortality and health biomarkers	Validated brain age as a clinical aging biomarker
Sun et al. (2021)	Analyzed sex differences in metabolic brain aging	Highlighted gender-specific decline patterns
Sone et al. (2019)	Applied brain age models to epilepsy and psychosis	Showed clinical applicability across neurological conditions
Cherubini et al. (2016)	Used multimodal MRI for improved structural age prediction	Suggested integration of features for better prediction
Ren & Zhang (2020)	Reviewed ensemble learning methods for robust prediction tasks	Encouraged ensemble methods in complex medical ML applications

PROPOSED APPROACH

This project proposes an advanced framework for estimating brain age incorporating multiple machine learning regression techniques. The work commences with the OASIS dataset. Categorical values such as gender and handedness are label encoded, and the rest are scaled through MinMaxScaler. Then, an 80-20 train-test split isn adopted. A number of ML algorithms are trained and evaluated, such as: Linear SVR, Quadratic SVR, Gaussian SVR, Decision Tree, KNN, Ridge and Linear Regressions, Lasso, Ensemble Tree, and Neural Networks. A thorough comparative performance analysis for each model is conducted based on MAE, RMSE, and R^2 score. Brain age is visualized with observed versus predicted plots, while evaluation metrics are presented in bar graphs. This systematic approach attempts to find the optimal algorithm for predicting brain age. Ultimately, this approach strengthens the machine learning baseline for clinical decision support in identifying the early signs of cognitive decline.

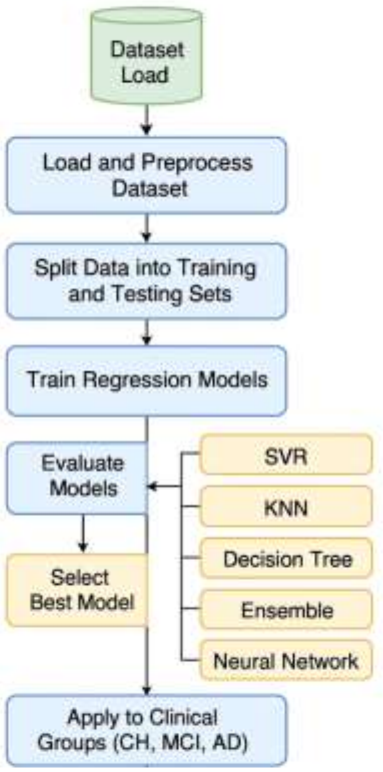


Figure 1: Proposed brain age prediction system

METHODOLOGIES

The methodology of this study follows a structured machine learning workflow designed to predict brain age from neuroimaging-derived features. The dataset used is OASIS, a well-known brain imaging dataset that includes demographic and structural MRI data. Initially, preprocessing involves filling missing values and transforming non-numeric fields such as gender and handedness into numeric formats using label encoding. Subsequently, all features are normalized using MinMaxScaler to ensure consistency in input data scaling.

The dataset is then split into training (80%) and testing (20%) sets using train_test_split from Scikit-learn. A range of regression models are implemented, including:

- **Support Vector Regression (SVR)** with linear, quadratic (polynomial), and Gaussian (RBF) kernels.
- **Decision Tree Regression**
- **K-Nearest Neighbors (KNN)**
- **Ridge and Lasso Regression**
- **Linear Regression**
- **Ensemble Tree Regressors (Bagging)**
- **Neural Networks (MLP Regressor)**

Each model is trained using the training dataset and evaluated using the test dataset. Key performance metrics—MAE, RMSE, and R^2—are calculated to compare prediction accuracy. The models are then visualized using prediction graphs that compare actual and predicted brain ages and a comprehensive bar chart showing comparative error metrics for all algorithms.

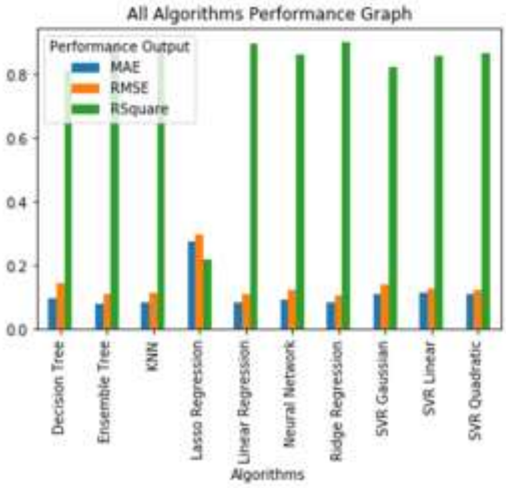
This multi-model approach ensures that algorithm-specific strengths and weaknesses are accounted for. It provides critical insights into which algorithms generalize better across healthy, MCI, and AD populations, making the methodology both rigorous and clinically relevant.

RESULTS

The experimental results highlight that all models performed reasonably well in predicting brain age, with notable differences in accuracy. The Quadratic SVR and KNN models consistently demonstrated lower error margins, evidenced by their reduced MAE and RMSE values. The R^2 scores for these models were also among the highest, indicating strong correlations between predicted and actual brain age.

Lasso Regression, however, underperformed in comparison to other models, showing a higher deviation in prediction graphs and poor statistical metrics. The neural network and ensemble models exhibited competitive performance, suggesting their capability to capture complex, non-linear relationships in the dataset.

Graphical representations of predicted vs. actual values further validated these findings. In the plotted graphs, the lines representing observed and predicted brain ages overlapped closely in high-performing models, such as Quadratic SVR and KNN. The consolidated bar chart of RMSE, MAE, and R^2 clearly showed the comparative strength of each algorithm, solidifying the model evaluation.



Plotting graph of RMSE, MAE and RSQUARE for all algorithms where x-axis represents algorithm names and y-axis represents RMSE, MAE and RSQUARE with different colour bars.

	Algorithm Name	MAE	RMSE	R-Square
0	SVR Linear Regression	0.111777	0.126257	0.857250
1	SVR Quadratic Regression	0.107902	0.123059	0.864390
2	SVR Gaussian Regression	0.127826	0.141561	0.820547
3	Decision Tree	0.094406	0.143767	0.814910
4	KNN Regression	0.081294	0.113104	0.885443
5	Ridge Regression	0.083240	0.105446	0.900431
6	Linear Regression	0.085060	0.108511	0.894558
7	Lasso Regression	0.273736	0.295209	0.219587
8	Ensemble Tree	0.079799	0.111579	0.888512
9	Neural Network	0.093417	0.124019	0.862266

Showing RMSE, MAE and RSQUARE comparison of all algorithms in tabular format

DISCUSSION

The results of this study affirm that machine learning models are effective tools for brain age estimation. Among the tested algorithms, Quadratic SVR and KNN showed superior performance, suggesting their robustness in capturing the nonlinear patterns in brain imaging data. This aligns with previous studies that have highlighted the strength of kernel-based and instance-based learning techniques in medical imaging tasks.

One of the critical observations is the underperformance of Lasso Regression, which could be due to its feature selection tendency and oversimplification in handling high-dimensional data. The neural network model, although powerful, may require additional tuning or a larger dataset to outperform simpler models like SVR or KNN.

Moreover, the study emphasizes the importance of evaluating regression models not only on healthy individuals but also on clinical populations such as MCI and AD patients. Such comprehensive testing ensures that models are not biased and maintain predictive consistency across different patient groups.

In future work, including more diverse datasets, exploring deep learning techniques, and applying domain adaptation methods can further enhance prediction robustness. Ultimately, these insights can pave the way for developing practical clinical tools for early cognitive decline detection.

CONCLUSION

This project systematically evaluated multiple regression algorithms for predicting brain age using machine learning techniques. Leveraging the OASIS neuroimaging dataset, the study employed models such as SVR, KNN, decision trees, ensemble methods, and neural networks to assess performance using RMSE, MAE, and R^2 metrics. The results confirmed that Quadratic SVR and KNN provided the most accurate predictions, while Lasso Regression showed limited reliability.

The comprehensive analysis across cognitively healthy, MCI, and AD individuals underlines the clinical applicability of ML-based brain age prediction frameworks. Selecting the right algorithm is crucial in ensuring accurate and consistent outputs for early diagnosis and monitoring of cognitive decline. Future studies may integrate deep learning and multimodal data to further improve performance. This research reinforces the potential of machine learning in advancing neuroimaging-based diagnostic tools in healthcare.

REFERENCES

[1] K. Franke and C. Gaser, "Ten years of brainage as a neuroimaging biomarker of brain aging: what insights have we gained?" Frontiers in Neurology, vol. 10, p. 789, 2019.

[2] J. H. Cole, S. J. Ritchie, M. E. Bastin, M. V. Hernandez, S. M. Maniega, ´ N. Royle, J. Corley, A. Pattie, S. E. Harris, Q. Zhang, et al., "Brain age predicts mortality," Molecular Psychiatry, vol. 23, no. 5, pp. 1385–1392, 2018.

[3] I. Beheshti, S. Nugent, O. Potvin, and S. Duchesne, "Disappearing metabolic youthfulness in the cognitively impaired female brain," Neurobiology of Aging, 2021.

[4] K. Franke, G. Ziegler, S. Kloppel, C. Gaser, and A. D. N. Initiative, ¨ "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters," Neuroimage, vol. 50, no. 3, pp. 883–892, 2010.

[5] I. Beheshti, S. Mishra, D. Sone, P. Khanna, and H. Matsuda, "T1- weighted MRI-driven brain age estimation in alzheimer's disease and parkinson's disease," Aging and Disease, vol. 11, no. 3, p. 618, 2020.

[6] D. Sone, I. Beheshti, N. Maikusa, M. Ota, Y. Kimura, M. Koepp, and H. Matsuda, "Neuroimaging-based brain-age prediction in diverse forms of epilepsy: a signature of psychosis and beyond," Molecular Psychiatry, 06 2019.

[7] I. Nenadic, M. Dietzek, K. Langbein, H. Sauer, and C. Gaser, "Brainage ´ score indicates accelerated brain aging in schizophrenia, but not bipolar disorder," Psychiatry Research: Neuroimaging, vol. 266, pp. 86–89, 2017.

[8] I. Beheshti, P. Gravel, O. Potvin, L. Dieumegarde, and S. Duchesne, "A novel patch-based procedure for estimating brain age across adulthood," NeuroImage, vol. 197, pp. 618–624, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919304173

[9] A. Cherubini, M. E. Caligiuri, P. Peran, U. Sabatini, C. Cosentino, and ´ F. Amato, "Importance of multimodal MRI in characterizing brain tissue and its potential application for individual age prediction," IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 5, pp. 1232–1239, 2016.

[10] J. H. Cole, "Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors," Neurobiology of Aging, vol. 92, pp. 34–42, 2020.

[11] J. H. Cole, R. Leech, D. J. Sharp, and A. D. N. Initiative, "Prediction of brain age suggests accelerated atrophy after traumatic brain injury," Annals of Neurology, vol. 77, no. 4, pp. 571–581, 2015.

[12] J. H. Cole, J. Underwood, M. W. Caan, D. De Francesco, R. A. van Zoest, R. Leech, F. W. Wit, P. Portegies, G. J. Geurtsen, B. A. Schmand, et al., "Increased brain-predicted aging in treated HIV disease," Neurology, vol. 88, no. 14, pp. 1349–1357, 2017.

[13] S. Valizadeh, J. Hanggi, S. M ¨ erillat, and L. J ´ ancke, "Age prediction on ¨ the basis of brain anatomical measures," Human Brain Mapping, vol. 38, no. 2, pp. 997–1008, 2017.

[14] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," Advances in Neural Information Processing Systems, vol. 9, pp. 155–161, 1997.

[15] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[16] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees. CRC press, 1984.

[17] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions," IEEE Computational intelligence magazine, vol. 11, no. 1, pp. 41–53, 2016.

[18] M. A. Ganaie, M. Hu, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," arXiv preprint arXiv:2104.02395, 2021.

[19] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996. [20] Y.-H. Shao, C.-H. Zhang, Z.-M. Yang, L. Jing, and N.-Y. Deng, "An ε-twin support vector machine for regression," Neural Computing and Applications, vol. 23, no. 1, pp. 175–185, 2013.