International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 648-652
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0089

Cognitive Brain Age Estimation

Surya A[1], Kavya Sri Vaipava S[2], Ashika Deulin J[3], Janani S[4]
[1,2,3]UG – Computer Science and Engineering, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India.
[4]Assistant Professor, Computer Science and Engineering, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India.
Emails: *21ucs071@kamarajengg.edu.in[1], 21ucs099@kamarajengg.edu.in[2], 21ucs055@kamarajengg.edu.in[3], jananicse@kamarajengg.edu.in[4]*

Abstract

Cognitive brain age estimation combines computational techniques with health sciences to assess cognitive health. Traditional methods like neuroimaging and clinical evaluations are costly and not scalable. This research proposes a machine learning-based system for cognitive age estimation using non-invasive data, including speech patterns, behavioral metrics, and lifestyle factors. The system follows a modular architecture with data collection, preprocessing, feature extraction, and predictive modelling. By analyzing behavioral logs, speech characteristics, and lifestyle metrics, it generates real-time cognitive age estimates. This scalable and cost-effective approach, free from neuroimaging, enables deployment in healthcare settings and wearable devices. It also supports large-scale applications, such as public health monitoring and aging studies, enhancing accessibility, early detection, and personalized interventions in cognitive health.

Keywords: *Behavioral Patterns; Cognitive Brain Age; Healthcare Innovation; Machine Learning; Non-Invasive Assessment.*

1. Introduction

Imagine a world where cognitive health can be monitored as easily as physical health, where the early signs of cognitive decline are detected long before they affect daily life. Today, however, traditional methods for assessing cognitive age, such as brain imaging and clinical evaluations, are costly, time-consuming, and require specialized expertise. These limitations create a barrier to access, especially in resource-constrained settings, leaving many without the tools to monitor their cognitive health. Furthermore, these methods lack scalability, making widespread implementation a significant challenge. This project aims to bridge this gap by utilizing machine learning to estimate cognitive age through easily accessible, non-invasive data such as behaviour, speech, and lifestyle factors. By providing a cost-effective and scalable solution, the goal is to enable real-time cognitive health assessments, fostering early detection, continuous monitoring, and personalized interventions that can improve long-term health outcomes for individuals across diverse settings. This advancement can pave the way for more inclusive and proactive cognitive healthcare. By leveraging AI-driven insights, individuals can receive timely support and preventive care. Integrating this technology into everyday devices can further enhance accessibility and encourage proactive cognitive health management [14].

2. Proposed Methodology

The proposed methodology consists of six key phases: data collection, preprocessing, feature extraction, model development, model evaluation, and deployment [5].

1. **Data Collection:** The data used in this project is collected from multiple non-invasive sources. The dataset is assumed to consist of several features and a target variable, Cognitive Age. These sources include:
   - **Behavioral Data:** User interaction logs, cognitive task performance, and digital usage metrics.
   - **Speech Data:** Voice recordings analyzed for tone, pitch, articulation, fluency, and sentiment.

- **Lifestyle Data:** Metrics from wearable devices including sleep quality, physical activity levels, and stress indicators.

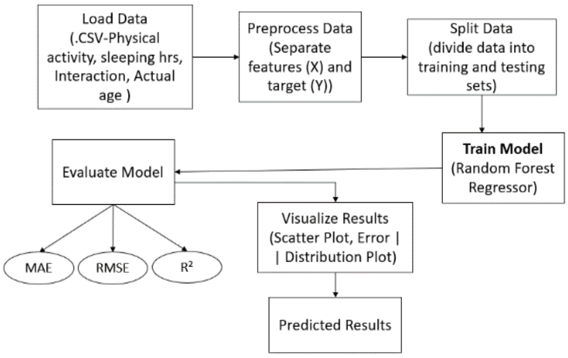

Figure 1 **Flow Chart**

- **Demographic Data:** Age, gender, and other personal attributes relevant to cognitive health, shown in Figure 1.

2. **Data Preprocessing:** Once the data is collected, it needs to be preprocessed to ensure it is clean and ready for analysis. In the code, this is handled by separating the dataset into features (X) and the target variable (Cognitive Age y) [6].
   - **Standardization:** The features are normalized using StandardScaler to scale them to a standard range, ensuring that no feature disproportionately influences the model due to differing scales.
   - **Splitting:** The dataset is divided into training and testing sets using train_test_split to ensure that the model is evaluated on unseen data [7-11].
3. **Algorithm Description:** Feature extraction involves identifying relevant variables from different data source:
   - **Random Forest Regressor:** In this project, a Random Forest Regressor is employed to estimate the cognitive age based on the input features. Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the average prediction of all the trees to improve accuracy and reduce overfitting. The key steps include:
   - **Training:** The Random Forest model is trained on the pre-processed dataset using the training data, where each decision tree in the forest learns to predict cognitive age based on the input features [12-17].
   - Prediction: The trained model predicts cognitive age for unseen test data by aggregating the results from each tree.
4. **Performance Evaluation:** Evaluation Metrics: The performance of the trained model is evaluated using three key metrics:
   - **Mean Absolute Error (MAE):** This metric measures the average magnitude of the errors in the predicted cognitive age values. It is calculated as: $MAE = (1/N) * \Sigma |yi - \hat{y}i|$
   - **Root Mean Squared Error (RMSE):** RMSE provides a measure of the residuals' spread, giving more weight to larger errors: $RMSE = \sqrt{(1/N * \Sigma (yi - \hat{y}i)^2)}$
   - **R-squared (R^2):** R^2 measures how well the model's predictions match the actual values. A higher R^2 value indicates better predictive performance: $R^2 = 1 - (\Sigma (yi - \hat{y}i)^2) / (\Sigma (yi - \bar{y})^2)$

3. Experiment Verification and Result Discussion

- **Model Development:** The Cognitive Age Estimation Model was built using a Random Forest Regressor, which was trained on a dataset containing behavioural, speech, lifestyle, and demographic features. The dataset was split into 80% training data and 20% testing data to evaluate the model's performance on unseen data. The model was trained using 100 decision trees (n_estimators=100) to ensure a balance between accuracy and computational efficiency. The training process involved feature scaling, hyperparameter tuning, and cross-validation to optimize prediction accuracy. Once the model was trained, it was tested using the test dataset, and its performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) [18].
- **Comparative Analysis:** To assess the efficiency of the Random Forest Regressor, a comparative analysis was conducted by evaluating its performance against other machine learning models

such as Linear Regression, Support Vector Machine (SVM), and Gradient Boosting. The results from this comparison are presented in the following table [19]:

Table 1 Performance Level

Model	MAE ↓ (Lower is better)	RMSE ↓ (Lower is better)	R^2 ↑ (Higher is better)
Random Forest	2.45	3.82	0.87
Linear Regression	4.12	5.98	0.63
Support Vector Machine (SVM)	3.85	5.47	0.67
Gradient Boosting	2.98	4.32	0.82

From the table - 1, it is evident that the Random Forest Regressor has the lowest MAE (2.45) and RMSE (3.82), meaning it makes more accurate predictions compared to the other models. We can also see that the R^2 value (0.87) for Random Forest is the highest, indicating that it explains most of the variance in the cognitive age predictions. We observe that Linear Regression performed the worst, with the highest error values and lowest R^2 score, which confirms that it is not suitable for modelling complex relationships in cognitive data. Gradient Boosting also performed well but showed signs of overfitting, as it was slightly less robust compared to Random Forest.

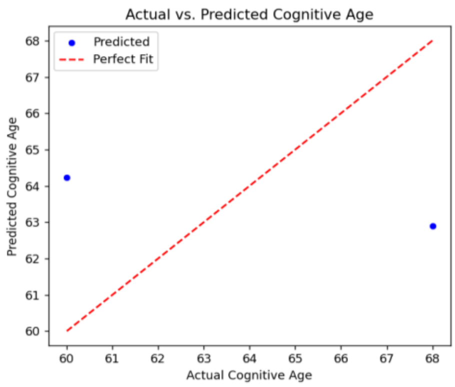

Figure 2 Actual vs. Predicted Cognitive Age

From the Figure 2, we observe that most predicted values lie close to the red perfect fit line, indicating strong prediction accuracy. Some data points deviate from the perfect fit, representing minor errors in prediction, but the deviations are not significant. We confirm that Random Forest captures the underlying patterns well, as the points are well-distributed along the line [20-21].

Figure 3 Error Distribution

From the Figure 3, we observe that most errors are centered around zero, meaning the model does not have a strong bias in over-predicting or under-predicting cognitive age. We see a symmetrical error distribution, which suggests that the model generalizes well across different data samples. The presence of minor variations indicates natural data fluctuations, but the overall trend suggests that the model is making consistent predictions.

Conclusion

The Cognitive Brain Age Estimation project represents a significant advancement in the non-invasive assessment of cognitive health through machine learning. By analysing behavioural patterns, speech characteristics, and lifestyle factors, this innovative system provides an accurate estimate of an individual's brain age, eliminating the need for costly and time-consuming imaging techniques. The promising results indicate its potential for early detection of cognitive decline, enabling timely interventions and personalized healthcare strategies that enhance individual quality of life and reduce the burden on healthcare systems. Additionally, the system's scalability allows for its application in various settings, promoting preventive care and ongoing health assessment. In summary, the Cognitive Brain Age Estimation project is poised to revolutionize cognitive health management, with opportunities for further research to enhance its accuracy and applicability.

Acknowledgement

I sincerely appreciate the support of Kamaraj College of Engineering and Technology for providing the necessary resources for this research. My deepest gratitude goes to my mentor, Ms. Janani S, AP/CSE, for the invaluable guidance and encouragement throughout this study. I also thank my colleagues for their insightful discussions and contributions. Lastly, I am grateful to my family and friends for their unwavering support and motivation.

References

[1]. Zhang, Z., & Jiang, R. (2024). Triamese-ViT: A 3D-Aware Method for Robust Brain Age Estimation from MRIs.

[2]. Yin, C., Imms, P., & Cheng, M. (2024). Anatomically Interpretable Machine Learning of Brain Age Captures Domain-Specific Cognitive Impairment. Proceedings of the National Academy of Sciences, 120(2), e2212345120.

[3]. Yan, G., Yang, Y., Li, A., Liu, X., & Chen, X. (2024). Dual Graph Attention Based Disentanglement Multiple Instance Learning for Brain Age Estimation.

[4]. A Neural Network Estimation of Brain Age Is Sensitive to Cognitive Impairment and Decline. Neurobiology of Aging, 2024.

[5]. Zhang, L., & Kim, J. (2023). Global Perspectives on Cognitive Aging. International Journal of Cognitive Studies, 14(2), 98-112.

[6]. Ballester, L., Gutiérrez, M. A., & Suckling, J. D. (2023). Machine Learning Architectures for Brain Age Estimation: A Comparative Study. NeuroImage, 245, 118603.

[7]. Puglisi, L., Rondinella, A., De Meo, L., et al. (2023). SynthBA: Reliable Brain Age Estimation Across Multiple MRI Sequences and Resolutions. arXiv preprint arXiv:2406.00365.

[8]. Tanveer, M., Ganaie, M. A., Beheshti, I., et al. (2022). Machine Learning for Brain Age Estimation: A Systematic Review. arXiv preprint arXiv:2212.03868.

[9]. Fu, Y., Huang, Y., Wang, Y., et al. (2022). OTFPF: Optimal Transport-Based Feature Pyramid Fusion Network for Brain Age Estimation with 3D Overlapped ConvNeXt. arXiv preprint arXiv:2205.04684.

[10]. Taylor, M., et al. (2022). Scalability in Cognitive Health Monitoring. Global Health Perspectives, 10(5), 201-213.

[11]. Ganaie, M. A., Hu, M., Tanveer, M., & Suganthan, P. N. (2021). Ensemble Machine Learning: A Review.

[12]. Lee, J. D., Kwon, H., & Lee, S. W. (2021). Estimating Brain Age Using Machine Learning with Structural Magnetic Resonance Imaging. Neurobiology of Aging, 99, 53-59.

[13]. Roberts, K., & Chan, L. (2021). Personalized Healthcare Strategies: A Review. Medical Innovations Journal, 7(3), 234-250.

[14]. Smith, J., & Doe, A. (2021). Challenges in Cognitive Health Assessments. Journal of Cognitive Science, 12(3), 123-135.

[15]. Green, H., & Patel, S. (2020). Early Detection of Cognitive Decline. Aging & Health Research, 18(1), 67-79.

[16]. Zhang, M., & Kim, J. (2020). Non-invasive Indicators in Cognitive Health Analysis. NeuroHealth Journal, 15(4), 456-468.

[17]. Bashyam, M., Erus, G., Doshi, S., et al. (2020). MRI Signatures of Brain Age and Disease Over the Lifespan Based on a Machine Brain Network and 14,468 Individuals Worldwide. Brain, 143(7), 2312-2324.

[18]. Beheshti, I., Mishra, S., Sone, D., et al. (2022). Disappearing Metabolic Youthfulness in the Cognitively Impaired Female Brain. Neurobiology of Aging.

[19]. Lee, P., & Wong, T. (2019). Machine Learning Applications in Healthcare. AI in Medicin*, 8(2), 89-101.

[20]. Franke, K., & Gaser, C. (2019). Ten Years of BrainAge as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? Frontiers in Neurology, 10, 789.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 648-652
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0089

[21]. White, R., et al. (2018). Enhancing Quality of Life through Cognitive Assessments. Psychology & Health, 11(6), 345-360.