Pertanika Proceedings 1 (6): 97 - 104 (2025)

PERTANIKA JOURNALS

PERTANIKA PROCEEDINGS

Journal homepage: http://www.pertanika.upm.edu.my/

# Optimising Cognitive Assessment: Key Variable Identification for Efficiency using Ensemble Learning

**Salami Safran Olugbenga[1,2], Fakhrul Zaman Rokhani[1]*, Halimatus Sakdiah Minhat[1], Siti Anom Ahmad[1], Syamsiah Mashohor[1], Helen Zhao[3], and Maw Pin Tan[4]**

[1]*Faculty of Engineering University Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia*
[2]*Faculty of Engineering, Lagos State University of Science and Technology, Lagos, Nigeria*
[3]*Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[4]*School of Smart Health and Wellness, Health Medical College, Zhejiang Dongfang Polytechnic, China*

## ABSTRACT

Mild Cognitive Impairment (MCI) signifies an abnormal cognitive decline that surpasses the natural deterioration associated with aging. The Montreal Cognitive Assessment (MoCA) is a leading tool for MCI detection, evaluating twelve variables to diagnose various cognitive domains. This study aims to optimize the MoCA by identifying the most critical variables, thus simplifying the test while maintaining diagnostic accuracy. Using an Ensemble model comprising Gradient Boosting and Random Forest techniques, eight key variables were identified, enhancing the MoCA's efficiency without compromising its effectiveness.

*Keywords:* Cognitive impairment (CI), feature selection, frailty, machine learning, MCI, MOCA

## INTRODUCTION

Mild Cognitive Impairment (MCI) stands as an early, decisive warning sign of possible cognitive decline, frequently preceding more severe neurodegenerative conditions such as Alzheimer's disease(AD) (Boettcher et al., 2020). It occupies a transitional cognitive state between typical age-related decline and more debilitating conditions like dementia, where individuals may exhibit measurable cognitive deficits—particularly in memory, attention, and

ARTICLE INFO

*Article history:*
Received: 30 September 2025
Published: 28 November 2025

DOI: https://doi.org/10.47836/pp.1.6.016

*E-mail addresses*:
abdulsalasolugbenga@gmail.com (Salami Safran Olugbenga)
fzr@upm.edu.my (Fakhrul Zaman Rokhani)
halimatus@upm.edu.my (Halimatus Sakdiah Minhat)
sanom@upm.edu.my (Siti Anom Ahmad)
syamsiah@upm.edu.my (Syamsiah Mashohor)
helen@zjdfp.edu.cn (Helen Zhao)
mawpin@um.edu.my (Maw Pin Tan)
* Corresponding author

executive function—yet retain a degree of independence. Approximately 15% of individuals with MCI progress to AD annually (Liu et al., 2020; Quek et al., 2023). This alarming statistic highlights the urgent need for precise and timely diagnostic instruments that enable healthcare professionals to act quickly and prevent progression. Various tools, including the Mini Mental State Examination (MMSE) and Clinical Dementia Rating (CDR), have been identified in multiple literatures for assessing neurocognitive disorders (Boettcher et al., 2020; Ghoraani et al., 2021; Idris & Badruddin, 2021; Kumar et al., 2021; Louka et al., 2022; Mantovani et al., 2020; Zhou et al., 2022). The Montreal Cognitive Assessment (MoCA) has been widely acknowledged as a better tool for detecting MCI. The MoCA test, involving a 30-point assessment over 15 minutes, is highly regarded for evaluating cognitive abilities across multiple domains (Corral et al., 2024). In comparison with MMSE, researchers adjudged MoCA to be significantly better, more sensitive and more effective in detecting MCI (Roeck et al., 2019; Tan et al., 2021). Recent adaptations, like the MoCA-Blind, cater to vision- impaired individuals by excluding visual components (Effendi-Tenang et al., 2020; Wittich et al., 2010).

The Montreal Cognitive Assessment is widely recognised as an effective tool for detecting MCI by assessing multiple cognitive domains through twelve variables. Although MoCA has demonstrated value, its twelve parameters covering various cognitive areas add to its complexity, making it difficult to administer in fast-paced clinical settings. The mental effort required can cause fatigue and errors in older adults, highlighting the need for alternative approaches that balance simplicity with diagnostic accuracy. (Roeck et al., 2019; Tan et al., 2021). The challenge is to identify the most important variables in MoCA that can accurately predict cognitive impairment while minimizing the time and complexity of the assessment. The key question is; how can the MoCA be improved by pinpointing the most critical variables that preserve its diagnostic accuracy for cognitive impairment, thereby reducing the overall length and difficulty of the evaluation? This research aims to address this issue by creating a more concise version of the MoCA using an ensemble model that employs Gradient Boosting and Random Forest techniques to identify essential variables without sacrificing the tool's effectiveness and accuracy. To our knowledge, this is the first time a machine learning model has been used to develop a short-form MoCA.

## MATERIALS AND METHODS

The study utilized data from the Transforming Cognitive Frailty into Self-Sufficiency (AGELESS) and Malaysian Elders Longitudinal Research (MELOR) studies (Aravindhan et al., 2022; Asmuje et al., 2023). The dataset included 2177 participants after excluding those diagnosed with dementia. Data preprocessing involved eliminating cases with excessive null values; participants with missing values greater than 20% were removed, resulting in 1645 cases. Participants' cognitive impairment was categorized into three classes: normal Cognition (NC: MoCA scale; 28-30), (MCI: Moca scale; 20-25), and

Subjective Cognitive Decline (SCD: MoCA scale; 26-27) (Boettcher et al., 2020; Lin et al., 2021; Rotenberg et al., 2020). The MoCA variables were standardised for analysis. The descriptive analysis shows that the **NC** group has a younger mean age (69.26 years, SD = 6.16) and is predominantly female (mean gender value = 0.406, SD = 0.491), with a lower mean ethnicity value (0.57, SD = 0.89), indicating less ethnic diversity. In contrast, the MCI group shows an older mean age (72.41 years, SD = 6.87), a nearly equal gender split (mean = 0.487, SD = 0.500), and the highest ethnic diversity (mean = 1.26, SD = 1.06). The SCD group has a mean age of 72.04 years (SD = 6.60), a balanced gender distribution (mean = 0.485, SD = 0.500), and moderate ethnic diversity (mean = 0.91, SD = 1.02), suggesting increased symptoms as cognitive decline progresses (Table 1). The seven cognitive domains are described in Table 2; visuospatial ability (visuospatial 1-3), naming, attention (attention 1-3), language (language 1&2), abstraction, memory (delay recall), and orientation.

Table 1
*Descriptive statistics of demography of participants*

| **Participant Description** | **Mild Cognitive Impairment (MCI)** | | |
|---|---|---|---|
| | **(n=1645) NC (n=773)** | **SCD (n= 398)** | **MCI (n=474)** |
| Age, μ(σ) | 69.26(6.16) | 72.04(6.60) | 72.419(6.87) |
| Gender | | | |
| Female, No (%) | 459 (59.38) | 205 (51.51) | 243 (51.27) |
| Male, No (%) | 314 (40.62) | 193 (48.49) | 231 (48.73) |
| Ethnicity | | | |
| Chinese, No (%) | 500 (64.68) | 192 (48.24) | 161 (33.97) |
| Indian, No (%) | 134 (17.34) | 74 (18.59) | 85 (17.93) |
| Malay, No (%) | 118 (15.27) | 111 (27.89) | 177 (37.34) |
| Others, No (%) | 21 (2.72) | 21 (5.28) | 51 (10.76) |
| Education Level | | | |
| <= 6 years of Formal Education No (%) | 44 (5.69) | 77 (19.35) | 175 (36.92) |
| > 6 years of Formal Education No (%) | 729 (94.31) | 321 (80.65) | 299 (63.08) |

NC = normal cognition; MCI = mild cognitive impairment; SCD = subjective cognitive decline; SD = standard deviation

Table 2
*Description of cognitive domains*

| **MOCA Variable** | **Description** | **Scale** |
|---|---|---|
| **Visuospatial_1** | Number order arrangement | 1 |
| **Visuospatial_2** | Copy cube | 1 |
| **Visuospatial_3** | Draw a clock (Ten past eleven) | 3 |

Table 2 *(continue)*

| MOCA Variable | Description | Scale |
|---|---|---|
| **Naming** | Name the following; [ ] Lion [ ] Elephant [ ] Camel | 3 |
| **Attention_1** | Read a list of digits; Forward [ ] 2, 1, 8, 5, 4 Backward [ ] 7, 4, 2 | 2 |
| **Attention_2** | Read the list of numbers; 6, 2, 1, 3, 7, 8, 1, 1, 9, 7, 6, 2, 1, 6, 1, 7, 4, 5, 1, 1, 1, 9, 1, 7, 9, 6, 1, 1, 2 | 1 |
| **Attention_3** | Serial 7 subtraction starting at 100 | 3 |
| **Language_1** | Repeat sentences; 1 point for each sentence successfully repeated | 2 |
| **Language_2** | Fluency; Name as many animals as possible in one minute) | 1 |
| **Abstraction** | Describe the similarity between; Train - Bicycle, Watch - Ruler | 2 |
| **Delay recall** | Recall words ; [ ] Face [ ] Silk [ ] Church [ ] Rose [ ] Red | 5 |
| **Orientation** | [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] Country | 6 |
| Total Scale | | 30 |

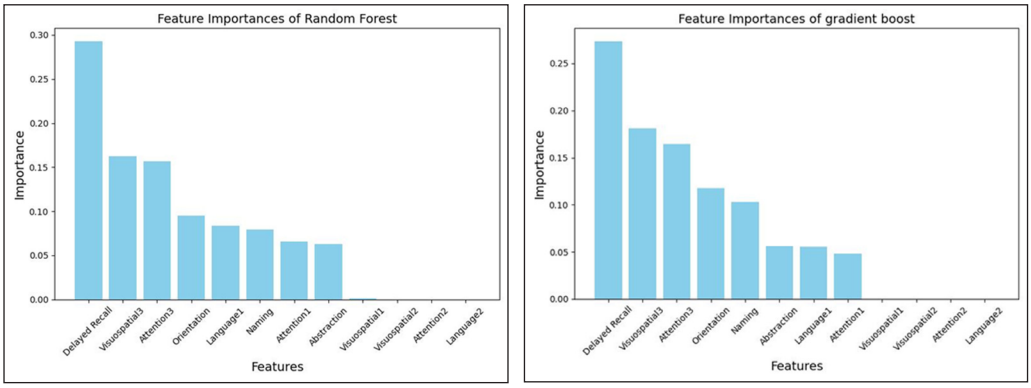
## Machine Learning Model

Two ensemble learning models, Gradient Boost and Random Forest, were employed for feature selection. The ensemble model used in this study—integrating Gradient Boosting and Random Forest techniques—is particularly well-suited for this task. Gradient Boosting constructs multiple weak models (typically decision trees) and combines them to form a stronger predictive model (Emami & Martínez-Muñoz, 2025). This method optimizes complex models like the MOCA by identifying and emphasizing the most important variables. Also, Random Forests, which consist of an ensemble of decision trees, provide robustness and flexibility, making them ideal for handling complex datasets with multiple variables (Li et al., 2020). Together, these techniques facilitate a thorough approach to refining the MOCA, ensuring that the final model retains diagnostic accuracy while enhancing usability. The data was split into training (80%) and testing (20%) sets. The models were evaluated based on their ability to predict cognitive impairment. The selected features were then applied on seven traditional machine learning models such as; Logistic Regression (LR), Linear Descriptive Analysis (LDA), K-Nearest Neighbor (KNN), Classification and Regression Tree (CART), Support Vector Machine (SVM), Random Forest Classifier (RFC) and Gaussian Naïve Baiye (GNB).

## RESULTS

The ensemble models, Gradient Boost and Random Forest, identified Delayed Recall, Visuospatial 3, and Attention 3 as the most significant variables (see Figure 1 (a & b)). The Gradient Boost model achieved an accuracy of 81.2%, while the Random Forest model reached 83.5%. The study retained all seven cognitive domains but found four

variables (Visuospatial 1 & 2, Attention 2, and Language 2) to be redundant, supporting a streamlined MoCA with eight key variables. Unlike previous work that reduced the number of domains from seven to four (Tan et al., 2021), compromising accuracy, our approach maintains all seven domains, only reducing variables within them, resulting in better predictive performance. The eight selected variables were then trained using various traditional machine-learning models and compared results with the full MoCA in predicting cognitive impairment. The results show that the modified MoCA performs comparably to the full MoCA across all metrics in six models and even outperforms it in the Gaussian Naïve Bayes model, with 73% accuracy against 39% for the full MoCA (Table 3). These findings confirm that the modified MoCA retains its effectiveness in predicting cognitive impairment (CI) with a reduced variable count from twelve to eight.




a b

*Figure 1.* Visualisation of important features in (a) Random forest and (b) Gradient boost

Table 3
*Machine learning results of modified and full versions of the Montreal Cognitive Assessment*

| Modified MoCA prediction of MCI | | | | | Full MoCA prediction of MCI | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Models** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **Models** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| LR | 79 | 75 | 75 | 75 | LR | 79 | 75 | 75 | 75 |
| LDA | 80 | 77 | 75 | 76 | LDA | 80 | 78 | 75 | 76 |
| KNN | 81 | 80 | 79 | 79 | KNN | 82 | 80 | 79 | 80 |
| CART | 85 | 84 | 82 | 83 | CART | 85 | 84 | 82 | 83 |
| SVM | 82 | 80 | 80 | 80 | SVM | 81 | 80 | 80 | 80 |
| RFC | 85 | 84 | 82 | 83 | RFC | 85 | 84 | 82 | 83 |
| GNB | 73 | 70 | 68 | 69 | GNB | 39 | 70 | 50 | 38 |

MoCA=Montreal Cognitive Assessment; MCI=mild cognitive impairment; LR: Logistic Regression, LDA: Linear Descriptive Analysis, KNN: K-Nearest Neighbor, CART: Classification And Regression Tree, SVM: Support Vector Machine, RFC: Random Forest Classifier, GNB: Gaussian Naïve Baiye

## DISCUSSION AND CONCLUSION

The findings indicate that the MoCA test can be optimized by concentrating on eight significant variables in the cognitive domain without sacrificing the accuracy of predicting cognitive impairment. This enhances its efficiency and reduces administration time. While the optimized MoCA tool promises reduced administration time, its practical implementation in clinical settings may face challenges, such as user acceptance by clinicians and adaptability across different healthcare systems. Future research should determine a new scale for the modified MoCA and validate the findings to confirm the optimized MoCA's effectiveness. This streamlined version promises to enable quicker and equally reliable cognitive assessments, benefiting both clinical geriatricians and patients.

## ACKNOWLEDGEMENT

I want to express my heartfelt gratitude to the Malaysian Ministry of Higher Education for their generous support through the Long Term Research Grant Scheme (LRGS), which made my participation in the NCA 2024 conference possible.

## REFERENCES

Aravindhan, K., Mat, S., Hamid, T. A., Shahar, S., Abdul Majeed, A. B., Teh, P. L., Ramasamy, K., Singh, D. K. A., & Tan, M. P. (2022). Development of virtual surveys for the COVID-19 wave of the AGELESS longitudinal study in Malaysia. *Gerontology*, *68*(5), 551-555. https://doi.org/10.1159/000517946

Asmuje, N. F., Mat, S., Myint, P. K., & Tan, M. P. (2023). Ethnic-specific sociodemographic factors as determinants of cognitive performance: Cross-sectional analysis of the malaysian elders longitudinal research study. *Dementia and Geriatric Cognitive Disorders*, *51*(5), 396-404. https://doi.org/10.1159/000526904

Boettcher, L. N., Hssayeni, M., Rosenfeld, A., Tolea, M. I., Galvin, J. E., & Ghoraani, B. (2020). Dual-task gait assessment and machine learning for early-detection of cognitive decline. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020*, 3204-3207. https://doi.org/10.1109/EMBC44109.2020.9175955

Corral, S., Gaspar, P. A., Castillo-Passi, R. I., Mayol Troncoso, R., Mundt, A. P., Ignatyev, Y., Nieto, R. R., & Figueroa-Muñoz, A. (2024). Montreal cognitive assessment (MoCA) as a screening tool for cognitive impairment in early stages of psychosis. *Schizophrenia Research: Cognition*, *36*, Article 100302. https://doi.org/10.1016/j.scog.2024.100302

Effendi-Tenang, I., Tan, M. P., Khaliddin, N., Jamaluddin Ahmad, M., Amir, N. N., Kamaruzzaman, S. B., & Ramli, N. (2020). Vision impairment and cognitive function among urban-dwelling malaysians aged 55 years and over from the Malaysian Elders Longitudinal Research (MELoR) study. *Archives of Gerontology and Geriatrics*, *90*, Article 104165 https://doi.org/10.1016/j.archger.2020.104165

Emami, S., & Martínez-Muñoz, G. (2025). Condensed-gradient boosting. *International Journal of Machine Learning and Cybernetics*, *16*(1), 687-701. https://doi.org/10.1007/s13042-024-02279-0

Ghoraani, B., Boettcher, L. N., Hssayeni, M. D., Rosenfeld, A., Tolea, M. I., & Galvin, J. E. (2021). Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning. *Biomedical Signal Processing and Control*, *64*, Article 102249. https://doi.org/10.1016/j.bspc.2020.102249

Idris, S., & Badruddin, N. (2021). Classification of cognitive frailty in elderly people from blood samples using machine learning. In *IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1-4). IEEE. https://doi.org/10.1109/BHI50953.2021.9508514

Kumar, S., Du, C., Graham, S., & Nguyen, T. (2021). Using machine learning to predict frailty from cognitive assessments. In *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1648-1652). IEEE. https://doi.org/10.1109/EMBC46164.2021.9630386

Li, J., Tian, Y., Zhu, Y., Zhou, T., Li, J., Ding, K., & Li, J. (2020). A multicenter random forest model for effective prognosis prediction in collaborative clinical research network. *Artificial Intelligence in Medicine*, *103*, Article 101814. https://doi.org/10.1016/j.artmed.2020.101814

Lin, X., Liu, F., Wang, B., Dong, R., Sun, L., Wang, M., & Bi, Y. (2021). Subjective cognitive decline may be associated with post- operative delirium in patients undergoing total hip replacement: The PNDABLE study. *Frontiers in Aging Neuroscience*, *13*, Article 680672. https://doi.org/10.3389/fnagi.2021.680672

Liu, J., Pan, Y., Wu, F. X., & Wang, J. (2020). Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification. *Neurocomputing*, *400*, 322-332. https://doi.org/10.1016/j.neucom.2020.03.006

Louka, A. M., Tsagkaris, C., Christoforou, P., Khan, A., Alexiou, F., Simou, P., Haranas, I., Gkigkitzis, I., Georgios, Z., Jha, N. K., Uddin, M. S., Shen, B., Kamal, M. A., Ashraf, G. M., & Alexiou, A. (2022). Current trends of computational tools in geriatric medicine and frailty management. *Frontiers in Bioscience - Landmark*, *27*(8), Article 232. https://doi.org/10.31083/j.fbl2708232

Mantovani, E., Zucchella, C., Schena, F., Romanelli, M. G., Venturelli, M., & Tamburin, S. (2020). Towards a redefinition of cognitive frailty. *Journal of Alzheimer's Disease*, *76*(3), 831-843. https://doi.org/10.3233/JAD-200137

Quek, L. J. V., Heikkonen, M. R., & Lau, Y. (2023). Use of artificial intelligence techniques for detection of mild cognitive impairment: A systematic scoping review. *Journal of Clinical Nursing, 32*(17-18), 5752-5762. https://doi.org/10.1111/jocn.16699

Roeck, E. E. D., De Deyn, P. P., Dierckx, E., & Engelborghs, S. (2019). Brief cognitive screening instruments for early detection of Alzheimer's disease: A systematic review. *Alzheimer's Research and Therapy*, *11*(1), Article 21. https://doi.org/10.1186/s13195-019-0474-3

Rotenberg, S., Maeir, A., & Dawson, D. R. (2020). Changes in activity participation among older adults with subjective cognitive decline or objective cognitive deficits. *Frontiers in Neurology*, *10*, Article 1393. https://doi.org/10.3389/fneur.2019.01393

Tan, J. P., Wang, X., Zhang, S., Zhao, Y., Lan, X., Li, N., Wang, L. N., & Gao, J. (2021). Accuracy of the short-form montreal cognitive assessment Chinese versions. *Frontiers in Aging Neuroscience*, *13*, Article 687824. https://doi.org/10.3389/fnagi.2021.687824

Wittich, W., Phillips, N., Nasreddine, Z. S., & Chertkow, H. (2010). Sensitivity and specificity of the Montreal cognitive assessment modified for individuals who are visually impaired. *Journal of Visual Impairment and Blindness*, *104*(6), 360-368. https://doi.org/10.1177/0145482x1010400606

Zhou, Y., van Campen, J., Hortobágyi, T., & Lamoth, C. J. (2022). Artificial neural network to classify cognitive impairment using gait and clinical variables. *Intelligence-based Medicine*, *6*, Article 100076. https://doi.org/10.1016/j.ibmed.2022.100076