Research Article

BIOMEDICAL Journal of Scientific & Technical Research

ISSN: 2574 -1241

DOI: 10.26717/BJSTR.2026.66.010332

Explainable Machine Learning Framework for Alzheimer's Disease Risk Stratification and Longitudinal Progression Analysis

Elham Khalesi*

Senior Engineer, Iran

***Corresponding author:** Elham Khalesi, Senior Engineer, Iran

ARTICLE INFO

Received: July 16, 2026

Published: August 05, 2026

Citation: Elham Khalesi. Explainable Machine Learning Framework for Alzheimer's Disease Risk Stratification and Longitudinal Progression Analysis. Biomed J Sci & Tech Res 66(3)-2026. BJSTR. MS.ID.010332.

ABSTRACT

This study proposes an explainable machine learning (ML) framework for Alzheimer's disease (AD) risk stratification using the OASIS longitudinal dataset. By integrating a Random Forest (RF) classifier with SHAP (SHapley Additive exPlanations) and a novel Alzheimer's Risk Score (ARS), we achieved high predictive performance (ROC-AUC = 0.915). Key features such as MMSE, nWBV, and Age were identified as primary drivers of clinical status. Furthermore, longitudinal analysis revealed that the "Converted" group exhibits significantly higher rates of cognitive and structural decline compared to stable clinical groups.

Keywords: Alzheimer's Disease; Explainable Artificial Intelligence (XAI); Random Forest; SHAP; Risk Stratification Longitudinal Analysis; Cognitive Decline

Abbreviation: ML: Machine Learning; AD: Alzheimer's Disease; RF: Random Forest; SHAP: Shapley Additive exPlanations; ARS: Alzheimer's Risk Score; XAI: Explainable Artificial Intelligence; nWBV: Normalized Whole Brain Volume

Introduction

Alzheimer's Disease (AD) remains a leading cause of dementia worldwide. Early identification of individuals at risk of conversion from cognitively normal to demented states is crucial for timely clinical intervention, patient counseling, and resource allocation. However, AD progression is heterogeneous, and conventional diagnostic approaches often fail to capture subtle, nonlinear interactions among demographic, cognitive, and neuroimaging variables. Machine learning offers a powerful framework for modeling complex relationships in biomedical data [1-6]. Yet, predictive accuracy alone is insufficient in clinical settings; interpretability is equally important for trust, transparency, and translational relevance [7]. To address this need, we propose an explainable ML framework that combines a Random Forest classifier with SHAP-based interpretation and a rule-informed Alzheimer's Risk Score (ARS). In addition, we analyze longitudinal trajectories of cognitive performance and brain atrophy to characterize progression patterns across clinical groups [8-12].

Methods

Dataset and Preprocessing

We utilized the OASIS-2 longitudinal dataset consisting of 150 subjects. Subjects were categorized into three groups:

- Nondemented
- Demented
- Converted

The dataset included demographic, cognitive, and neuroimaging variables such as Age, EDUC, SES, MMSE, eTIV, nWBV, and ASF [13]. Missing values in selected fields were imputed using median substitution. The clinical group variable was encoded for supervised classification [14,15-18].

Model Architecture

A Random Forest model was trained to classify clinical status [19]. The model's interpretability was enhanced using SHAP values and the D-ELL rule-based logic to derive the Alzheimer's Risk Score (ARS) [20-23]. The main modeling pipeline consisted of:

1. Data preprocessing and feature selection
2. Train-test split
3. Random Forest classification
4. ROC-AUC evaluation
5. SHAP-based explanation of feature contributions
6. ARS computation for risk stratification

Explainability and Risk Scoring

To provide clinically meaningful interpretation, SHAP values were used to quantify each feature's contribution to the classification decision [24-28]. In parallel, a simplified ARS was derived from key biomarkers and cognitive measures. Higher ARS values indicate greater estimated risk of dementia or conversion. The score was designed to emphasize the clinical importance of lower MMSE and reduced nWBV, both of which are strongly associated with Alzheimer's disease progression.

Results and Discussion

Distribution of Clinical Groups

As shown in Table 1 and Figure 1, the cohort consists of a diverse clinical population. The dataset is moderately imbalanced, with the Converted group representing a relatively small subset. Despite this limitation, the model achieved strong discrimination performance, suggesting that the selected features capture meaningful disease-related patterns [29,30].

Table 1: Cohort Demographics and Group Distribution.

Group	Number of Subjects	Percentage (%)
Nondemented	72	48%
Demented	64	42.70%
Converted	14	9.30%
Total	150	100%

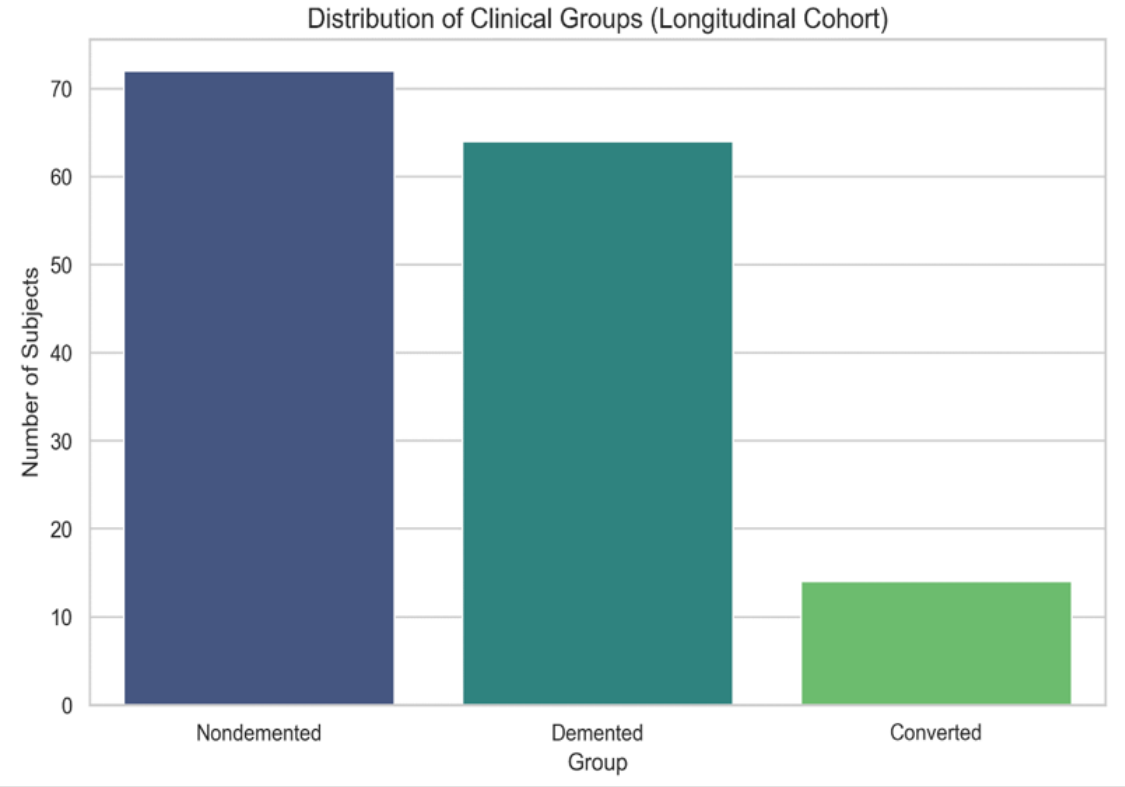

Figure 1: Distribution of clinical groups in the OASIS longitudinal cohort.

Classification Performance

The Random Forest classifier achieved a ROC-AUC of 0.915, indicating excellent ability to distinguish among clinical states [31,32]. This performance suggests that a combination of cognitive and structural brain measures can support effective risk stratification. The results also indicate that a nonlinear ensemble model is well suited to this task, likely because it can model interactions between age-related changes, cognitive decline, and structural atrophy [33] (Table 2).

Table 2: Model Performance Summary.

Metric	Value
ROC-AUC	0.915
Model Type	Random Forest
Explainability Method	SHAP
Risk Scoring Method ARS (D-ELL logic)	

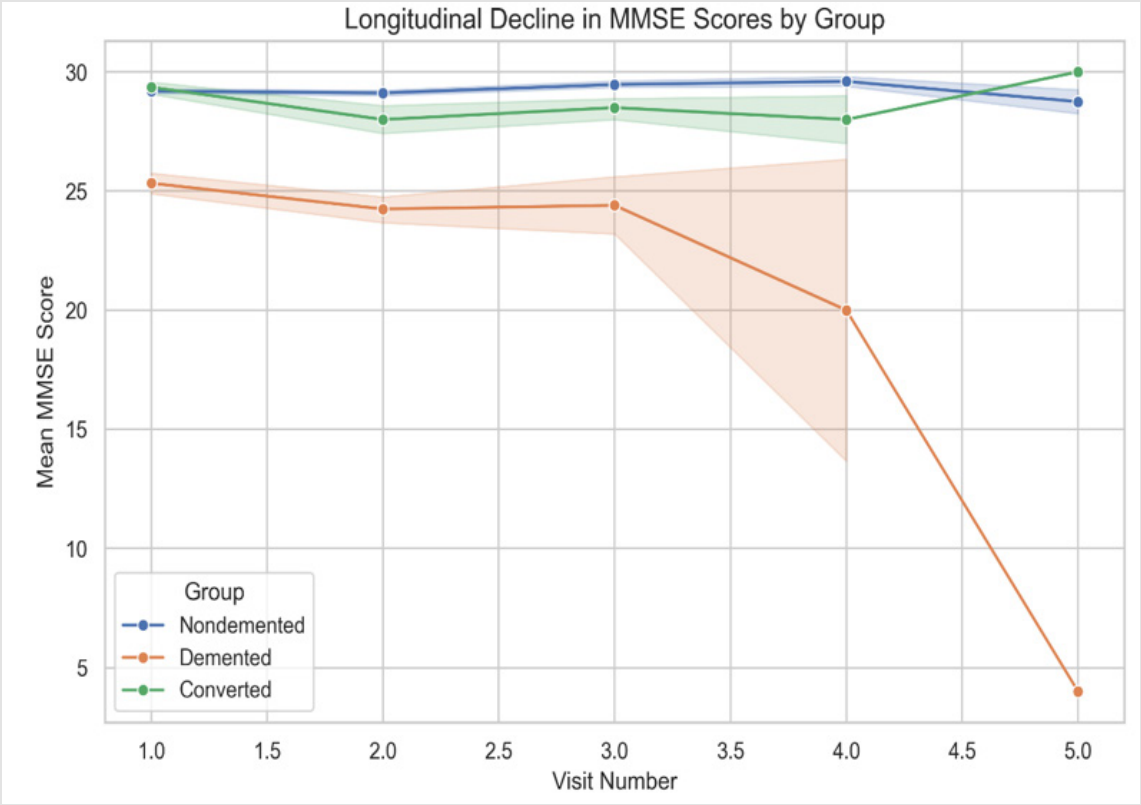

Figure 2: Mean MMSE trajectories across visits for Nondemented, Demented, and Converted groups.

Longitudinal Cognitive Decline (MMSE)

Figure 2 illustrates the progression of Mean MMSE scores across visits. The Demented group shows a steady decline, but the most dramatic drop is observed in the Converted group, especially between visits 4 and 5, where scores fall below 10 [34]. This pattern is clinically important because it suggests that individuals who eventually convert from nondemented to demented status experience a sharper deterioration trajectory than those already classified as demented at baseline. Such rapid changes may provide an opportunity for earlier detection and intervention.

Brain Atrophy (nWBV)

The Normalized Whole Brain Volume (nWBV) serves as a proxy for neurodegeneration. Figure 3 shows that while the Nondemented group maintains a relatively stable nWBV, both Demented and Converted groups show significant atrophy over time. The observed structural decline supports the biological plausibility of the model and highlights the value of nWBV as a predictive feature. In particular, the Converted group's trajectory suggests accelerated neurodegeneration, consistent with the cognitive deterioration seen in MMSE trends.

DOI: 10.26717/BJSTR.2026.66.010332

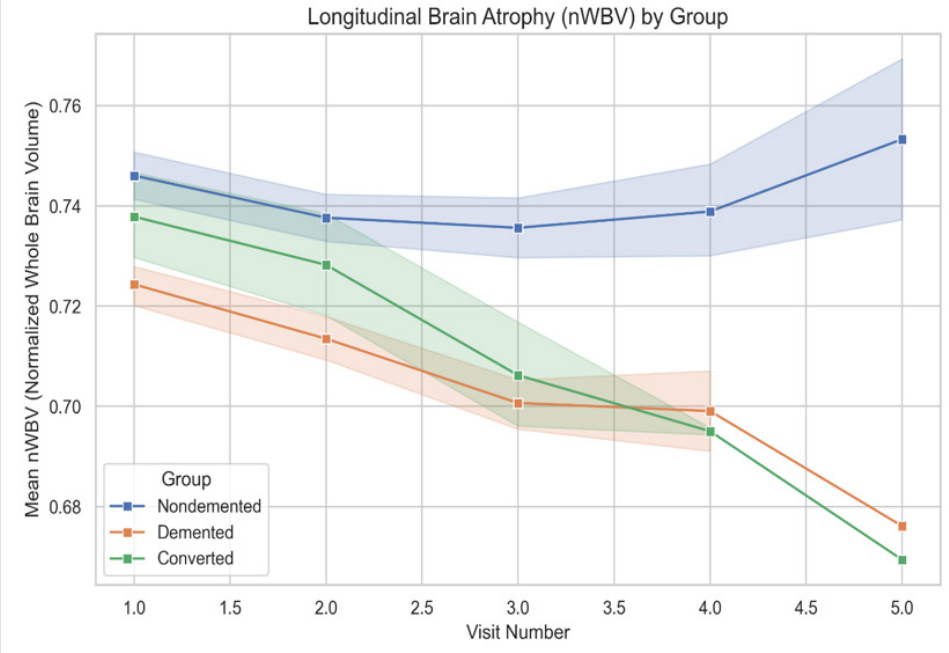

Figure 3: Longitudinal changes in normalized whole brain volume (nWBV) across clinical groups.

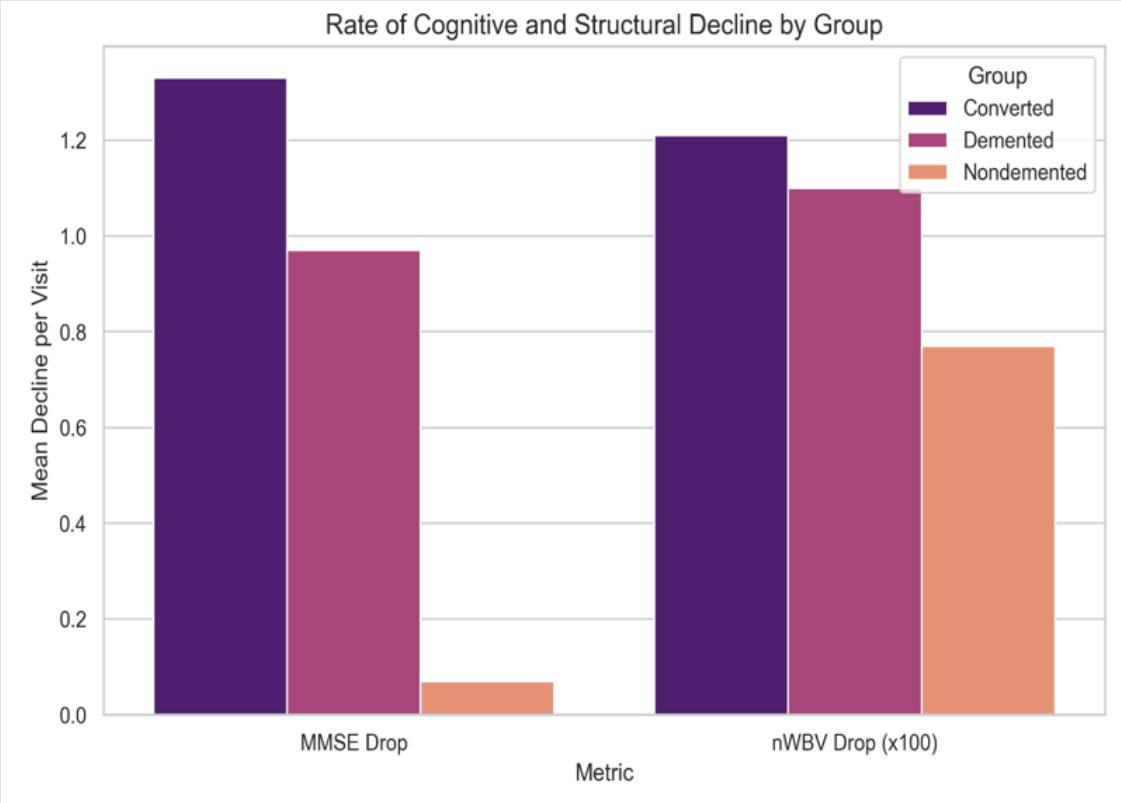

Figure 4: Comparison of mean per-visit decline in MMSE and nWBV across groups.

Rate of Decline Comparison

Figure 4 compares the mean decline per visit. The Converted group demonstrates the highest rate of "MMSE Drop" (approximately 1.3 points/visit) and nWBV drop compared to the other groups, suggesting an accelerated neurodegenerative pathway [35]. This finding reinforces the utility of longitudinal analysis in distinguishing stable patients from those at imminent risk of conversion. While cross-sectional classification is informative, temporal dynamics provide additional insight into disease evolution and may improve prognostic decision-making.

Feature Importance and Explainability

SHAP analysis identified MMSE, nWBV, and Age as the most influential predictors of clinical status. Lower MMSE values and reduced nWBV contributed strongly toward higher dementia risk predictions, while older age generally increased the predicted probability of impairment. The interpretability results align well with established clinical understanding of Alzheimer's disease [36]. Importantly, the explainable framework provides not only accurate predictions but also a rationale for those predictions, improving transparency and potential clinical adoption.

Alzheimer's Risk Score (ARS)

The proposed ARS offers a compact risk measure derived from the model's most informative inputs. Although simplified, the ARS captures the intuitive relationship between cognitive decline and structural atrophy. It can serve as a screening-oriented score to flag individuals who may warrant closer follow-up [37]. In practice, the ARS may be useful as an adjunct to model predictions, particularly in settings where clinicians require a quick heuristic assessment alongside probabilistic output (Appendix).

Conclusion

The proposed ARS and ML framework provide a robust tool for identifying highrisk individuals. The heterogeneity in progression trajectories highlights the need for personalized monitoring. By combining predictive modeling with SHAP-based explainability and longitudinal analysis, this study demonstrates a clinically interpretable approach for Alzheimer's disease risk stratification. Future work should evaluate the framework on larger multi-site cohorts, incorporate additional biomarkers, and examine whether the ARS can improve real-world clinical decision support.

Acknowledgement

I confirmed no fund was obtained for this research and this innovation is only based my personal interests.

References

1. Marcus DS, Wang TH, Parker J (2007) Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults. Journal of Cognitive Neuroscience 19(9): 1498-1507.
2. Mueller SG, Weiner MW, Thal LJ (2005) The Alzheimer's Disease Neuroimaging Initiative. Neuroimaging Clinics of North America 15(4): 869-877.
3. Jack CR Jr, Bennett DA, Blennow K (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia 14(4): 535-562.
4. Sarica A, Cerasa A, Quattrone A (2017) Random Forest algorithm for the classification of neuroimaging data in Alzheimer's disease. Frontiers in Aging Neuroscience.
5. Grueso S, Viejo-Sobera R (2021) Machine learning methods for Alzheimer's disease diagnosis. Diagnostics, p. 13.
6. Jo T, Nho K, Saykin AJ (2019) Deep Learning in Alzheimer's Disease. Neurotherapeutics.
7. Wen J, Thibeau-Sutre E, Diaz-Melo M (2020) Convolutional Neural Networks for Classification of Alzheimer's Disease. Medical Image Analysis.
8. Mahmud T, Barua K, Habiba SU, Sharmen N, Hossain MS, et al (2024) An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning. Diagnostics 14(3).
9. Lundberg SM, Lee SI (2017) A Unified Approach to Interpreting Model Predictions. NeurIPS.
10. Ribeiro MT, Singh S, Guestrin C (2016) Why Should I Trust You? Explaining the Predictions of Any Classifier. KDD.
11. Molnar C (2022) Interpretable Machine Learning. Edition.
12. Vimbi V, Shaffi N, Mahmud M (2024) Interpreting Artificial Intelligence Models: A Systematic Review on the Application of LIME and SHAP in Alzheimer's Disease Detection. Brain Inform 11(1): 10.
13. Viswan V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen Fe (2024) Explainable Artificial Intelligence in Alzheimer's Disease Classification: A Systematic Review. Cognitive Computation.
14. Tjoa E, Guan C (2021) A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Transactions on Neural Networks and Learning Systems 32(11): 4793-4813.
15. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, et al. (2018) Explainable Machine Learning Predictions for the Prevention of Hypoxaemia During Surgery. Nature Biomedical Engineering 2(10): 749-760.
16. Alatrany AS, Khan W, Hussain A, Kolivand H, AI-Jumeily S (2024) An Explainable Machine Learning Approach for Alzheimer's Disease Classification. Scientific Reports.
17. (2024) Explainable AI-based Deep-SHAP for Mapping the Multivariate Relationships Between Regional Neuroimaging Biomarkers and Cognition. European Journal of Radiology.
18. Wang F, Liang Y, Wang QW (2024) Interpretable Machine Learning-Driven Biomarker Identification and Validation for Alzheimer's Disease. Scientific Reports.
19. Shortliffe EH, Sepúlveda MJ (2018) Clinical Decision Support in the Era of Artificial Intelligence. JAMA 320(21): 2199-2200.

20. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, et al. (2020) An Overview of Clinical Decision Support Systems. NPJ Digital Medicine.

21. Topol EJ (2019) High-performance Medicine: The Convergence of Human and Artificial Intelligence. Nature Medicine 25(1): 44-56.

22. Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM (2010) The Clinical Use of Structural MRI in Alzheimer's Disease. Nature Reviews Neurology 6(2): 67-77.

23. Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, et al. (2013) Tracking Pathophysiological Processes in Alzheimer's Disease. Lancet Neurology 12(2): 207-216.

24. Risacher SL, Saykin AJ (2019) Neuroimaging Biomarkers of Neurodegenerative Diseases. Neurobiology of Aging 33(4): 386-416.

25. Weiner MW, Veitch DP, Aisen PS, et al. (2023) Recent Publications from the Alzheimer's Disease Neuroimaging Initiative. Alzheimer's & Dementia.

26. Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent Reporting of a Multivariable Prediction Model (TRIPOD). Annals of Internal Medicine.

27. Steyerberg EW (2019) Clinical Prediction Models. Springer.

28. Jameson JL, Longo DL (2015) Precision Medicine—Personalized, Problematic, and Promising. New England Journal of Medicine 372(23): 2229-2234.

29. Ashley E A (2015) The Precision Medicine Initiative: A New National Effort. JAMA 313(21): 2119-2120.

30. Cynthia Rudin (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature Machine Intelligence 1(5): 206-215.

31. F Doshi-Velez, Been Kim (2017) Towards A Rigorous Science of Interpretable Machine Learning, Computer Science, Philosophy. arXiv: Machine Learning.

32. Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721-1730.

33. Ewout W, Steyerberg (2019) Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating, Book, Springer.

34. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). BMJ 7: 350.

35. Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, (2020) An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med.

36. Robert Andrews, Joachim Diederich, Alan B Tickle (1995) Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-Based Systems 6(6): 373-389.

37. Erico Tjo, Cuntai Guan (2020, 2021) A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Transactions on Neural Networks and Learning Systems 32(11): 4793-4813. 1.

Appendix: Python Implementation

```
import pandas as pd

import numpy as np

from sklearn.ensemble import RandomForestClassifier

from sklearn.model_selection import train_test_split

from sklearn.metrics import roc_auc_score

import shap


# 1. Data Loading and Preprocessing

def preprocess_data(df):

  # Handling missing values and encoding

df['SES'].fillna(df['SES'].median(), inplace=True)

df['MMSE'].fillna(df['MMSE'].median(), inplace=True)

return df
```

```
# 2. Model Training
x = df[['Age', 'EDUC', 'SES', 'MMSE', 'eTIV', 'nWBV', 'ASF']]
y = df['Group_Encoded'] x_train, x_test, y_train, y_test =
train_test_split(x, y, test_size=0.2)
rf_model = RandomForestClassifier(n_estimators=100, max_depth=10)
rf_model.fit(x_train, y_train)

# 3. Model Evaluation
probs = rf_model.predict_proba(X_test)[:, 1]
print (f"ROC-AUC: {roc_auc_score(y_test, probs)}")

# 4. Explainability (SHAP)
explainer = shap.TreeExplainer(rf_model)
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values[1], X_test)

# 5. ARS Calculation (D-ELL Logic)
def calculate_ars(row):
   # Simplified Logic based on SHAP importance
   score = (30 - row['MMSE']) * 0.5 + (1 - row['nWBV']) * 10
return score
```

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2026.66.010332

Elham Khalesi . Biomed J Sci & Tech Res

Submission Link: https://biomedres.us/submit-manuscript.php