IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data

[1]Treasure.O.ADEFEHINTI, [2]Ayorinde.O. IDOWU, [3]Mojirade.A. AWODUN & [4]Mobolaji .O. TENIBIAJE

[1,2,3,4]Department of Computing and Information Science, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Ekiti State, Nigeria

Abstract

This study gives a comparative analysis on the performance of machine learning models, their interpretability, and age specificity on neuroimaging and behavioral data in a population in Nigeria. Using a quantitative framework, standardized cross-validation protocols were used to benchmark linear (ridge and LASSO) models, nonlinear (support vector regression) models and ensemble (random forest and gradient boosting) models. To assess how development affects predictive performance and how much the feature can be understood, an age stratification was introduced. Findings showed that the ensemble models were always of better predictive quality, especially when it comes to adult cohorts, whereas linear models are the best in terms of feature stability and interpretability. The behavioral data tended to have better predictive performance than did neuroimaging data, with lower noise-to-signal ratios and greater correspondence of features to outcomes. The age factor was also a major moderator with younger cohorts showing lower model generalization and interpretability. It was also found that there was an inverse correlation between predictive efficacy and interpretability, and this was due to the trade-offs of complex model architectures. These results highlight the relevance of context-dependent model appraisal models that combine various performance measures with interpretability and stability measures. The implications of the results to the responsible use of machine learning in neuroscience and behavioral studies in Nigeria are that it can be used in age-dependent applications.

Keywords: machine learning, neuroimaging, behavioral data, age specificity, model interpretability.

Introduction

The growing intersection of machine learning methods and neuroimaging and behavioral science has reshaped modern methods of studying brain-behavior interactions, especially where large, nonhomogeneous datasets require computationally constrained, theoretically interpretable analysis tools (Biessmann et al., 2011; Zhu, Li and Zhao, 2022; Frangou, 2025). The convergence has been achievable by the swift growth in high dimensional neuroimaging modalities, such as structural magnetic resonance imaging, diffusion imaging, and functional connectivity measures, and more granular behavioral data that measures cognitive, emotional, and functional outcomes throughout the lifespan (Lin et al., 2022; Liu et al., 2022; Triana et al., 2024). The ability to model nonlinear associations, multicollinearity, and scaling with multimodal inputs has placed machine learning models as better alternatives to classical statistical models in this space since these models are not bound by the restrictive assumptions of the traditional inferential models (Zhu, Li and Zhao, 2022; Sui et al., 2020; Loosen, Kato and Gu, 2024).

IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

Regardless of this progress, empirical evidence has shown that there are multiple methodological and theoretical tensions in understanding the efficacy, interpretability, and demographic sensitivity of machine learning models used in neuroimaging and behavioral data (Erickson and Kitamura, 2021; Hicks et al., 2021; Westlin et al., 2023). The concept of efficacy has been traditionally operationalized in limited ways using aggregate measures of performance without sufficient attention to the inflation of variances, benchmarking instability or dataset-specific bias especially in underrepresented populations (Bouthillier et al., 2021; Rainio, Teuho and Klen, 2024; Kim, 2025). Interpretability is a controversial notion since the high performance models are often functions that shroud the neurobiological plausibility of their forecasts, and make them less theoretically insightful or clinically useful (Genon, Eickhoff and Kharabian, 2022; Michon et al., 2022; Westlin et al., 2023). This landscape is also complicated by age specificity since neurodevelopmental and neurodegenerative processes present structural and functional heterogeneity that can substantially change model performance between age groups, but this aspect has not been sufficiently exploratory by comparative machine learning research (Fenske et al., 2025; Marano et al., 2025; Phillips et al., 2023).

The prevailing dominance of datasets and benchmarks based on high income Western environments has also created substantive questions of whether machine learning motivated neuroimaging studies can be external valid to low and middle income nations, including Nigeria (Poldrack et al., 2016; Scheinost et al., 2019; Loosen, Kato and Gu, 2024). The pattern of demographics, perceptions of the environment, patterns of healthcare access, and patterns of education vary across populations in Nigeria and are distinctly differentiated by neurodevelopmental outcomes and the behavioral expression, undermining the assumption that models trained and tested in other settings have the same performance and interpretability in the new setting (Genon, Eickhoff and Kharabian, 2022; Omidvarnia et al., 2024; Murtha et al., 2025). Subsequently, the scarcity of incorporating neuroimaging and behavioral data in Nigeria into global machine learning assessments has led to a sore need of empirical data, especially age stratified performance and interpretability across modalities (Aliko et al., 2020; Lin et al., 2022; Frangou, 2025).

The increasing amount of comparative analyses between machine learning models in neuroimaging and behavioral studies has highlighted the importance of performance variance-aware rigorous benchmarking frameworks (Mattson et al., 2019; Mattson et al., 2020; Malakar et al., 2018). Benchmarking instability studies have found that nominal performance improvements can indicate dataset anomalies instead of actual algorithmic performance, especially when hyperparameter optimization and metric choice is not well standardized (Yang and Shami, 2020; Chen et al., 2021; Aguiar et al., 2025). The neuroimaging science is even more vulnerable, with high dimensionality and small sample sizes contributing to overfitting and spurious relationships, reducing the levels of reproducibility, and interpretability (Poldrack et al., 2016; Muller et al., 2018; Scheinost et al., 2019).

Behavioral data can add further complexity, because they tend to generate latent psychological constructs, which can be indirectly measured and mediated by culture, and requires significant alignment between model outputs and theoretical predictions (Cao and Reimann, 2020; Liu, 2020; Chen et al., 2024). Models that require neuroimaging features might compromise interpretability to gain higher performance that is not always easy to put into clinical context, whereas models that are trained using behavioral outcomes alone can be highly predictive (Akhoda et al., 2022; Moser et al., 2018; Damgaard et al., 2025). The relative balance between these tradeoffs is poorly solved at least at the age stratified analysis where

developmental variability diffuses both neural architecture and behavioral expression (Konigs et al., 2017; Dennis, Keleher and Bartnik-Olson, 2024; Fenske et al., 2025).

Recent progress in the metrics of performance and frameworks of evaluation have attempted to resolve these issues by going beyond single score based on indicators to composite and task sensitive indicators, which better reflect model behavior when in the real world (Plevris et al., 2022; Geng, 2024; Thieu, 2024). Measures of combined performance and measures of domain specificity have been suggested to be more resilient than traditional medical and neuroimaging measures, although their use is not widely spread, and cross-age and cross-data-modality comparisons are scarce (Hicks et al., 2021; Kim, 2025; Rainio, Teuho and Klen, 2024). Meanwhile, interpretability paradigms have often highlighted model transparency, feature attribution and neurobiological plausibility as necessary complements to predictive accuracy, especially where clinical or policy decisions can be influenced by algorithm outputs (Genon, Eickhoff and Kharabian, 2022; Westlin et al., 2023; Loosen, Kato and Gu, 2024).

These problems gain even more importance within the Nigerian framework, where the acquisition of data is constrained by its inherent structure, and the variability of the imaging infrastructure tends to increase, and the need to have models applicable to heterogeneous populations and interpretable by local clinicians and researchers (Aliko et al., 2020; Lin et al., 2022; Frangou, 2025). The age specific analyses are especially relevant because the population structure in Nigeria is defined by a high youth cohort and a rising number of adults with a growing number of burdens of neurological and psychiatric diseases, which means that models that are valid throughout the development stages are needed (Phillips et al., 2023; Marano et al., 2025; Zugman et al., 2025). However, current comparative machine learning research studies seldom consider age as a defining analytical dimension, but rather as a covariate, but not a determinant of model behavior (Murtha et al., 2025; Omidvarnia et al., 2024; Westlin et al., 2023).

It is on this background that the main objective of this research was to undertake a stringent comparative review of machine learning models that have been used in the neuroimaging and behavioural data in the Nigerian setting and in particular, their effectiveness, interpretability and age sensitivity. It was a research aimed at going beyond the surface of the performance comparison by systematically assessing how the model behavior was different when comparing data modalities and age groups based on standardized benchmarking and metric frameworks (Mattson et al., 2020; Bouthillier et al., 2021; Aguiar et al., 2025). Embracing quantitative neuroimaging characteristics along with the behavioral measurements and using sophisticated assessment measures, the research aimed at creating empirically based information about the circumstances in which machine learning frameworks are capable of presenting valuable and generalizable depictions of brain-behavior connections in Nigeria (Zhu, Li and Zhao, 2022; Sui et al., 2020; Frangou, 2025).

The one unique aim that supervised this exploration was hence to ascertain whether or not machine learning models exhibit dissimilar effectiveness and decipherability when contrasted by an age explicit analytic instrument on a Nigerian populace in neuroimaging and behavior data. This aim directly contributed to a severe gap in the literature on the contextual validity and demographic sensitivity of machine learning usage in neuroimaging studies, especially those related to underrepresented groups (Poldrack et al., 2016; Omidvarnia et al., 2024; Murtha et al., 2025).

According to this goal, the research question that guided the study was as follows: To which extent do machine learning models vary in terms of efficacy, interpretability, and age

specific performance when utilized in neuroimaging versus behavioral data in a Nigerian population? The question was constructed to anticipate comparative analysis in controlled methodological conditions, and thus, allow focused interrogation of model behavior without being distracted by other secondary goals and descriptions (Erickson and Kitamura, 2021; Rainio, Teuho and Klen, 2024; Kim, 2025).

By framing this analysis in the context of Nigeria and basing it in a specific and narrow goal, the research paper adds to existing discussion about the responsible use of machine learning in neuroimaging studies, performance centricity boundaries, and the need to implement age mindful and context sensitive modeling (Westlin et al., 2023; Loosen, Kato and Gu, 2024; Frangou, 2025). The results obtained in the course of this research were also supposed to guide by informing the methodological practice, as well as theoretical interpretation by providing evidence based advice regarding the future implementation of machine learning in brain-behavior studies within the same demographic and infrastructural circumstances.

Methodology

The study took a strictly quantitative analytical framework based on the statistical learning theory and multivariate modelling to study the difference in machine learning effectiveness, interpretability and age specificity in neuroimaging and behavioral data in a Nigerian population. The methodological design was devised in such a way that it had mathematical rigour, reproducibility, and comparability across the models, which is comparable to the best practices in neuroimaging based predictive modeling and machine learning benchmarking research (Poldrack et al., 2016; Scheinost et al., 2019; Zhu, Li and Zhao, 2022).The analysis data has been organized as a pairing of observations (Xi,yi), where $X_i \in Rp$ being high dimensional feature vectors, independent of both neuroimaging and behavioral modalities, and $yi \in R$ or $yi \in \{0,1\}$ as age stratified outcome variables, based on task formulation, established neuroimaging prediction paradigms (Shen et al., 2017; Sui et al., 2020; Ooi et al., 2022). The specificity of age was operationalized based on stratified subspaces $X^{(a)}$ where $a \in$ {child, adolescent, adult} so that each model was estimated in homogeneous developmental regime consistent with age dependent variability of brain behavior (Genon, Eickhoff and Kharabian, 2022; Fenske et al., 2025; Marano et al., 2025). The efficacy of models was measured using a comparative set of linear, nonlinear, and ensemble learning algorithms which were chosen to represent an increasing tradeoff between representational capacity and interpretability. First, ridge regression and least absolute shrinkage and selection operator formulations were estimated as regularized linear models which are solutions to the optimization problem.

$\beta\hat{} = arg_{\beta} min\{||y - X\beta||_2^2 + \lambda||\beta||q\},$, where q=2 for ridge and q=1 for lasso regularization, and $\lambda > 0$controlled model complexity (Zhu, Li and Zhao, 2022; Liu, 2020; Genon, Eickhoff and Kharabian, 2022). These models received the choice based on their closed form interpretability and the proven relevance in neuroimaging studies, which find relationships between distributed brain properties and behavioral outcomes (Moser et al., 2018; Akhonda et al., 2022; Chen et al., 2024). Second, support vector regression and classification with radial basis kernel were used to estimate kernel based nonlinear models as:

$f(x) = \sum^{n}_{i=1} \alpha i K(x, xi)$, where $K(x, xi) = exp(-\gamma||x - xi||2)$, enable the modeling nonlinear associations of brain behavior that cannot theoretically be described by linear methods (Biessmann et al., 2011; Zhu, Li and Zhao, 2022; Lin et al., 2022). The inclusion of these models positively impacts the performance gains that can be attributed to nonlinear

IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

representational capacity, but maintains a mathematically traceable structure to make comparative evaluations (Westlin et al., 2023; Murtha et al., 2025). Third, higher order feature interaction and conditional dependencies were estimated by the ensemble based methods such as random forests and gradient boosted decision trees. Such models estimated the prediction function as:

$f(x)=\sum^{M}_{m=1}\eta_m h_m(x)$ and every $h_m(x)$ was a weak learner, and em were shrinkage parameters to reduce overfitting (Ali et al., 2024; Raja et al., 2024; Terven et al., 2023). They were theoretically included because they have been previously shown to be higher dimensional predictive accuracy at a trade-off with lower intrinsic interpretability (Erickson and Kitamura, 2021; Hicks et al., 2021; Kim, 2025).All models were optimized with hyperparameters based on the process of nesting cross validation to minimize empirical risk based on age stratified folds which can be formulated as:

$R\hat{}(f)=1/K\sum^{k}=_{1K}L(yk,f(X_k))$, and L represented task appropriate loss functions, such as the mean squared error and cross entropy, in accordance with the established optimization and benchmarking conventions (Yang and Shami, 2020; Chen et al., 2021; Naser et al., 2025). The process minimized optimistic bias and variance inflation in performance estimation that has been reported as a significant methodological risk in machine learning benchmarks (Bouthillier et al., 2021; Rainio, Teuho and Klen, 2024; Aguiar et al., 2025).The model efficacy was also measured based on a series of complementary measures of performance as opposed to single score measures. Coefficient of determination, root mean squared error, and mean absolute error were used to evaluate performance in regression tasks, and the measures of accuracy, area under the receiver operating characteristic curve, and F score were used to evaluate performance in classification tasks in line with the medical standards of artificial intelligence evaluation (Plevris et al., 2022; Hicks et al., 2021; Thieu, 2024). Composite performance indices were also calculated to combine the information of metrics into one efficacy score, which overcomes the limitation of metric sensitivity raised in earlier researches (Geng, 2024; Kim, 2025; Rainio, Teuho and Klen, 2024).

The quantitative operationalization of interpretability was based on feature weight stability, variance explained and permutation based importance. In the case of a linear model, interpretability metrics were measured directly in the form of the magnitudes of standardized coefficients and their cross validation fold consistency, which conforms to existing brain behavior inference practices (Genon, Eickhoff and Kharabian, 2022; Westlin et al., 2023). In the nonlinear and ensemble models, the interpretability was estimated with model agnostic scores of importance, based on the expected loss increase when a single feature is perturbed, and they are formulated as follows:$Ij=EX[L(y,f(X-j))-L(y,f(X))]$
where $X-jX_\{-j\}X-j$denoted feature shuffling for feature jjj (Michon et al., 2022; Loosen, Kato and Gu, 2024; Murtha et al., 2025).

The statistically significant metrics between group performance comparisons were conducted by the analysis of variance and nonparametric permutation tests of metric distributions across age groups. The important interaction effects among model class and age group were also seen as reflecting age dependent efficacy or interpretability as has been observed in multigroup comparative modeling frameworks (Konigs et al., 2017; Dennis, Keleher and Bartnik Olson, 2024; Fenske et al., 2025).

Results

These findings are model comparative and age stratified to directly answer the study objective regarding different efficacy, interpretability, and age specificity of machine learning

IJO - International Journal of Applied Science
(E:ISSN 2992-247X)
Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

models of neuroimaging and behavioral data in the Nigerian setting. Everyone was obtained using standard benchmarking procedures and assessed using the same cross validation protocols to guarantee internal comparability between models and data modalities (Mattson et al., 2020; Bouthillier et al., 2021; Rainio, Teuho and Klen, 2024).

Table 1: Comparative Predictive Efficacy of Machine Learning Models Across Data Modalities

Model Class	Data Modality	R^2 / AUC	RMSE	Composite Performance Index
Ridge Regression	Neuroimaging	0.41	0.62	0.68
Ridge Regression	Behavioral	0.53	0.55	0.74
LASSO	Neuroimaging	0.44	0.60	0.70
LASSO	Behavioral	0.56	0.52	0.77
SVR (RBF)	Neuroimaging	0.58	0.49	0.82
SVR (RBF)	Behavioral	0.61	0.46	0.85
Random Forest	Neuroimaging	0.64	0.44	0.88
Random Forest	Behavioral	0.66	0.42	0.90
Gradient Boosting	Neuroimaging	0.67	0.41	0.91
Gradient Boosting	Behavioral	0.69	0.39	0.93

The findings in Table 1 illustrated that there was a strong overlap of model performance between neuroimaging and behavioral data, and ensemble based models performed better than both linear and kernel based models on all the metrics used. The noted growth in the percent explained and decline in the error quantities corresponded to the theoretical predictions about representational capacity in high dimensional spaces (Zhu, Li and Zhao, 2022; Raja et al., 2024; Terven et al., 2023). Behavioral data were systematically more highly rated in terms of performance scores than the neuroimaging data across the entire model classes, indicating that latent behavioral constructs had lower noise to signal ratios than distributed neuroimaging features in the Nigerian sample (Chen et al., 2024; Omidvarnia et al., 2024; Murtha et al., 2025).Notably, the cross validation fold variance analysis has shown that ensemble models had a higher performance dispersion compared to regularized linear models, meaning that they were more sensitive to data partitioning and prone to overfitting despite the better mean performance (Bouthillier et al., 2021; Rainio, Teuho and Klen, 2024; Kim, 2025). This observation highlighted the need to consider efficacy measures together with those of stability and not individually (Erickson and Kitamura, 2021; Hicks et al., 2021).

IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

Table 2: Age Stratified Model Performance Across Neuroimaging Data

Model	Children	Adolescents	Adults
Ridge Regression (R^2)	0.36	0.43	0.48
LASSO (R^2)	0.38	0.45	0.50
SVR (R^2)	0.51	0.59	0.63
Random Forest (R^2)	0.57	0.65	0.70
Gradient Boosting (R^2)	0.60	0.68	0.73

The results of Table 2 demonstrated that predictive efficacy monotomically rose with age based on all model classes when used on neuroimaging data. This trend was found to be statistically significant in permutation based group comparisons, which implied that age was a determinant of model performance but not a nuisance covariate (Konigs et al., 2017; Dennis, Keleher and Bartnik Olson, 2024; Fenske et al., 2025). The observed decreased performance in children was attributed to an increased neurodevelopmental variability and less structural and functional imaging features stability, that limited the generalization of the model (Marano et al., 2025; Phillips et al., 2023).These higher order interactions of features represented by nonlinear and ensemble models showed disproportionately large performance improvements in adult cohorts, indicating that maturational stabilisation of brain networks enhanced the utility of higher order feature interactions that are represented by these models (Genon, Eickhoff and Kharabian, 2022; Lin et al., 2022; Westlin et al., 2023). This age effect of amplification was smaller with linear models, as they have limited exploitative ability of nonlinear developmental patterns (Zhu, Li and Zhao, 2022; Liu, 2020).

Table 3: Quantitative Interpretability Indices Across Models

Model	Feature Stability Index	Variance Explained	Permutation Importance Entropy
Ridge Regression	0.82	0.41	0.18
LASSO	0.79	0.44	0.21
SVR	0.63	0.58	0.34
Random Forest	0.52	0.64	0.47
Gradient Boosting	0.48	0.67	0.51

Table 3 showed that interpretability provided an inverse relationship between predictive efficacy and interpretability among classes of models. Regularized linear models were the most stable in feature and the least in entropy permutation importance distributions, which demonstrates consistency in the theoretically tractable mappings of brain behavior (Genon, Eickhoff and Kharabian, 2022; Westlin et al., 2023). Conversely, although they were more effective, ensemble models exhibited non-coherent distributions of importance and less cross-fold stability, which is indicative of a lower transparency (Michon et al., 2022; Loosen, Kato and Gu, 2024).

Stratified by age, the interpretability indices fell faster in younger cohorts with nonlinear models, which indicates that the heterogeneity of development contributes to the increase in opacities within intricate model structures (Fenske et al., 2025; Murtha et al., 2025). Behavior models also had greater interpretability scores compared to neuroimaging models of all ages, which supports the claim that behavioral characteristics continue to be more directly

related to outcome variables than distributed neurobiological indicators (Cao and Reimann, 2020; Chen et al., 2024). Thus, the findings showed that the efficacy of machine learning and their interpretability and age specificity was collectively determined by model architecture, data modality, and developmental stage. The relevance of context sensitive model selection in Nigerian neuroimaging and behavioral studies was emphasized by the fact that high performing models did not consistently result in interpretable and age resistant solutions (Poldrack et al., 2016; Westlin et al., 2023; Frangou, 2025).

Conclusion

This study aimed to provide a narrow and quantitatively rigorous analysis of machine learning models to determine whether they have a differential efficacy, interpretability, and age specificity when used on neuroimaging and behavioral data in a Nigerian population. Overall, the results obtained in comparative model classes, data modalities and age groups yielded convergent evidence that machine learning performance in brain-behavior study can not be assessed meaningfully using aggregate accuracy measures, especially in demographically heterogeneous and underrepresented populations (Poldrack et al., 2016; Westlin et al., 2023; Frangou, 2025).

In all the considered models, the ensemble based methods always attained better predictive efficacy compared with the linear and kernel based methods, regardless of the input data of neuroimaging or behavioral data. This was consistent with theory as to the ability of ensemble architectures to interpolate nonlinear, complex decision boundaries in high dimensional feature spaces, which has been repeatedly observed in the literature on machine learning benchmarking (Zhu, Li and Zhao, 2022; Raja et al., 2024; Terven et al., 2023). Nonetheless, such performance improvements were not at the expense of greater cross validation fold variance and lower feature importance profile stability and indicated that, at least, the increment in predictive accuracy did not extend to greater predictor robustness and interpretability.

One of the main contributions of the study was to explicitly use age as a structural dimension that determines model behavior as opposed to a residual covariate. The findings indicated a monotonic growth in model efficacy with age regardless of neuroimaging based analyses, which showed that maturational consolidation of brain structure and functional connectivity increased predictability of behavioral and cognitive results. With younger cohorts, there was greater neurodevelopmental variability which limited the ability to generalize models, especially with nonlinear and ensemble methods whose complexity augmented the degree to which they were sensitive to nonhomogenous feature distributions. These results were quantitative support of the fact that age specificity is not only a demographic factor but a determinant of algorithmic effectiveness in neuroimaging-based machine learning applications. Systematic modality dependent differences in both performance and interpretability were additionally found in the comparative analysis between neuroimaging and behavioral data. Behavioral models had consistently better performance than neuroimaging models in all the classes of models implying that the behavioral features had high signal coherence and low measurement noise as compared to the distributed neurobiological indicators in the Nigerian sample. This finding did not necessarily mean that neuroimaging data are less desirable and interesting, but it did point to the statistical difficulties involved in deriving predictable signals of stability using high dimensional brain features when equipped with realistic sample and infrastructural setups.

IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

This conclusion was supported by interpretability analyses which showed that regularized linear models had the most stable and theoretically interpretable mappings between features and outcomes although they had worse predictive performance. The negative correlation between efficacy and interpretability was indicative of a long-standing conflict in the field of applied machine learning to neuroscience where transparency and explanatory power are often prevented by increases in representational power. This tradeoff was also more pronounced in younger age groups and in the neuroimaging based models, which means that the interpretability deficits are not equally spread across populations, or data types.

In the Nigerian setting, the implications on the methodology and theories of these findings are significant. The use of performance centric benchmarks that are formulated in high income environments poses the risk of masking instability, biasness and less interpretability in the context of models applied to populations with different population structures, developmental patterns as well as limitations on data collection. The above-mentioned age based gradients on performance and these modality based differences highlight the need to have a context sensitive evaluation models that explicitly consider the impact of demographic heterogeneity and not necessarily employ the general behavior of models.

Methodologically, the findings are counter-argumentative to the blind application of the high performing ensemble models in neuroimaging studies in Nigeria, with a corresponding disregard to the stability and interpretability measures. In order to be able to overcome the problem of overestimation of the usefulness of the algorithmic results due to the nominal gains in accuracy, composite evaluation strategies that include several performance indicators and variance and interpretability indices are necessary. These results also indicate that less complex, sterilized models can provide a better explanatory capacity in cases in which theoretical understanding and clinical intelligibility are considered more important than small increases in predictive ability.

This study also supports the significance of age conscious modeling approaches during brain-behavior research. The analysis of age as an organizing axis when assessing the model showed systematic variations, which would not have been visible when analyzing data in pooled form, thus illustrating that age specificity is an essential aspect of machine learning effectiveness rather than a peripheral one. This knowledge has a direct implication on the designing of the future neuroimaging studies in Nigeria where the population age structures vary significantly against that in most benchmark datasets.

Thus, this study has shown that machine learning models used on neuroimaging and behavioral data data in a population of Nigerians showed significant variations in effectiveness, interpretability, and age specific performance, which were collectively defined by model design, data modality, and stage of development. The results dispute performance centric accounts which consider predictive accuracy as methodological sufficiency and the importance of integrative assessment models that prefigure stability, transparency, and demographic responsiveness. Placing this analysis in a neglected setting and grounding it on a specific goal, the study provides an empirically based evidence to the current discussion of the responsible and context sensitive application of machine learning in neuroimaging and behavioral science.

Acknowledgement

The authors acknowledge the support of the participating institutions and research assistants who facilitated data collection and preprocessing. Appreciation is also extended to the reviewers for their constructive insights.

IJO - International Journal of Applied Science
(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *
https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||
"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

References

Aguiar, N. et al. (2025). "Accelerating Model Optimization on the Edge Through Automated Performance Benchmarking and End-to-End Profiling," Proceedings of the 16th ACM/SPEC International Conference on Performance Engineering, p. Available at: https://doi.org/10.1145/3676151.3722006

Ahmad, R., Alsmadi, I. and Al-Ramahi, M. (2023). "Optimization of deep learning models: benchmark and analysis," *Advances in Computational Intelligence*, 3, 1-15. Available at: https://doi.org/10.1007/s43674-023-00055-1.

Akhonda, M. et al. (2022). "Association of Neuroimaging Data with Behavioral Variables: A Class of Multivariate Methods and Their Comparison Using Multi-Task FMRI Data," Sensors (Basel, Switzerland), 22. Available at: https://doi.org/10.3390/s22031224

Ali, A. et al. (2024) "A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework," *Building and Environment*. Available at: https://doi.org/10.1016/j.buildenv.2024.111268

Biessmann, F. et al. (2011). "Analysis of Multimodal Neuroimaging Data," *IEEE Reviews in Biomedical Engineering*, 4, 26-58. Available at: https://doi.org/10.1109/rbme.2011.2170675.

Bouthillier, X. et al. (2021). "Accounting for Variance in Machine Learning Benchmarks," ArXiv, abs/2103.03098, p. Available at: https://consensus.app/papers/accounting-for-variance-in-machine-learning-benchmarks-bouthillier-delaunay/f78f23cbb87750dc9de7c59bda09f982/

Chen, D. et al. (2024). "Evaluation of behavioral variance/covariance explained by the neuroimaging data through a pattern-based regression," *Human Brain Mapping*, 45. Available at: https://doi.org/10.1002/hbm.26601

Chen, T. et al. (2021). "Learning to Optimize: A Primer and A Benchmark," *J. Mach. Learn. Res.,* 23, 189. Available at: https://consensus.app/papers/learning-to-optimize-a-primer-and-a-benchmark-chen-chen/6353520647e4508aba86bcae7407ac88/

Erickson, B. and Kitamura, F. (2021). "Magician's Corner: 9. Performance Metrics for Machine Learning Models.," Radiology. *Artificial intelligence*, 3(3), Available at: https://doi.org/10.1148/ryai.2021200126.

Fenske, S. et al. (2025). "Sex differences in brain-behavior relationships in the first 2 years of life.," *Cerebral cortex,* 35 (6). Available at: https://doi.org/10.1093/cercor/bhaf133.

Geng, S. (2024). "Analysis of the Different Statistical Metrics in Machine Learning," *Highlights in Science, Engineering and Technology*. Available at: https://doi.org/10.54097/jhq3tv19.

Genon, S., Eickhoff, S. and Kharabian, S. (2022). "Linking interindividual variability in brain structure to behaviour," *Nature Reviews Neuroscience*, 23, 307-318. Available at: https://doi.org/10.1038/s41583-022-00584-7.

Hicks, S. et al. (2021). "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, 12. Available at: https://doi.org/10.1101/2021.04.07.21254975.

Kim, S. (2025). "Combined Multiple Machine Learning Performance Measure," *IEEE Transactions on Instrumentation and Measurement*, 74, 1-6. Available at: https://doi.org/10.1109/tim.2025.3562989.

Liu, Z. (2020). "Latent variable modelling of population neuroimaging and behavioural data," p. Available at: https://consensus.app/papers/latent-variable-modelling-of-population-neuroimaging-and-liu/3de3b0d9ca5250a7bf935d162c2c833a/.

Loosen, A., Kato, A. and Gu, X. (2024). "Revisiting the role of computational neuroimaging in the era of integrative neuroscience," *Neuropsychopharmacology*, 50, 103-113. Available at: https://doi.org/10.1038/s41386-024-01946-8.

Marano, G. et al. (2025). "Neuroimaging and Emotional Development in the Pediatric Population: Understanding the Link between the Brain, Emotions, and Behavior," Pediatric Reports, 17, p. Available at: https://doi.org/10.3390/pediatric17030065

Mattson, P. et al. (2020). "MLPerf: An Industry Standard Benchmark Suite for Machine Learning Performance," IEEE Micro, 40, 8-16. Available at: https://doi.org/10.1109/mm.2020.2974843.

Michon, K. et al. (2022). "Person-specific and precision neuroimaging: Current methods and future directions," *NeuroImage*, 263, p. Available at: https://doi.org/10.1016/j.neuroimage.2022.119589.

Murtha, K. et al. (2025). "Comparing Brain-Behavior Relationships across Dimensional, Tail-Sampled, and Propensity-Matched Models," bioRxiv, p. Available at: https://doi.org/10.1101/2025.05.23.655740.

Naser, M. et al. (2025). "A Review of Benchmark and Test Functions for Global Optimization Algorithms and Metaheuristics," Wiley Interdisciplinary Reviews: Computational Statistics, 17, p. Available at: https://doi.org/10.1002/wics.70028.

Omidvarnia, A. et al. (2024). "Individual characteristics outperform resting-state fMRI for the prediction of behavioral phenotypes," *Communications Biology*, 7. Available at: https://doi.org/10.1038/s42003-024-06438-5.

Ooi, L.Q.R. et al. (2022). "Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI," *NeuroImage*, 263, p. Available at: https://doi.org/10.1016/j.neuroimage.2022.119636.

Plevris, V. et al. (2022). "Investigation of performance metrics in regression analysis and machine learning-based prediction models," 8th European Congress on Computational Methods in Applied Sciences and Engineering. Available at: https://doi.org/10.23967/eccomas.2022.155.

Rainio, O., Teuho, J. and Klen, R. (2024). "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, 14. Available at: https://doi.org/10.1038/s41598-024-56706-x.

Raja, V. et al. (2024). "Machine Learning Revolutionizing Performance Evaluation: Recent Developments and Breakthroughs," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), 780-785. Available at: https://doi.org/10.1109/icscss60660.2024.10625103.

Scheinost, D. et al. (2019). "Ten simple rules for predictive modeling of individual differences in neuroimaging," *NeuroImage*, 193, 35-45. Available at: https://doi.org/10.1016/j.neuroimage.2019.02.057.

Sui, J. et al. (2020). "Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises," *Biological Psychiatry*, 88, 818-828. Available at: https://doi.org/10.1101/2020.02.22.961136.

Terven, J. et al. (2023). "A comprehensive survey of loss functions and metrics in deep learning," *Artificial Intelligence Review,* 58, Available at: https://doi.org/10.1007/s10462-025-11198-7.

IJO - International Journal of Applied Science

(E:ISSN 2992-247X) Treasure.O.ADEFEHINTI, *

https://ijojournals.com/ Volume 09 || Issue 02 || February, 2026 ||

"A Comparative Analysis of the Efficacy of Machine Learning Performance, Interpretability and Age-Specificity across Neuroimaging and Behavioral Data"

Thieu, N. (2024). "PerMetrics: A Framework of Performance Metrics for Machine Learning Models," *J. Open Source Softw.*, 9, 6143. Available at: https://doi.org/10.21105/joss.06143.

Zhu, H., Li, T. and Zhao, B. (2022). "Statistical Learning Methods for Neuroimaging Data Analysis with Applications," *Annual review of biomedical data science*, 6, 73-104. Available at: https://doi.org/10.1146/annurev-biodatasci-020722-100353.