<u>Original Paper</u>

Artificial Intelligence–Enhanced Multi-Algorithm R Shiny Application for Predictive Modeling and Analytics: Case Study of Alzheimer Disease Diagnostics

Han Wenzheng[1*], MS; Edmund F Agyemang[1*], MS; Sudesh K Srivastav[1], PhD; Jeffrey G Shaffer[1], PhD; Samuel Kakraba[1,2*], PhD

[1]Department of Biostatistics and Data Science, Celia Scott Weatherhead School of Public Health and Tropical Medicine at Tulane University, New Orleans, LA, United States
[2]Tulane Center for Aging, School of Medicine, Tulane University, New Orleans, LA, United States
*these authors contributed equally

Corresponding Author:

Samuel Kakraba, PhD
Department of Biostatistics and Data Science
Celia Scott Weatherhead School of Public Health and Tropical Medicine at Tulane University
1440 Canal St
New Orleans, LA 70112
United States
Phone: 1 5049882475
Email: skakraba@tulane.edu

Abstract

Background: Artificial intelligence (AI) has demonstrated superior diagnostic accuracy compared with medical practitioners, highlighting its growing importance in health care. SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) is an innovative AI-based application for Alzheimer disease (AD) prediction using handwriting analysis.

Objective: This study aimed to develop and evaluate a noninvasive, cost-effective AI tool for early AD detection, addressing the need for accessible and accurate screening methods.

Methods: The study used principal component analysis for dimensionality reduction of handwriting data, followed by training and evaluation of 10 diverse AI models, including logistic regression, naïve Bayes, random forest, adaptive boosting, support vector machine, and neural network. Model performance was assessed using accuracy, sensitivity, precision, specificity, F_1-score, and area under the curve (AUC) metrics. The DARWIN (Diagnosis Alzheimer With Handwriting) dataset, comprising handwriting samples from 174 participants (89 patients with AD and 85 healthy controls), was used for validation and testing.

Results: The neural network classifier achieved an accuracy of 91% (95% CI 0.79-0.97) and an AUC of 94% on the test set after identifying the most significant features for AD prediction. These performance results surpass those of current clinical diagnostic tools, which typically achieve around 81% accuracy. SMART-Pred's performance aligns with recent AI advancements in AD prediction, such as Cambridge scientists' AI tool achieving 82% accuracy in identifying AD progression within 3 years, using cognitive tests and magnetic resonance imaging scans. The variables "air_time" and "paper_time" consistently emerged as critical predictors for AD across all 10 AI models, highlighting their potential importance in early detection and risk assessment. To augment transparency and interpretability, we incorporated the principles of explainable AI, specifically using Shapley Additive Explanations, a state-of-the-art method to emphasize the features responsible for our model's efficacy.

Conclusions: SMART-Pred offers noninvasive, cost-effective, and efficient AD prediction, demonstrating the transformative potential of AI in health care. While clinical validation is necessary to confirm the practical applicability of the identified key variables, the findings of this study contribute to the growing body of research on AI-assisted AD diagnosis and may lead to improved patient outcomes through early detection and intervention.

Introduction

Artificial intelligence (AI) is revolutionizing medical diagnosis, offering unprecedented accuracy and efficiency in health care [1]. AI algorithms can analyze vast amounts of complex medical data, including images, patient records, and genetic information, to detect diseases earlier and more precisely than traditional methods [1-3]. In some cases, AI outperforms human practitioners in diagnostic accuracy, particularly in fields like radiology and pathology [4-8]. For instance, AI systems can identify subtle patterns in electrocardiograms to detect genetic heart conditions and analyze retinal images to diagnose diabetic retinopathy in its early stages [9-13]. By augmenting human expertise, AI not only improves diagnostic speed and accuracy but also has the potential to extend quality health care to remote areas, ultimately leading to better patient outcomes and more personalized treatment plans [1,3,4,14-17].

Alzheimer disease (AD) is a progressive brain disorder that gradually impairs memory and cognitive abilities, ultimately affecting the capacity to perform even the most basic tasks [18-22]. In most cases, especially those involving late-onset AD, symptoms usually appear in individuals during their mid-60s. Alzheimer disease and related dementias (ADRD) represent the primary cause of dementia among older adults [23-26]. In contrast, early-onset ADRD, which is relatively uncommon, can develop in people aged 30-65 years [27]. Aging is common to all cells and organelles in living organisms [28-31]. In recent times, there has been a disproportionate increase in the occurrence of ADRD, Huntington disease, Parkinson disease, amyotrophic lateral sclerosis, and other neurodegenerative diseases [32-35]. It is estimated that about 50 million people across the globe are living with ADRD [36,37]. ADRD are associated with a large family and caregiver burden. In 2018, the United States spent as much as US $277 billion on care for patients with ADRD. It is estimated that over 100 million people will be living with ADRD by the year 2050, for which the associated cost of treatment will be US $1200 billion [38-40]. Even though the prevalences of most diseases, such as breast cancer, prostate cancer, heart disease, stroke, and HIV, have decreased from 2000 to 2015, the prevalence of ADRD has increased by 123% within the same period [19].

Although many interventions, including novel nonsteroidal anti-inflammatory drugs and small molecules, have been explored for the treatment of ADRD [41-44], there is currently no cure for ADRD. Consequently, early diagnosis and the identification of ADRD biomarkers are critical to tailoring appropriate interventions for ADRD, reinforcing the need to design optimal methods for the prediction of ADRD and identify individuals at higher risk of ADRD in a timely manner [14,45-49].

The rapid growth of big data has necessitated the development of robust AI-driven algorithms for predictive modeling [50-58]. These algorithms are essential for efficiently handling large datasets and extracting relevant information through machine learning (ML) techniques [59]. Illustratively, ML algorithms, specifically hierarchical clustering, have been applied to visualize and examine the effects of point mutations that are associated with different phenotypes of diseases, like COVID-19, cystic fibrosis, and sickle cell anemia, resulting in an improved understanding of how point mutations lead to various variants of diseases [60-62]. AI and ML models have also found their place in disease prediction and risk analysis [54,63,64] for various diseases (neurodegenerative disorders, cardiovascular diseases, renal diseases, infectious diseases, diabetes, chronic diseases, neurological and psychiatric disorders, obesity, COVID-19, skin diseases, brain tumors, thyroid diseases, and liver diseases) [64-67]. However, no single algorithm can optimally solve every problem, highlighting the need for diverse ML approaches in predictive modeling and analytics. R, a widely adopted programming language for statistical data analysis, has become a powerful tool in data science [68]. The introduction of the Shiny package by RStudio revolutionized web development within the R ecosystem, making it accessible to both experienced data scientists and those with limited programming experience [69,70].

Methods

Development and Framework of the Shiny Multi-Algorithm R Tool for Predictive Modeling Application

SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) is an innovative AI-based R Shiny application developed to address the increasing demand for accessible data science tools and to simplify the complexity of ML workflows [67,71,72]. This user-friendly platform automates essential steps, such as data preprocessing, model training, hyperparameter tuning, model validation, and performance evaluation. It features a web interface that allows users to upload and preview datasets, configure variables, and compute descriptive statistics. SMART-Pred incorporates 10 classic ML algorithms, including logistic regression, random forest, adaptive boosting (AdaBoost), support vector machine (SVM), and neural network, for robust classification tasks. The tool also offers interactive data frames for comparing model evaluation metrics, alongside visualization capabilities like multimodel receiver operating characteristic (ROC) curve comparisons, feature importance ranking (by each selected algorithm saved as a CSV file in the same working directory in addition to the graphical display), and the Shapley Additive Explanations (SHAP) framework, which can provide the direction of influence of each feature. With one-click server deployment, SMART-Pred ensures accessibility anytime and anywhere, creating a convenient and intuitive environment for ML modeling that bridges the gap between complex AI

algorithms and practical applications, potentially accelerating research and innovation across data-driven fields.

The DARWIN (Diagnosis Alzheimer With Handwriting) dataset [73] was used to validate SMART-Pred's classification capabilities and benchmark its performance against prior studies. Specifically curated for early AD detection, the DARWIN dataset comprises handwriting data from 174 participants (89 patients with AD and 85 healthy controls) across 452 columns, including unique identifiers, classification labels ("P" for patient and "H" for healthy), and 451 features extracted from 25 structured tasks (graphic, copy, memory, and dictation exercises) with 18 features per task. Collected using standardized protocols (eg, Wacom tablets at 200 Hz), this dataset captures fine motor and cognitive variations linked to neurodegeneration, offering a balanced sample to minimize algorithmic bias and richer multidimensional data compared to single-task datasets. Unlike neuroimaging-focused alternatives like Alzheimer's Disease Neuroimaging Initiative (ADNI), the DARWIN dataset supports noninvasive, cost-effective, and immediately deployable screening in clinical settings, aligning with SMART-Pred's goal of accessible early AD detection [73, 74]. Its feature richness and task diversity mitigate overfitting risks and provide direct insights into cognitive-motor integration, a hallmark of AD progression, while prioritizing actionable, low-resource biomarkers over complex biological data, thus enabling it to offer pragmatic solutions for early diagnosis in diverse health care environments. The SMART-Pred application and data are provided in Multimedia Appendix 1.

The ultimate objective of this study is to leverage handwriting characteristics from the DARWIN dataset to distinguish between patients with AD and healthy individuals, contributing to earlier, more accessible diagnostic methods that can inform tailored interventions. We hypothesized that the integration of user-configurable and automated ML methods within the SMART-Pred workflow will enable the reliable classification of patients with AD versus healthy individuals based on handwriting features from the DARWIN dataset, demonstrating the potential of handwriting-based digital biomarkers for early, scalable AD detection. SMART-Pred's end-to-end ML workflow for classification tasks is structured across 4 core components: data input and preview, data preprocessing, model training and validation, and model evaluation. This framework integrates user-configurable options with automated processing to ensure flexibility and methodological rigor. The process starts with data upload and preprocessing, advances through model training with multiple algorithmic choices, and culminates in comprehensive performance evaluation and visualization, as detailed in the flowchart presented in Figure 1. By democratizing access to advanced ML techniques through an intuitive interface, SMART-Pred aims to enhance research capabilities and clinical applicability in AD detection and beyond.

Figure 1. Schematic workflow of the SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) application. The comprehensive workflow of the SMART-Pred application is organized into four principal stages: (1) uploading and previewing the dataset, (2) preprocessing data, (3) model training and validation, and (4) model evaluation. Graphical elements indicate process types: dotted boxes represent user-configurable steps (eg, data selection, feature handling, and model settings), while solid boxes denote system-fixed operations. The color scheme clarifies each stage's nature: pink for data-related processes, green for machine learning operations, and blue for model outputs. This visual summary highlights both the modular structure and the flexible customization available within SMART-Pred for streamlined, transparent predictive analytics. AdaBoost: adaptive boosting; PCA: principal component analysis; ROC: receiver operating characteristic; SHAP: Shapley Additive Explanations; XGBoost: extreme gradient boosting.
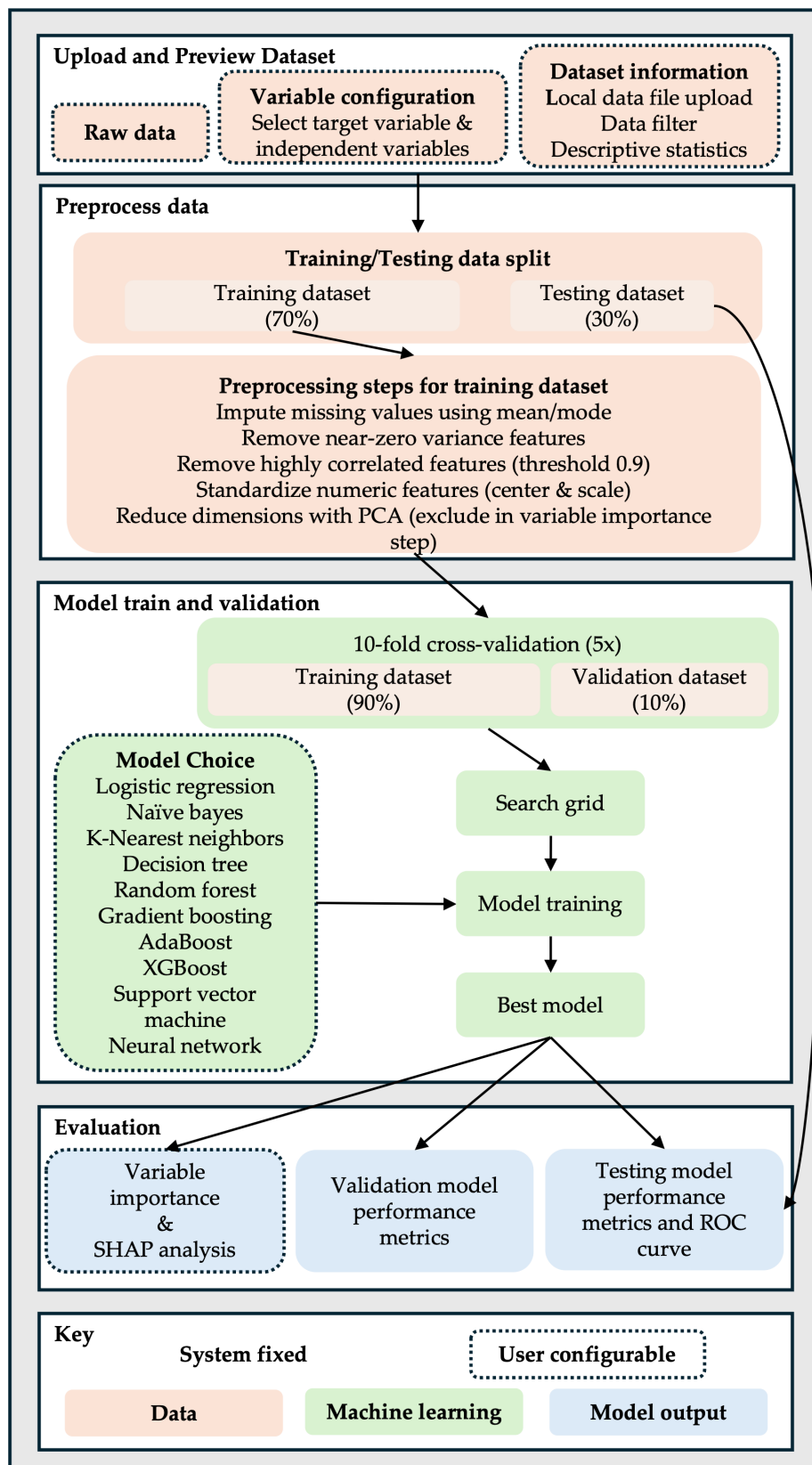
SHAP-Based Interpretability Method

SHAP is a model-agnostic interpretability technique grounded in cooperative game theory, which quantifies how much each feature contributes to a model's prediction by computing Shapley values, ensuring consistent and locally accurate explanations for individual outcomes. In its general, model-agnostic form, SHAP approximates these values using a weighted linear regression approach (Kernel SHAP), though specialized algorithms exist for specific model types. SHAP is applied to interpret predictions from all 10 ML classifiers trained on AD data, revealing the relative impact of each input feature on classification outcomes [75,76]. The explanation model g approximates the original complex model f by representing the prediction as a linear combination of feature coalitions given by:

$$g(z') = \phi_0 + \sum_{i=1}^{N} \phi_i z_i', \quad z_i' \in \{0, 1\}^N$$

where $z`$ denotes a binary vector indicating the presence or absence of features in the coalition, N is the total number of AD features, and ϕ_i corresponds to the Shapley value, reflecting the individual feature's contribution [77].

Through optimization, the Shapley value ϕ_i provides a theoretically grounded and consistent measure of feature importance, enhancing the interpretability of the ML classifier's predictions.

Data Input and Preview

SMART-Pred allows users to upload local data files through a web interface. Upon uploading, users can preview the dataset in an interactive table format (Figure 2) and access basic descriptive statistics (Figure 3). For descriptive statistics, the interface provides options for variable and measure configuration, where users can choose specific statistical measures: mean, SD, minimum, median, and maximum for numerical variables, and mode and unique count for character variables. After data upload, users proceed to configure the analysis parameters through a structured variable selection process (Figure 4). First, users specify the analysis type by selecting either "Classification" or "Regression" from the task type dropdown menu. For the target variable selection, users choose the dependent variable that represents the outcome to be predicted. The interface then presents a comprehensive list of available independent variables in a scrollable selection box, including various measurement features and other kinematic parameters. Users can either select individual variables or use the "Select all independent variables" checkbox for comprehensive analysis. This structured approach ensures proper specification of the prediction task and relevant features for subsequent analysis. The interface also provides checkboxes for selecting from 10 ML models (Figure 5).

Figure 2. Data input and preview interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The left panel shows data input controls, including CSV file upload. The right panel displays an interactive data preview table with pagination controls (10 entries per page) and search functionality. The interface enables users to configure analytic parameters while maintaining an overview of the dataset structure. ROC: receiver operating characteristic; SHAP: Shapley Additive Explanations.

Figure 3. Descriptive statistics interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The left panel includes variable and statistics selection controls: separate boxes for numeric variables (eg, air_time1 and disp_index1) and character variables (eg, ID and class), along with corresponding statistics selection boxes (mean and std for numerical variables; mode and unique count for character variables) and a "Calculate" button to run the process. The right panel displays the calculated statistics in a paginated table format with sortable columns, showing descriptive measures for all selected variables. The interface includes navigation tabs at the top for accessing different analysis views and a search functionality for filtering results. ROC: receiver operating characteristic; SHAP: Shapley Additive Explanations.
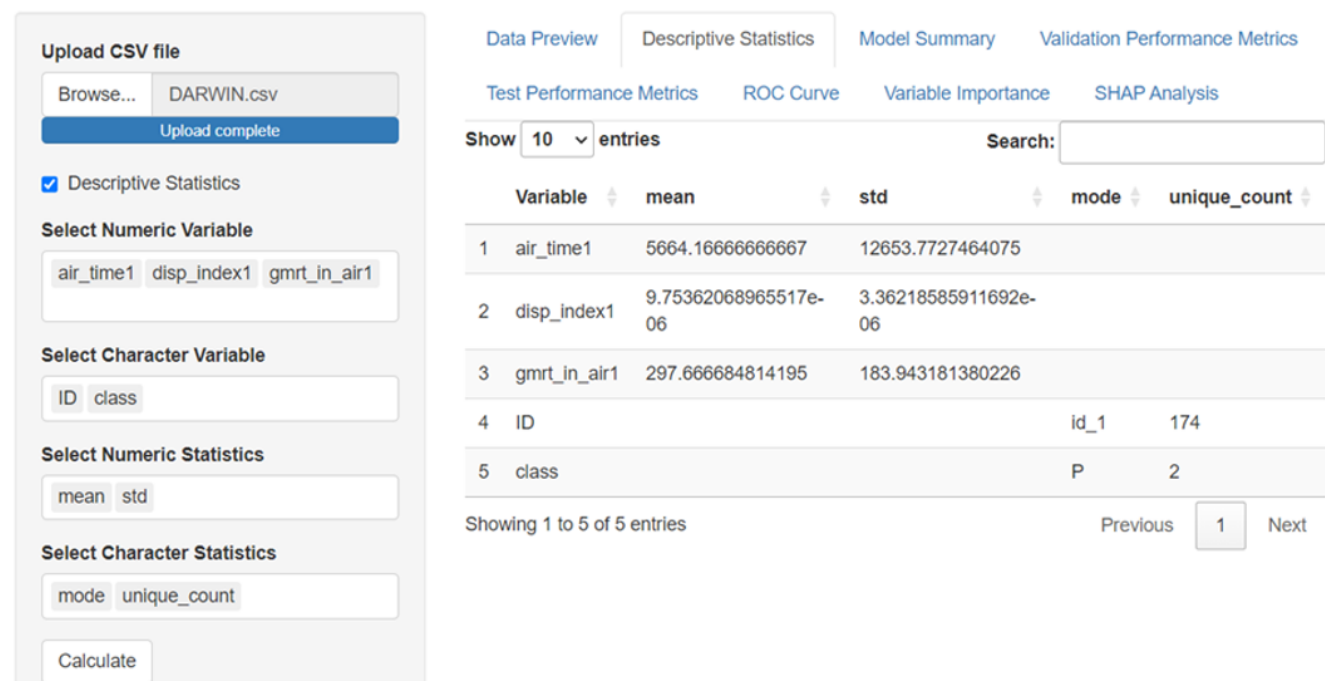
Figure 4. Task and variable selection interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The interface consists of 3 main components: a task type dropdown menu for selecting between classification and regression analysis, a target variable dropdown menu for specifying the dependent variable (eg, "class"), and a scrollable selection box displaying all available independent variables. Users can either select individual variables or use the "Select all independent variables" checkbox for batch selection. The interface displays variable names in a structured format, allowing users to view and select from all the features (eg, air_time1, disp_index1, and gmrt_in_air1). AdaBoost: adaptive boosting; KNN: k-nearest neighbors; SVM: support vector machine; XGBoost: extreme gradient boosting.

Figure 5. Model selection panel of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The interface provides checkboxes for selecting from 10 machine learning models: logistic regression, naïve Bayes, k-nearest neighbors (KNN), decision tree, random forest, gradient boosting, adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), support vector machine (SVM), and neural network. A "variable importance" option enables feature significance analysis, a "SHAP analysis" option estimates Shapley Additive Explanations (SHAP) values to emphasize the direction of the features responsible for the model's efficacy, and a "Run" button initiates the model training process. The panel allows multiple model selection for comparative analysis.

Data Preprocessing

The preprocessing pipeline (Figure 1) implements a systematic approach to data preparation. The preprocessing pipeline in SMART-Pred begins with a stratified 70-30 split to preserve class balance between patients with AD and controls, ensuring representative distributions in the training (70%) and test (30%) sets. For missing values, mean/mode imputation is systematically applied: numerical features use subgroup-specific means (eg, imputing based on AD stage or age cohort) to retain biological context, while categorical variables use mode replacement or an "Unknown" category to avoid information loss. This approach balances simplicity with clinical relevance, as arbitrary deletion would disproportionately affect smaller AD subgroups.

Principal component analysis (PCA) was strategically implemented to address the DARWIN dataset's high dimensionality (451 features), focusing on spatiotemporal handwriting variables like *air_time* and *paper_time*. PCA condensed these into orthogonal components that captured 95% cumulative variance, eliminating redundant kinematic features (eg, overlapping stroke velocity metrics) while preserving discriminative power for AD-related motor-cognitive patterns. The threshold of 95% variance retention was chosen to minimize noise from minor handwriting variations unrelated to neurodegeneration. Prior to PCA, feature filtering removed near-zero variance metrics (<5% unique values) and highly correlated variables (Pearson $r>0.9$), reducing multicollinearity that could distort component interpretation.

For non-DARWIN datasets with missing values, SMART-Pred uses context-aware imputation: numerical features use median values stratified by diagnostic status to avoid dilution of AD-specific patterns, while categorical variables adopt a hybrid approach-mode replacement for features with <10% missingness and "Missing" flagging for higher rates. This dual strategy ensures robustness across heterogeneous data sources while maintaining compatibility with the tool's multi-algorithm framework. The preprocessing workflow optimizes computational efficiency without sacrificing clinical interpretability, enabling focused analysis of neurodegeneration-sensitive handwriting biomarkers.

Model Training and Validation

The model training process uses a robust validation strategy involving 10-fold cross-validation repeated 5 times. The system has 10 classic ML algorithms (Table 1). The models range from traditional approaches (eg, logistic regression and naïve Bayes) to advanced ensemble methods (eg, random forest and gradient boosting) and deep learning (eg, neural network), and the selection has been driven by the need to balance algorithmic diversity, clinical utility, and

technical robustness. These models have been chosen to cover a broader spectrum of approaches: linear models (eg, logistic regression) provide interpretability for clinical stakeholders, ensemble methods (eg, random forest and AdaBoost) address nonlinear relationships while reducing overfitting, and deep learning models (eg, neural network) capture complex interactions. Criteria, such as accuracy, computational efficiency, and explainability, have guided the selection. Domain-specific validation ensures that the models excel at analyzing temporal handwriting dynamics and mitigates class imbalance. The inclusion of both high-precision neural networks and interpretable models optimizes the tradeoff between complexity and clinical transparency, while cross-validation confirms consistent performance. This multi-algorithm strategy ensures adaptability to diverse patient profiles and aligns with SMART-Pred's goal of delivering a robust, deployable tool for early AD detection. Detailed information about each algorithm's implementation and theoretical foundation has been provided previously [25-33]. Hyperparameter optimization is performed through a grid search process to identify the best model configuration.

Table 1. Machine learning models implemented in the SMART-Pred[a] application.

Algorithm	R function	Reference
Logistic regression	glm	[78]
Naïve Bayes	Naïve_bayes	[79]
K-nearest neighbor	knn	[80]
Decision tree	rpart	[81]
Random forest	rf	[82]
Gradient boosting	gbm	[83]
Adaptive boosting	ada	[84]
Extreme gradient boosting	xgbTree	[85]
Support vector machine	svmRadial	[80]
Neural network	nnet	[86]

[a]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.

Evaluation Output

The evaluation module provides a comprehensive assessment through multiple metrics and visualizations. The system generates validation model performance metrics (Figures 6 and 7), including accuracy [87], sensitivity [88], specificity [88], and F_1-score [89], for model selection; test model performance metrics (Figure 8), including accuracy [34], sensitivity [35], specificity [35], F_1-score [36], and ROC [37], for final evaluation; and ROC curves (Figure 9) for comparing model performance. Additionally, variable importance analysis (Figure 10) is conducted to identify key predictive features across different models. Lastly, SHAP analysis is conducted to interpret the predictions made by the 10 ML classifiers. All evaluation results are presented through interactive visualizations and data frames, allowing the examination of model performance from multiple perspectives.

Figure 6. Model summary interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) displaying detailed model training results. The interface presents comprehensive information for each trained model and an error message if a problem occurs during model training, including model type (eg, generalized linear model and naïve Bayes), dataset characteristics (eg, 121 samples and 451 predictors), preprocessing steps (eg, impute_mean, impute_mode, nzv, corr, center, scale, and pca), and cross-validation details (10-fold, repeated 5 times). Performance metrics are displayed in a structured format, showing receiver operating characteristic (ROC), sensitivity, and specificity values for each model's cross-validation results. SHAP: Shapley Additive Explanations.
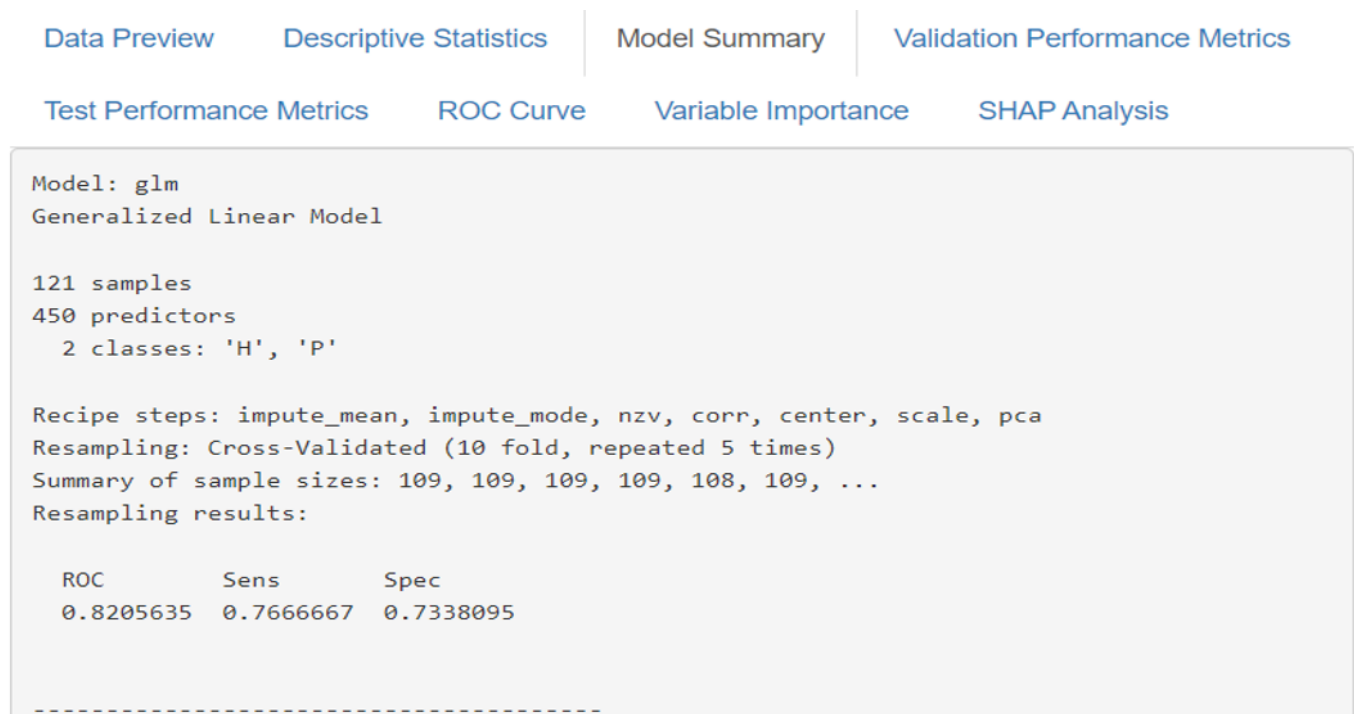
Data Preview   Descriptive Statistics   Model Summary   Validation Performance Metrics

Test Performance Metrics   ROC Curve   Variable Importance   SHAP Analysis

```
Model: glm
Generalized Linear Model

121 samples
450 predictors
  2 classes: 'H', 'P'

Recipe steps: impute_mean, impute_mode, nzv, corr, center, scale, pca
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 109, 109, 109, 109, 108, 109, ...
Resampling results:

  ROC        Sens       Spec
  0.8205635  0.7666667  0.7338095


-------------------------------------------------
```

Figure 7. Validation performance metrics comparison interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The table displays comprehensive evaluation metrics for 6 selected machine learning models (logistic regression, naïve Bayes, k-nearest neighbors, random forest, support vector machine, and neural network), including accuracy, recall, precision, specificity, F_1-score, and receiver operating characteristic (ROC) value. Each model's optimized hyperparameters are shown in the Best_Params column. The interface is accessible through the "Validation Performance Metrics" tab, allowing users to compare model performance across multiple evaluation criteria. SHAP: Shapley Additive Explanations.

Data Preview   Descriptive Statistics   Model Summary   Validation Performance Metrics

Test Performance Metrics   ROC Curve   Variable Importance   SHAP Analysis

Model	Accuracy	Accuracy (95% CI)	Recall	Precision	Specificity	F1_Score	ROC	Best_Params
glm	0.75	(0.712-0.783)	0.77	0.73	0.73	0.75	0.82	parameter=none
naive_bayes	0.67	(0.638-0.692)	0.94	0.60	0.41	0.73	0.87	laplace=0, usekernel=FALSE, adjust=1
knn	0.64	(0.614-0.659)	0.97	0.58	0.32	0.72	0.84	k=9
rf	0.78	(0.764-0.803)	0.83	0.75	0.74	0.79	0.90	mtry=41
svmRadial	0.72	(0.695-0.737)	0.85	0.66	0.59	0.74	0.86	sigma=0.00716678304819292, C=1
nnet	0.81	(0.803-0.824)	0.82	0.80	0.80	0.81	0.91	size=1, decay=0.1

Figure 8. Test performance metrics comparison interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The table displays comprehensive evaluation metrics for 6 machine learning models (logistic regression, naïve Bayes, k-nearest neighbors, random forest, support vector machine, and neural network), including accuracy, recall, precision, specificity, F_1-score, and receiver operating characteristic (ROC) values. Each model's optimized hyperparameters are shown in the Best_Params column. The interface is accessible through the "Test Performance Metrics" tab, allowing users to compare model performance across multiple evaluation criteria. SHAP: Shapley Additive Explanations.

Data Preview	Descriptive Statistics	Model Summary	Validation Performance Metrics				

Test Performance Metrics	ROC Curve	Variable Importance	SHAP Analysis				

Model	Accuracy	Accuracy (95% CI)	Recall	Precision	Specificity	F1_Score	ROC	Best_Params
glm	0.74	(0.597-0.847)	0.65	0.77	0.81	0.71	0.78	parameter=none
naive_bayes	0.64	(0.498-0.769)	0.85	0.59	0.44	0.70	0.82	laplace=0, usekernel=FALSE, adjust=1
knn	0.62	(0.479-0.752)	0.96	0.57	0.30	0.71	0.80	k=9
rf	0.79	(0.659-0.892)	0.77	0.80	0.81	0.78	0.89	mtry=41
svmRadial	0.83	(0.702-0.919)	0.81	0.84	0.85	0.82	0.83	sigma=0.00716678304819292, C=1
nnet	0.79	(0.659-0.892)	0.69	0.86	0.89	0.77	0.88	size=1, decay=0.1

Figure 9. Receiver operating characteristic (ROC) curve comparison interface of SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The plot displays the true positive rate against the false positive rate for 6 classification models, with each model represented by a different colored line. The diagonal reference line represents random prediction (area under the curve [AUC]=0.5). The interface is accessible through the "ROC Curve" tab, allowing users to compare model performance across multiple evaluation criteria. KNN: k-nearest neighbors; SVM: support vector machine.
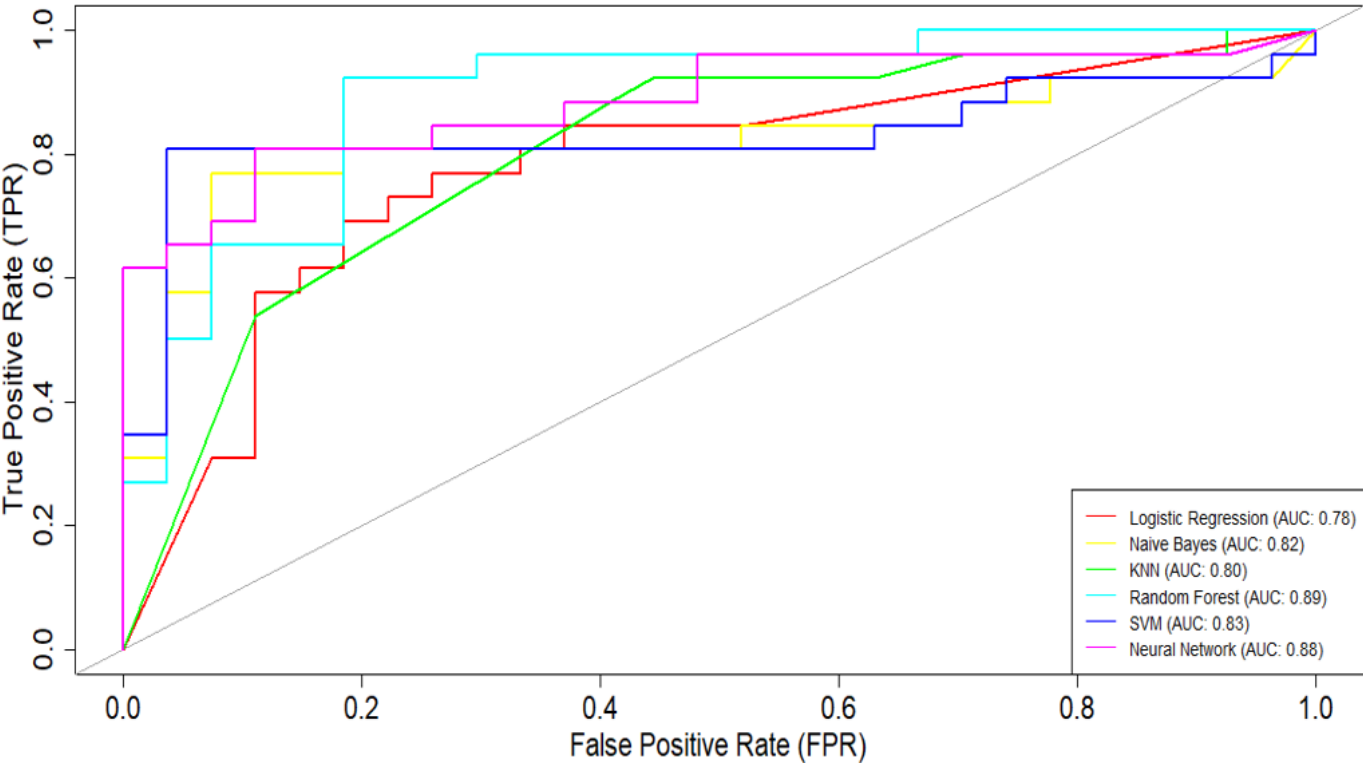
Figure 10. Top 10 variable importance visualization interface in SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The display consists of panels representing different models (eg, logistic regression, naïve Bayes, k-nearest neighbors [KNN], random forest, support vector machine [SVM], and neural network), with each showing bar charts of variable importance scores (y-axis, 0-100 scale) for the most influential predictors (x-axis). Each model type is distinguished by a unique color scheme, allowing for cross-model comparison of feature importance patterns. The interface is accessible through the "Variable Importance" tab, allowing users to compare variable importance across multiple evaluation criteria. ROC: receiver operating characteristic; SHAP: Shapley Additive Explanations.
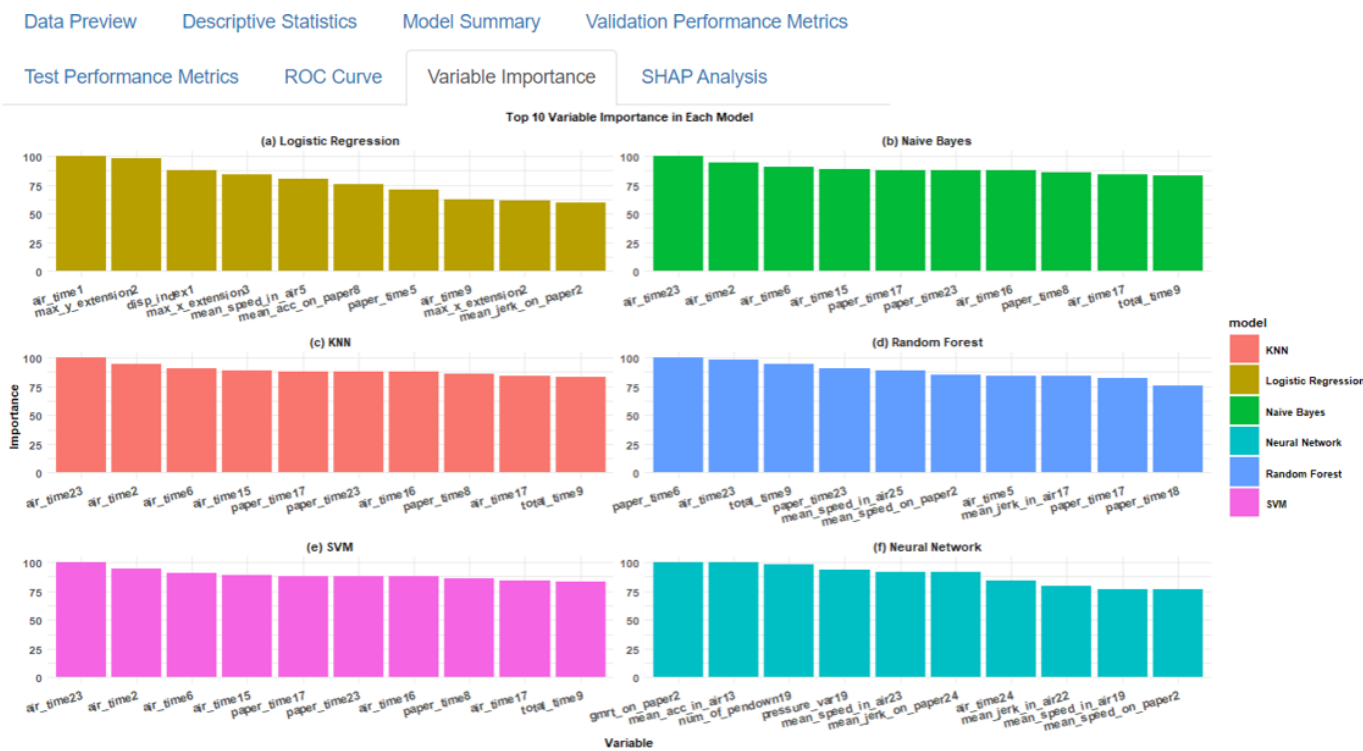
Ethical Considerations

This study used secondary, deidentified data and therefore did not require Institutional Review Board approval. All data used in this study have been anonymized to protect privacy.

Results

Preliminary Findings

In SMART-Pred, the validation results (Table 2) demonstrated the diverse array of models analyzed, revealing significant variations in performance metrics, such as accuracy, sensitivity, specificity, F_1-score, and area under the curve (AUC), for the DARWIN dataset among the models. The neural network and AdaBoost classifiers emerged as top performers, exhibiting high accuracy scores of 81% and 80%, respectively. Moreover, AdaBoost and neural network achieved commendable AUC scores (92% and 91%, respectively), suggesting a robust capability to distinguish between positive and negative AD cases. Conversely, models, such as k-nearest neighbors (KNN) and naïve Bayes, were noted for lower overall performance, revealing potential limitations in handling the DARWIN dataset's intricacies.

Table 2. Validation performance metric comparison across the 10 machine learning models in the SMART-Pred[a] application for the full feature model.

Model name	Accuracy, value (95% CI)	Recall	Precision	Specificity	F_1-score	AUC[b]
Logistic regression	0.75 (0.71-0.78)	0.77	0.73	0.73	0.75	0.82
Naïve Bayes	0.67 (0.64-0.69)	0.94	0.60	0.41	0.73	0.87
K-nearest neighbors	0.64 (0.61-0.66)	0.97	0.58	0.32	0.72	0.84
Decision tree	0.69 (0.67-0.71)	0.58	0.73	0.79	0.65	0.80
Random forest	0.78 (0.76-0.80)	0.83	0.75	0.74	0.79	0.90
Gradient boosting	0.80 (0.79-0.82)	0.81	0.80	0.80	0.80	0.91
Adaptive boosting	0.80 (0.79-0.81)	0.83	0.78	0.77	0.80	0.92
Extreme gradient boosting	0.80 (0.79-0.80)	0.80	0.79	0.79	0.79	0.91
Support vector machine	0.72 (0.70-0.74)	0.85	0.66	0.59	0.74	0.86
Neural network	0.81 (0.80-0.82)	0.82	0.80	0.80	0.81	0.91

[a]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.

Full Model With All Features

When the model was used for prediction in the test set (Table 3), the AdaBoost classifier stood out with a remarkable accuracy of 85%, a balanced sensitivity of 88%, an F_1-score of 85%, and an AUC of 92%, illustrating its ability to generalize well beyond the validation dataset. Such consistency between the validation and test phases highlights the robustness of the AdaBoost model for practical applications. The extreme gradient boosting (XGBoost)

classifier also demonstrated reliable performance across the dataset, maintaining stable accuracy and specificity, further reinforcing its candidacy for real-world deployment. Many other algorithms in our pipeline, such as the gradient boosting and random forest classifiers, achieved commendable evaluation scores across the evaluation metrics, reaffirming the robustness of the pipeline for classification-based predictive modeling.

Table 3. Test performance metric comparison across the 10 machine learning models in the SMART-Pred[a] application for the full feature model.

Model name	Accuracy, value (95% CI)	Recall	Precision	Specificity	F_1-score	AUC[b]
Logistic regression	0.74 (0.60-0.85)	0.65	0.77	0.81	0.71	0.78
Naïve bayes	0.64 (0.50-0.77)	0.85	0.59	0.44	0.70	0.82
K-nearest neighbors	0.62 (0.48-0.75)	0.96	0.57	0.30	0.71	0.80
Decision tree	0.72 (0.58-0.83)	0.77	0.69	0.67	0.73	0.73
Random forest	0.79 (0.66-0.89)	0.77	0.80	0.81	0.78	0.89
Gradient boosting	0.85 (0.72-0.93)	0.81	0.88	0.89	0.84	0.89
Adaptive boosting	0.85 (0.72-0.93)	0.88	0.82	0.81	0.85	0.92
Extreme gradient boosting	0.81 (0.68-0.91)	0.73	0.86	0.89	0.79	0.88
Support vector machine	0.83 (0.70-0.92)	0.81	0.84	0.85	0.82	0.83
Neural network	0.79 (0.66-0.89)	0.69	0.86	0.89	0.77	0.88

[a]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.
[b]AUC: area under the curve.

On integrating findings from both evaluation phases, it becomes evident that ensemble models (random forest, gradient boosting, AdaBoost, and XGBoost) hold significant promise due to their balanced performance metrics and strong generalization capabilities [90-94]. These models provide a compelling combination of sensitivity and specificity, which is crucial for the diagnostics required in AD detection. The observed fluctuations in the performance of the neural network classifier from validation to testing (especially for recall) suggest that while it possesses inherent strengths, there is room for enhancement.

Our comparative analysis underscores ensemble models (random forest, gradient boosting, AdaBoost, and XGBoost) as the most robust candidates for AD diagnostics in the scope of this work, given their superior performance across both the validation and test datasets [95-98]. These ensemble models have good, balanced accuracy in this regard, making them particularly suited for clinical settings where

precise prediction and identification are critical. The insights garnered from this study pave the way for continued exploration into model optimization, potentially enhancing the applicability and accuracy of ML techniques in terms of advancing AD research and the early diagnosis of other diseases, which can be considered crucial in tailoring interventions to target diseases.

Based on the ROC curves (Figure 11), the ensemble models (AdaBoost: AUC=0.92; random forest: AUC=0.89; gradient boosting: AUC=0.89; XGBoost: AUC=0.88) had the strongest performance among the 10 ML models considered in this study. The findings suggest that the AdaBoost ensemble model is highly effective in correctly identifying positive and negative instances for the given classification task. AdaBoost was closely followed by random forest and gradient boosting, both of which demonstrated impressive AUC values of 0.89. The XGBoost and neural network classifiers both achieved an impressive AUC value of 0.88.

Figure 11. Receiver operating characteristic (ROC) curves across different machine learning (ML) models obtained from SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The plot illustrates the relationship between the true positive rate (y-axis) and false positive rate (x-axis) across 10 different ML models. Each colored curve represents a distinct model's performance, with the area under the curve (AUC) values shown in the legend. The diagonal gray line represents the random classifier baseline (AUC=0.5). The closer the ROC curve is to the upper left corner of the graph, the higher is the accuracy of the test, as in the upper left corner, sensitivity is 1 and the false positive rate is 0 (specificity=1). The ideal ROC curve thus has an AUC of 1.0. AdaBoost: adaptive boosting; KNN: k-nearest neighbors; SVM: support vector machine; XGBoost: extreme gradient boosting.
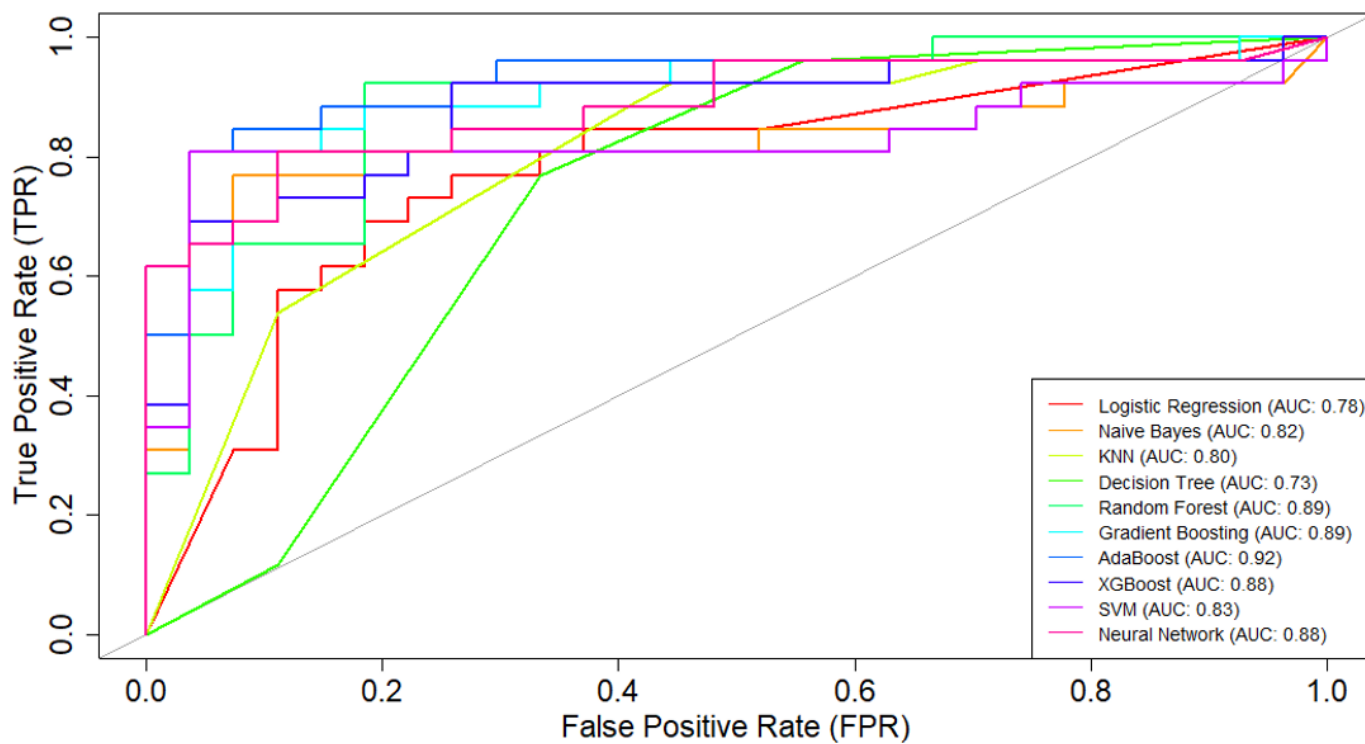
In contrast, the decision tree model exhibited the weakest performance, with an AUC of 0.73. While this is still better than random guessing, it suggests that the decision tree algorithm may not be the most suitable choice. The remaining models, including logistic regression, naïve Bayes, SVM, and KNN, had AUC values ranging from 0.78 to 0.83, indicating reasonably good predictive capabilities. The intersecting nature of the ROC curves highlights the better tradeoffs between different models [99-102]. Certain algorithms may excel at specific levels of the false positive rate, which can be an important consideration when selecting the most appropriate model for the task at hand [63].

Building upon our variable importance analysis of SMART-Pred (Figure 12), the partial variable set was constructed by selecting the most influential features among all the algorithms identified through our importance ranking analysis, including "air_time" and "paper_time" variables. Results for the reduced feature model analysis via SMART-Pred are shown in Tables 4 and 5.

Table 4. Validation performance metric comparison across the 10 machine learning models in the SMART-Pred[a] application for the reduced feature model.

Model name	Accuracy, value (95% CI)	Recall	Precision	Specificity	F_1-score	AUC[b]
Logistic regression	0.85 (0.82-0.88)	0.85	0.85	0.86	0.85	0.89
Naïve Bayes	0.85 (0.83-0.87)	0.87	0.83	0.84	0.85	0.93
K-nearest neighbors	0.83 (0.81-0.84)	0.95	0.76	0.71	0.84	0.93
Decision tree	0.82 (0.80-0.84)	0.77	0.85	0.87	0.81	0.89
Random forest	0.88 (0.87-0.90)	0.88	0.89	0.89	0.88	0.95
Gradient boosting	0.88 (0.87-0.89)	0.88	0.88	0.88	0.88	0.94
Adaptive boosting	0.89 (0.88-0.90)	0.89	0.88	0.88	0.88	0.96
Extreme gradient boosting	0.87 (0.87-0.88)	0.87	0.87	0.88	0.87	0.94
Support vector machine	0.86 (0.85-0.88)	0.85	0.87	0.87	0.86	0.93
Neural network	0.86 (0.85-0.87)	0.87	0.85	0.85	0.86	0.95

[a]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.
[b]AUC: area under the curve.

Table 5. Test performance metric comparison across the 10 machine learning models in the SMART-Pred[a] application for the reduced feature model.

Model name	Accuracy, value (95% CI)	Recall	Precision	Specificity	F_1-score	AUC[b]
Logistic regression	0.87 (0.74-0.95)	0.88	0.85	0.85	0.87	0.88
Naïve Bayes	0.81 (0.68-0.91)	0.85	0.79	0.78	0.81	0.84
K-nearest neighbors	0.81 (0.68-0.91)	0.92	0.75	0.70	0.83	0.90
Decision tree	0.81 (0.68-0.91)	0.88	0.77	0.74	0.82	0.81
Random forest	0.87 (0.75-0.95)	0.85	0.88	0.89	0.86	0.91
Gradient boosting	0.87 (0.75-0.95)	0.85	0.88	0.89	0.86	0.90
Adaptive boosting	0.85 (0.72-0.93)	0.81	0.88	0.89	0.84	0.91
Extreme gradient boosting	0.85 (0.72-0.93)	0.81	0.88	0.89	0.84	0.87
Support vector machine	0.89 (0.77-0.96)	0.85	0.92	0.93	0.88	0.92
Neural network	0.91 (0.79-0.97)	0.88	0.92	0.93	0.90	0.94

[a]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.
[b]AUC: area under the curve.

Figure 12. The top 10 variable importance findings across different models obtained from SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling). The subplots represent the following models: logistic regression, a basic classification algorithm; naïve Bayes, a probabilistic classifier based on Bayes' theorem; k-nearest neighbors (KNN), an instance-based learning algorithm; decision tree, a tree-structured classification model; random forest, an ensemble learning method based on multiple decision trees; gradient boosting, a boosting-based ensemble method; adaptive boosting (AdaBoost), an adaptive boosting algorithm; extreme gradient boosting (XGBoost), an efficient implementation of gradient boosting; support vector machine (SVM), a kernel-based classification algorithm; and neural network, a deep learning architecture. Each subplot represents a specific model's top 10 most important features, with normalized importance scores (y-axis, 0-100) plotted against feature names (x-axis). Different colors distinguish between models, and features are arranged in descending order of importance within each subplot. The importance scores are normalized within each model, with the most important feature scaled to 100.

Reduced Model With Important Features

The comparative analysis of the reduced feature model and full feature model showed that models trained on the selected partial variables (reduced feature model) consistently demonstrated equal or superior performance compared to those trained on all variables (full feature model), as seen in Tables 4 and 5. The comparison between the full and partial variable sets (reduced feature model) revealed patterns in both validation (Figure 13) and test (Figure 14) metrics.

Figure 13. Comparison of validation metrics between the full feature model (all variables included) and the reduced feature model (only the top 2 most important variables across all models in the full feature model) across machine learning models. Each subplot represents a different model, with 6 standardized performance metrics (accuracy, recall, precision, specificity, F_1-score, and area under the curve [AUC]) on the x-axis and their corresponding values (0-1.0) on the y-axis. Red bars represent models trained with all variables, while blue bars indicate models trained with the top 2 most important variables ("air_time" and "paper_time"). This visualization enables a direct comparison of model performance between the full and reduced feature sets across multiple evaluation metrics. AdaBoost: adaptive boosting; KNN: k-nearest neighbors; SVM: support vector machine; XGBoost: extreme gradient boosting.

**Figure 14.** Comparison of test metrics between the full feature model (all variables included) and the reduced feature model (only the top 2 most important variables across all models in the full feature model) across machine learning models. Each subplot represents a different model, with 6 standardized performance metrics (accuracy, recall, precision, specificity, F_1-score, and area under the curve [AUC]) on the x-axis and their corresponding values (0-1.0) on the y-axis. Red bars represent models trained with all variables, while blue bars indicate models trained with the top 2 most important variables ("air_time" and "paper_time"). This visualization enables a direct comparison of model performance between the full and reduced feature sets across multiple evaluation metrics. AdaBoost: adaptive boosting; KNN: k-nearest neighbors; SVM: support vector machine; XGBoost: extreme gradient boosting.

The pattern was particularly profound for models, such as SVM, naïve Bayes, KNN, and neural network, which showed substantial improvements across all metrics when using the selected important variables (most notable was the neural network's approximately 20% increment in the reduced feature model's test recall, from 69% to 88%). Ensemble methods like random forest, gradient boosting, AdaBoost, and XGBoost demonstrated slight improvements in performance across both variable sets, highlighting their inherent robustness regardless of feature selection [103,104]. Notably, the improvements observed in the validation metrics were largely mirrored in the test metrics, indicating that the benefits of using carefully selected important variables generalize well to unseen data.

The consistent enhancement across various performance indicators, including accuracy, recall, precision, F_1-score, and ROC values, in both the validation and test scenarios reinforces the effectiveness of our variable importance–based feature selection approach. These findings strongly validate our strategy of using variable importance ranking for feature selection, demonstrating that focusing on the most influential variables not only maintains or improves model performance but also potentially reduces computational complexity and simplifies the model optimization process [105-108]. In the reduced feature model, the SVM classifier achieved an accuracy of 89% (95% CI 0.77-0.96) with an AUC of 92%, while the neural network algorithm reached an accuracy of 91% (95% CI 0.79-0.97) with an AUC of 94%, after identifying the most significant features for AD prediction in the test set. The comparison of ROC curves is presented in Figure 15.

Figure 15. Comparison of receiver operating characteristic (ROC) curves obtained directly from SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) between models trained with the full feature set (A) and selected important features (B). Each subplot shows the true positive rate against the false positive rate for 10 machine learning models, with corresponding area under the curve (AUC) values in the legend. Different colors represent different models, and the diagonal gray line indicates random prediction (AUC=0.5). This side-by-side comparison demonstrates the impact of feature selection on model discrimination ability. The closer the ROC curve is to the upper left corner of the graph, the higher is the accuracy of the test, as in the upper left corner, sensitivity is 1 and the false positive rate is 0 (specificity=1). The ideal ROC curve thus has an AUC of 1.0. AdaBoost: adaptive boosting; KNN: k-nearest neighbors; SVM: support vector machine; XGBoost: extreme gradient boosting.

A
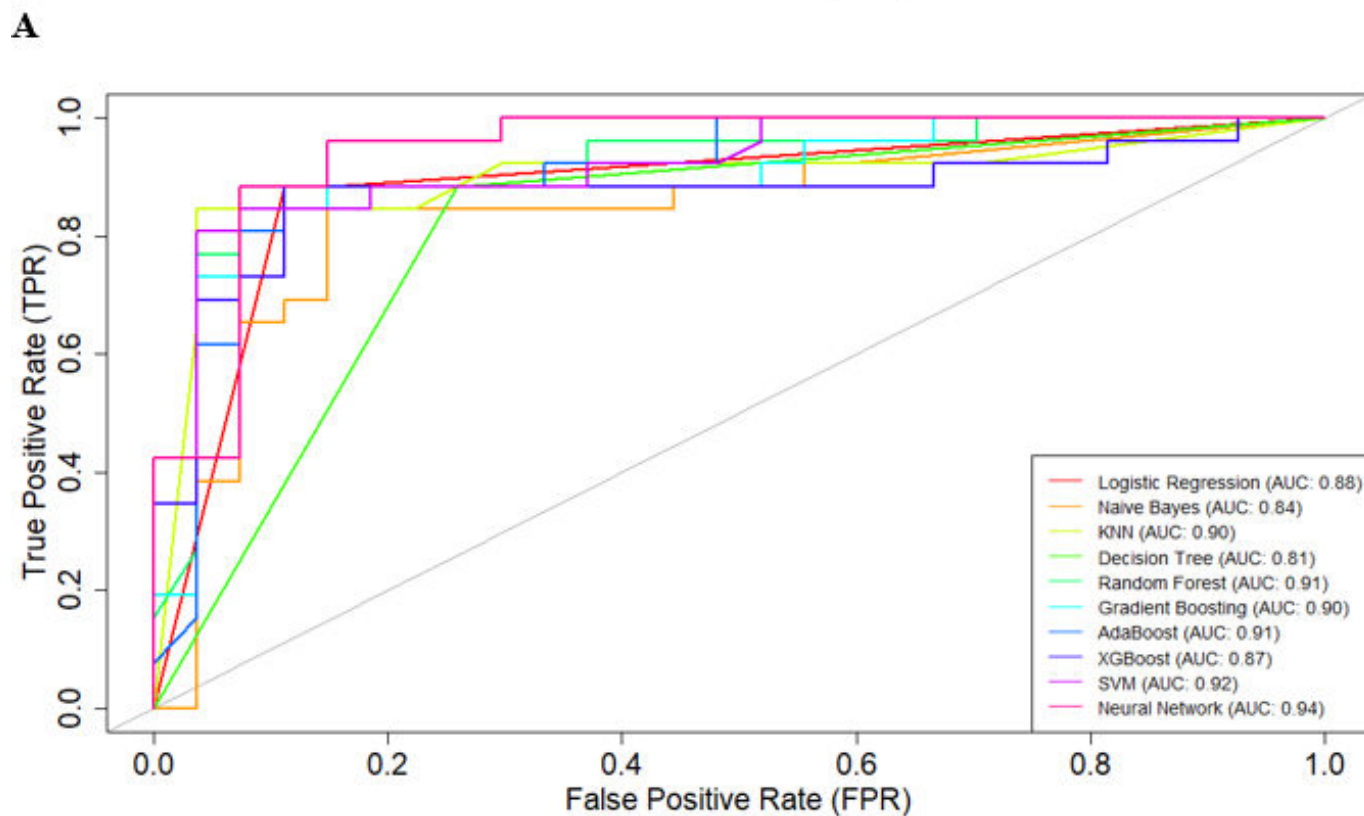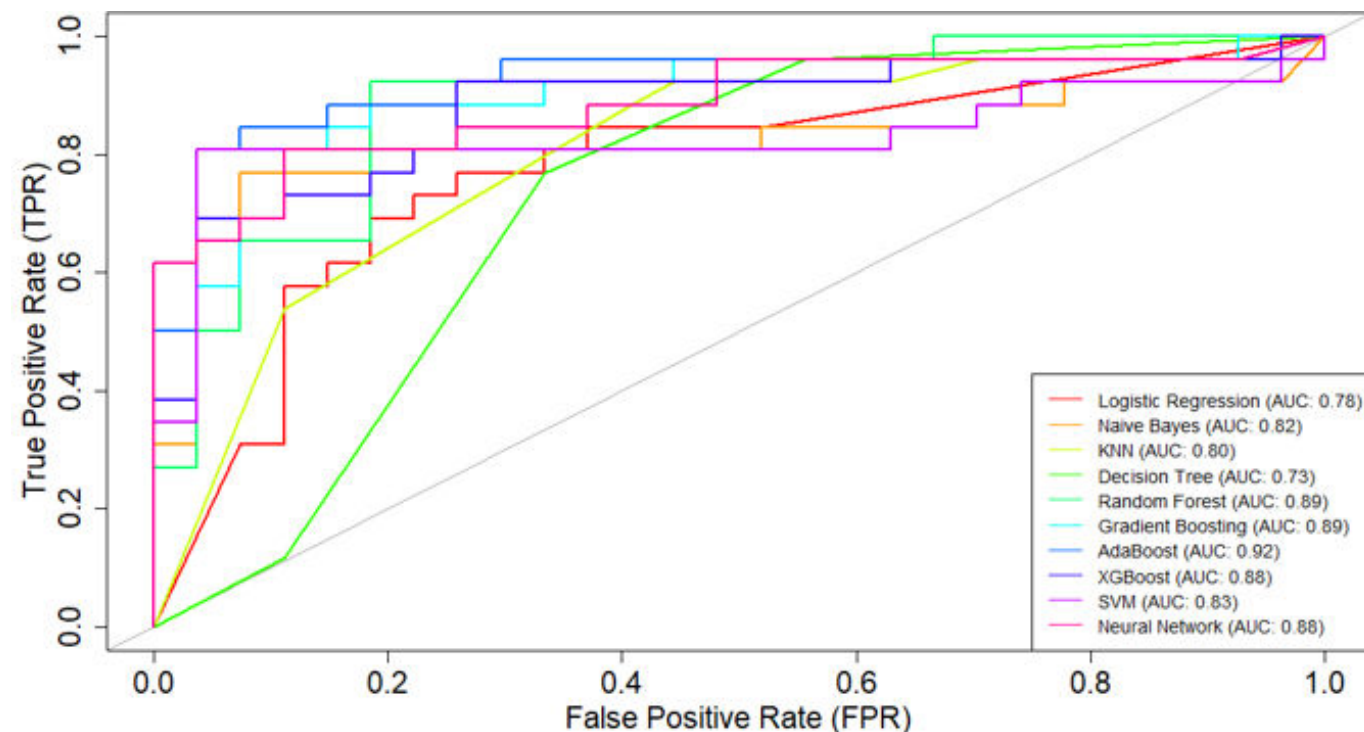
B

SHAP Analysis on the DARWIN Dataset

Interpreting the results and understanding the decision-making process of ML models has become paramount, especially in critical fields like medical diagnostics. In the context of AD detection, understanding which features or biological markers most influence a model's prediction (specifically direction of influence) can provide invaluable insights for medical professionals. In lieu of this, this study performed SHAP analysis. However, for brevity and ease of presentation, the SHAP analysis of the 2 best models (neural network and SVM classifiers) is presented in this section, while the SHAP analysis of the remaining 8 models has been presented in Multimedia Appendix 2.

The SHAP analysis of the neural network classifier indicated that *air_time24* and *air_time5* (mean SHAP values of –0.0029 and –0.0008, respectively) had the greatest negative influence on AD prediction, while *air_time2* and *paper_time8* (mean SHAP values of +0.0018 and +0.0012, respectively) had the greatest positive influence on AD prediction (Figure 16A).

Figure 16. Shapley Additive Explanations (SHAP) summary plots for (A) neural network and (B) support vector machine (SVM) classifiers. Each plot visualizes the impact of individual features on model output using SHAP values, which quantify both the magnitude and direction of each feature's contribution to the prediction. In these summary plots, features are ranked from top to bottom by their overall importance in the model's decision-making process. Each point represents a feature value for an instance in the dataset, with color gradients indicating whether the feature value is low (blue) or high (red). The horizontal axis shows the SHAP value, reflecting whether an increased feature value drives the model output higher (positive SHAP value) or lower (negative SHAP value). Comparisons between neural network and support vector classifier results in the figure highlight both consensus and differences in feature importance and decision logic across models. These SHAP visualizations provide insights into which features most influence model predictions, supporting the interpretability and transparency of the machine learning models used in this research.

A

B
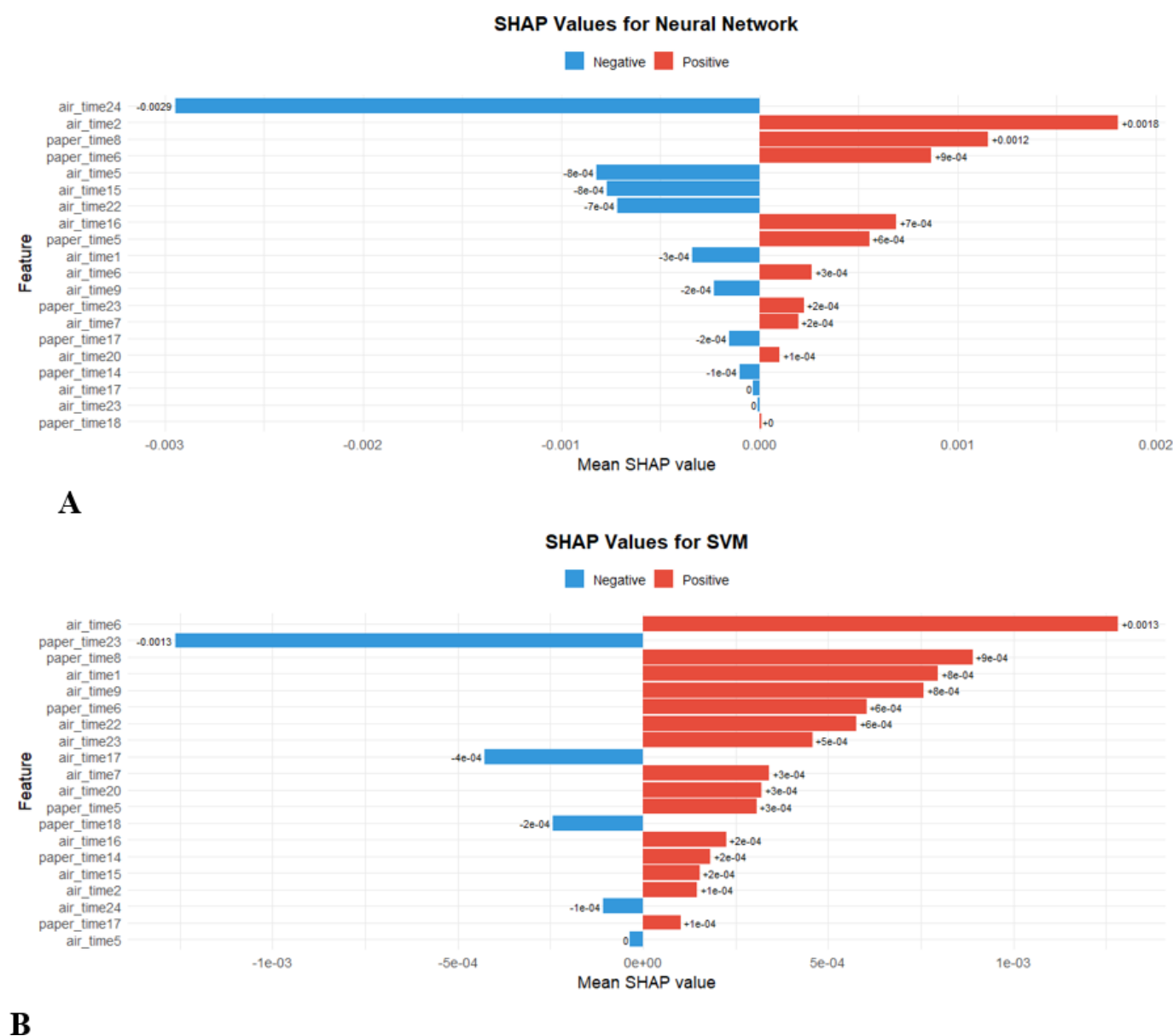
For the support vector classifier, *paper_time23* and *air_time17* (mean SHAP values of –0.0013 and –0.0004, respectively) had the greatest negative influence on AD prediction, while *air_time6* and *paper_time8* (mean SHAP values of +0.0013 and +0.0009, respectively) had the greatest positive influence on AD prediction (Figure 16B).

Discussion

Robustness of the SMART-Pred Application

AI is transforming medical practice and research by enhancing disease diagnostics, accelerating biomarker discovery, and revolutionizing disease surveillance [109-115]. In diagnostics, AI algorithms can examine complex medical data, such as imaging findings, genetic information, and clinical records, to identify subtle patterns that might be missed by human observers [116-120]. This capability leads to earlier and more accurate disease detection, potentially improving patient outcomes. In biomarker identification, AI's ability to process vast amounts of multiomic data is uncovering novel indicators of disease presence, progression, and treatment response [121,122]. These AI-discovered biomarkers are paving the way for personalized medicine approaches and more targeted therapeutic interventions. For disease surveillance, AI systems are proving to be invaluable in monitoring and predicting disease outbreaks on a global scale [123]. By analyzing diverse data sources, including social media, environmental factors, and health records, AI can identify emerging health threats in real-time, enabling rapid response and containment strategies [124,125]. Moreover, AI's impact extends to drug discovery, treatment optimization, and health care resource allocation, promising a future of more efficient, effective, and accessible health care delivery [126,127].

The SMART-Pred application demonstrates a robust, user-friendly, and customizable approach to variable importance analysis across multiple ML models. This innovative tool offers valuable insights into the predictive power of various features in disease diagnostics, biomarker identification, and disease surveillance, particularly in the context of AD, as demonstrated in this study. By using the "varImp()" function and SHP analysis in R in SMART-Pred, our study identified the top-most influential variables for each model, providing a comprehensive view of feature relevance in AD prediction.

Discussion of the Findings

A thorough analysis of variable importance across all models revealed consistent patterns and model-specific characteristics. The "air_time" features, particularly air_time23, air_time2, and air_time6, emerged as consistently important variables across multiple models, including naïve Bayes, KNN, AdaBoost, and SVM. Additionally, "paper_time" features (paper_time17 and paper_time23) showed significant importance across several models, indicating the crucial role of temporal features in prediction. The selection of air_time (pen movement duration above paper) and paper_time (pen contact duration) in SMART-Pred's reduced feature model

is rooted in their direct biological ties to AD pathology. Air_time reflects dysfunction of the prefrontal cortex, a region critical for motor planning and task initiation [128, 129]. AD-related neurodegeneration is reflected in handwriting performance, with patients showing longer overall writing times and substantially increased air times between pen strokes across everyday copying tasks, indicating hesitation and disrupted sequencing of movements as they struggle to translate cognitive intentions into fluent motor output [130-132]. Paper_time, conversely, correlates with motor execution deficits caused by degeneration of the basal ganglia and cerebellar region, which impairs fine motor control. Quantitative tablet-based studies have reported that kinematic and pressure features, such as time in air, speed, and complexity, can correctly classify healthy controls, patients with mild cognitive impairment, and patients with AD, with roughly 70%-80% accuracy in 3-group models and above 90% accuracy in several pairwise contrasts. However, it was observed that the accuracy of classification was significantly influenced by the characteristics of the groups being categorized and the particular task, with rates varying from 63.5% to 100%. Moreover, the diagnostic accuracy derived from kinematic measurements exhibited higher specificity in differentiating normal cognition from impaired cognition (mild cognitive impairment and AD), whereas enhanced sensitivity was seen in discriminating between the degrees of impaired cognition (mild cognitive impairment vs AD) [133-135]. More recent digital pen work has shown that handwriting characteristics distinguish patients with AD from controls, with the task accuracy of the XGBoost model achieving a maximum value of 96.55% and the AUC reaching up to 99.10%, proving that handwriting characteristic analysis is promising in auxiliary AD screening or AD diagnosis [136]. On the other hand, the spatiotemporal dynamics of simple loop handwriting separate patients with early-stage AD from healthy older adults, with about 74% classification accuracy [137]. These findings support total duration, time in air, and related temporal variability as sensitive quantitative markers that differentiate healthy aging, mild cognitive impairment, and AD and that can contribute to early staging and screening.

The validity of the variables is further supported by multimodal evidence. Neuroimaging and digital clock drawing studies have linked temporal handwriting features, including "time in air" and "time on paper," to global brain and gray matter volumes, while time-based handwriting metrics have shown significant negative correlations with global cognitive scores, such as the MoCA (Montreal Cognitive Assessment), supporting the interpretation of prolonged air_time and paper_time variability as markers of cognitive decline [138-142]. Compared with invasive biomarkers, such as cerebrospinal fluid Aβ42, which require lumbar puncture and typically achieve diagnostic sensitivities of around 80% at a cost of approximately US $1000 per test [143,144], and FDG-PET (fluorodeoxyglucose positron emission tomography) scans, which show similar median sensitivity but are substantially more expensive (often around US $3000 per scan) [145,146], digital clock drawing metrics, such as air_time and paper_time, provide noninvasive,

low-cost screening with high diagnostic accuracy for early cognitive impairment and AD [147], with our SMART-Pred protocol achieving 88% sensitivity at a cost of approximately US $5 per test (our study). Their ability to capture both executive dysfunction (via planning delays) and visuospatial deficits (via erratic stroke execution) provides a holistic view of AD's cognitive-motor integration breakdown, aligning with National Institutes of Health priorities for functional digital biomarkers [148-150]. This dual focus on mechanistic relevance and clinical practicality makes them ideal for scalable, early AD detection in routine care settings.

Consistent with the literature, our findings showed that random forest had a more balanced feature utilization [151,152], with paper_time6 showing maximum importance (100.00), followed by air_time23 (98.27) and total_time9 (94.82). The model also highlighted the significance of diverse features, such as paper_time23 (90.20), mean_speed_in_air25 (89.03), and mean_jerk_in_air17 (84.03), suggesting a more comprehensive feature integration approach [153-156]. Gradient boosting and XGBoost showed similar patterns but with distinct emphases [156]. Gradient boosting prioritized air_time16 (100.00), while XGBoost highlighted air_time15 (100.00) as the most important feature. The gradient boosting model incorporated additional variables like pressure_mean8 and mean_speed_on_paper2 in the predictions, while the XGBoost model incorporated additional variables like gmrt_in_air1 and total_time3 in the predictions. The neural network model exhibited a unique feature importance pattern, emphasizing variables that were less prominent in other models. While including air_time features, it notably prioritized gmrt_on_paper2 (100.00), mean_acc_in_air13 (98.61), and num_of_pendown19 (98.58), suggesting its ability to capture better patterns in the data, particularly those related to acceleration and movement characteristics. These findings regarding variable importance support the notion that interpretability is crucial in AI systems, especially in medical applications [157-161]. The ability to identify and understand underlying diagnostic biomarkers is not only beneficial for enhancing the diagnostic process but also plays a vital role in tailoring treatment plans for patients [162-168].

ROC curve analysis provided a comparative evaluation of various ML models when assessed with all available variables (full feature model) versus a carefully selected subset of the most important variables ("air_time" and "paper_time" variables; reduced feature model). In the full feature model, AdaBoost demonstrated exceptional predictive capability with an AUC of 0.92, while decision tree showed relatively lower performance with an AUC of 0.73. Strong performance was also noted for random forest and gradient boosting (both with an AUC of 0.89) and XGBoost and neural network (both with an AUC of 0.89). In the reduced feature model, neural network and SVM classifiers outperformed all other algorithms (AUC values of 0.94 and 0.92, respectively), revealing their effectiveness even with fewer inputs. Ensemble models, including random forest, gradient boosting, XGBoost, and AdaBoost, continued to perform well, achieving AUC values ranging from 0.87 to 0.91, which

reflect their robustness [169-172]. Interestingly, the decision tree's AUC improved to 0.81, indicating that it benefited significantly from targeted variable selection in the reduced feature model. Consistent with the literature, our results from the SMART-Pred analysis highlight the consistent strength of ensemble models, such as random forest, gradient boosting, XGBoost, and AdaBoost, across varying variable sets and support the strategic approach of leveraging variable importance ranking to enhance model performance [173-179]. By focusing on critical features, models can maintain or even improve accuracy, as demonstrated in this work. However, it should be emphasized that nothing in this work demonstrated causality between these variables and AD.

Comparison of SMART-Pred With Other AI Tools for AD Diagnostics

SMART-Pred offers a unique approach to AD diagnostics by employing handwriting analysis as a noninvasive biomarker, distinguishing itself from other AI tools in the field. Unlike most existing tools that use cognitive tests, magnetic resonance imaging scans, or blood-based markers, SMART-Pred leverages the DARWIN dataset, which includes handwriting samples from 174 participants (89 patients with AD and 85 healthy controls) across 25 tasks, identifying critical predictors like *air_time* and *paper_time*. This focus provides a novel, cost-effective method for early AD detection. The tool integrates 10 diverse ML models, including random forest, AdaBoost, and neural network, within a user-friendly R Shiny application, democratizing access to advanced AI for nonspecialists and enabling robust performance comparison and customization. With neural network achieving 91% accuracy and SVM achieving 89% accuracy on the test set of the reduced feature model, SMART-Pred surpasses many clinical diagnostic tools (typically having around 81% accuracy) using noninvasive data, and its emphasis on interpretability through variable importance ranking (eg, air_time and paper_time) via the "varImp()" function in R aids physician acceptance and hypothesis formulation.

SMART-Pred stands out due to its reliance on accessible handwriting data compared to resource-intensive methods like magnetic resonance imaging scans or genetic testing used by tools, such as the USC model (approximately 85% accuracy) [180] and the Cambridge AI tool (82% accuracy for progression prediction) [45]. While a quantum ML approach has reported an exceptional 99.89% accuracy, its practical application is limited by computational complexity [181]. On the other hand, SMART-Pred prioritizes ease of use and scalability. Unlike the UCSF model, which predicts AD risk up to 7 years in advance using clinical records (72% accuracy), SMART-Pred focuses on immediate diagnostic classification for direct clinical application. Its ability to bridge complex AI algorithms with user-friendly interfaces addresses a critical gap in clinical adoption, and this is further enhanced by its focus on interpretability, a challenge for many AI models. By pinpointing specific handwriting features as predictors, SMART-Pred not only facilitates early

detection but also contributes to understanding potential risk factors, paving the way for tailored interventions.

The interplay between AI interpretability and physician acceptance is a key factor in integrating AI systems into health care. SMART-Pred, a multifaceted R Shiny AI-based predictive modeling application, demonstrated robust predictive power in AD diagnosis using the DARWIN dataset. Incorporating 10 ML algorithms with tunable parameters, it achieved optimal results, with ensemble models like random forest and AdaBoost excelling across evaluation metrics. This provides an efficient, customizable approach to AI modeling, particularly for complex datasets and disease diagnosis, which is extendable to other classification-based scenarios. Its ability to identify potential risk factors makes it valuable for guiding hypothesis formulation and experimentation across health care domains. Research indicates that physicians are more likely to adopt AI technologies when the reasoning behind insights is clear, especially in complex fields like AD diagnosis and management [182-184].

Future studies could focus on clinical assessments to evaluate whether top variables like *air_time* and *paper_time* are sufficient for AD diagnosis, strengthening the synergy between AI outcomes and experimental findings. Additionally, SMART-Pred could be integrated into electronic health record systems to identify biomarkers of successful aging, enhance disease diagnosis and surveillance, and analyze medical records, addressing public health concerns through data-driven insights and personalized medicine strategies. The use of smart home data for predicting AD symptoms has also shown promise, highlighting the potential of diverse data sources in AD research, with SMART-Pred's handwriting analysis offering a complementary approach.

SMART-Pred demonstrated competitive performance compared to recent AI models for AD prediction. Its unique use of handwriting analysis and accessibility as an R Shiny application contribute significantly to the field. Future work could also explore integrating multiple data modalities to enhance its predictive accuracy and robustness. Additional clinical research is needed to evaluate the practical applicability of the top-ranked variables (*air_time* and *paper_time*) in real-world AD risk assessment and diagnosis.

A comprehensive and comparative analysis of the performance of AI tools for AD detection and prediction is provided in Table 6.

Table 6. Comparative performance of AI[a] tools for AD[b] detection and prediction.

AI tool	Primary data source	Accuracy	Key features/differentiators
SMART-Pred[c] (our study)	Handwriting analysis (DARWIN[d] dataset)	91.00% (neural network) with 95% CI: 0.79-0.97), AUC=94%	Multi-algorithm R Shiny application (10 ML models) with automated feature selection, explainable AI (SHAP), and cross-validation; noninvasive, cost-effective, user-friendly, generalizable framework applicable to diverse biomarkers and deployable beyond handwriting analysis
Cambridge AI tool [45]	Cognitive tests, MRI[e] scans	82.00% (progression prediction)	Predicts AD progression within 3 years; 3 times more accurate than standard clinical markers
USC model [180]	MRI scans, genetic data, biomarkers	85.00% (CNN[f]) with control and MCI[g]/AD groups	Uses deep learning for high accuracy in brain scan analysis; integrates diverse data types
Quantum ML[h] approach [181]	Unspecified (advanced computational methods)	99.89%	Exceptional accuracy but limited by complexity and resource needs; less practical for immediate use
Deep learning CNN [185]	Handwriting analysis (DARWIN dataset)	90.40%	Proposes a novel CNN as a cheap, fast, and accurate solution with a high accuracy, but with no XAI[i]
Particle swarm optimization (PSO) algorithm [186]	Handwriting analysis (DARWIN dataset)	90.57% (LR[j], AdaBoost[k], and RF[l], with RF dominating the accuracy battle with the best feature dimension of 20)	Uses a hybrid ML model for predicting AD using PSO-based feature selection, but with no XAI.
ML [73]	Handwriting analysis (DARWIN dataset)	85.29% (RF)	Proposes a single classifier merging the features from the best tasks, but with no XAI.
ML [187]	DARWIN dataset analysis	82.43% (RF)	Uses several ML techniques with selected tasks of different types and complexities to evaluate the healthy status of patients and a combination of different tasks, but with no XAI
ML [188]	Handwriting analysis (DARWIN dataset)	85% (XGBoost[m])	Uses several ensemble techniques by implementing a 5-fold cross-validation that

AI tool	Primary data source	Accuracy	Key features/differentiators
			ensured robust and unbiased results, but with no XAI
ML [189]	Handwriting analysis (DARWIN dataset)	85.86% (LR)	Uses several ML techniques involving extensive hyperparameter tuning, but with no XAI

[a]AI: artificial intelligence.
[b]AD: Alzheimer disease.
[c]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.
[d]DARWIN: Diagnosis Alzheimer WIth Handwriting.
[e]MRI: magnetic resonance imaging.
[f]CNN: convolutional neural network.
[g]MCI: mild cognitive impairment.
[h]ML: machine learning.
[i]XAI: explainable artificial intelligence.
[j]LR: linear regression.
[k]AdaBoost: adaptive boosting.
[l]RF: random forest.
[m]XGBoost: extreme gradient boosting.

Practical Design Implications of SMART-Pred

The implementation of SMART-Pred, an AI-based tool for early AD detection via handwriting analysis, holds transformative practical design implications across clinical diagnostics, health care technology, and neurodegenerative research. In clinical settings, SMART-Pred provides a noninvasive, cost-effective screening option, achieving up to 91% accuracy with the neural network model (with 95% CI: 0.79-0.97 and AUC=94%), which can be seamlessly integrated into primary care where access to advanced neuroimaging or cerebrospinal fluid testing is limited. Using simple handwriting tasks from the DARWIN dataset, the application enables early identification of motor-cognitive decline, potentially reducing diagnostic delays, especially in underserved populations. This supports value-based care by enhancing efficiency and patient outcomes, aligning with broader trends of AI adoption in health care, and prioritizing patients for further evaluation or early interventions.

Beyond clinical use, SMART-Pred influences health care technology by showcasing a user-friendly R Shiny interface that automates complex ML workflows (data preprocessing, model training, and evaluation) for nonspecialist clinical researchers and providers while addressing challenges like infrastructure compatibility and algorithmic bias. Its focus on explainability through feature importance ranking (*air_time* and *paper_time*) promotes transparency, building trust among clinicians who are wary of black-box systems. In neurodegenerative research, SMART-Pred inspires exploration of noninvasive biomarkers for other disorders like Parkinson disease and Huntington disease, and future validation on datasets like ADNI could lead to hybrid models combining handwriting with imaging or genetic data for enhanced precision. By prioritizing clinical utility, SMART-Pred supports future AI development in radiology and beyond, encouraging interdisciplinary collaboration and clinician involvement in design to address real-world health care needs and drive innovation in diagnostics.

Limitations and Future Work

The SMART-Pred tool has demonstrated promising capabilities for AD prediction through handwriting analysis, but it faces a few critical limitations in its basic version. Computationally, its exclusive reliance on CPU processing within the R environment restricts parallel computing capabilities, leading to inefficiencies when handling large datasets or complex ML algorithms like neural network [190]. This architectural constraint directly impacts scalability, particularly for resource-intensive tasks such as hyperparameter optimization and multimodal data integration. A fundamental reproducibility challenge stems from the inability to save key outputs (model metrics, ROC curves, and trained model states), hindering longitudinal studies and comparative analyses [191-194]. These limitations collectively restrict the tool's utility in real-world clinical workflows.

To address these issues, the authors propose an advanced version integrating Python-based frameworks like TensorFlow for GPU-accelerated neural network models and enhanced computational efficiency. Planned upgrades include flexible hyperparameter optimization, automated parameter distribution analysis, and comprehensive result-saving functionalities. Future validation will expand beyond the DARWIN dataset to incorporate multimodal data from the ADNI, encompassing neuroimaging, genetic, and biochemical biomarkers. This broader validation aims to assess generalizability across diverse populations and clinical settings while enabling cross-modal biomarker comparisons, a critical step toward establishing SMART-Pred as a clinically actionable tool for early AD detection.

Comparison With Prior Work

Our study on SMART-Pred aligns with the growing trend of using AI for the early detection and prediction of AD. The performance of our model (91% accuracy with neural network, 95% CI: 0.79-0.97; AUC=94%) compares favorably with recent advancements in the field. Table 7 presents a concise comparison of studies that used ML, explainable AI, and the DARWIN dataset for AD detection.

Table 7. Comparative analysis of ML[a] and XAI[b] studies using the DARWIN[c] dataset for Alzheimer disease detection.

Dataset	ML techniques used	XAI methods applied	Key insights and features identified	Performance metrics
DARWIN dataset analysis [195]	LR[d], RF[e], SVM[f], and EBM[g]	SHAP[h] and EBM interpretability	"air_time" (in-air movement duration) identified as the most decisive feature group	89% (SD 5%) accuracy with EBM on the reduced dimensionality dataset
DARWIN dataset analysis [196]	ML and stacking	SHAP XAI features importance and direction of influence	"air_time" and "paper_time" (pen contact duration) features dominate the feature importance analysis using SHAP	88.57% accuracy with the stacking ensemble model
DARWIN dataset analysis [197]	LR, AdaBoost[i], XGBoost[j], KNN[k], RF, SVM, etc	LIME[l] and ELI5[m]	"air_time" and "paper_time" features were the most important features using ELI5	91.00% accuracy with RF
SMART-Pred[n] (current study)	LR, AdaBoost, XGBoost, KNN, RF, SVM, neural network, etc	SHAP features importance and direction of influence	"air_time" and "paper_time" as primary biomarkers using SHAP XAI.	91.00% accuracy with neural network; 95% CI 0.79-0.97; AUC=94%

[a]ML: machine learning.
[b]XAI: explainable artificial intelligence.
[c]DARWIN: Diagnosis Alzheimer WIth Handwriting.
[d]LR: logistic regression.
[e]RF: random forest.
[f]SVM: support vector machine.
[g]EBM: explainable boosting machine.
[h]SHAP: Shapley Additive Explanations.
[i]AdaBoost: adaptive boosting.
[j]XGBoost: extreme gradient boosting.
[k]KNN: k-nearest neighbors.
[l]LIME: Local Interpretable Model-Agnostic Explanations.
[m]ELI5: Explain Like I'm 5.
[n]SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling.

The comparative analysis revealed 3 critical insights. First, the studies demonstrate feature consensus, identifying temporal handwriting features like "air_time" as pivotal predictors, thereby validating their biological relevance to AD motor-cognitive decline through their association with prefrontal cortex dysfunction and motor planning deficits. Second, SMART-Pred represents a methodological evolution from prior work by expanding model diversity (10 algorithms vs up to a maximum of 7 algorithms in earlier studies), incorporating neural networks to capture complex nonlinear patterns in handwriting dynamics, and advancing biomarker identification through dual focus on "air_time" (in-air movement duration) and "paper_time" (pen contact duration) rather than single-feature analysis. Third, despite architectural differences, SMART-Pred classifiers performed very well (with neural network achieving an accuracy of 91% [95% CI 0.79-0.97]), highlighting the DARWIN dataset's robustness for handwriting-based diagnosis while indicating that model selection impacts biomarker discovery more than absolute accuracy in this domain. This progression from single-feature interpretation to multi-algorithm, multibiomarker analysis reflects an intentional shift toward clinical translatability, balancing explainability with predictive power.

Conclusion of the Study

The SMART-Pred R Shiny application represents a significant advancement in AI-based predictive modeling for medical diagnostics, offering a user-friendly interface and comprehensive algorithm selection. This tool holds value for researchers and clinicians in personalized medicine and public health surveillance, demonstrating potential to enhance early detection and screening strategies for conditions such as AD and Parkinson disease. By bridging the gap between computational modeling and clinical application, SMART-Pred exemplifies the growing trend of leveraging interactive data visualization and AI-powered analysis in health care settings.

Future development of SMART-Pred will focus on clinical validation and address the challenges associated with imbalanced datasets. The roadmap includes expanding algorithmic support to encompass advanced ML techniques, image recognition architectures, and analysis capabilities for diverse data types such as text, time series, and graph data. Additionally, plans to improve visualization capabilities and user interface design aim to increase accessibility and power. These ongoing enhancements are designed to create a more flexible and comprehensive platform for data analysis across various scientific domains, solidifying SMART-Pred's role in advancing predictive modeling in health care and beyond.

Acknowledgments

Funding

Data Availability

The DARWIN (Diagnosis Alzheimer With Handwriting) dataset [72,73] was used to validate the classification capabilities of the SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) application. A fully reproducible workflow using the provided R codes, including all necessary packages and code snippets for SMART-Pred, is available in our GitHub repository [198]. These scripts are designed to facilitate reproducibility and allow users to replicate the machine learning classification tasks outlined in our study.

Authors' Contributions

Conceptualization: SK
Data curation: HW, EFA, SK
Formal analysis: HW, EFA, SK, SKS
Funding acquisition: SK
Investigation: HW, EFA, SK, SKS
Methodology: HW, EFA, SK
Project administration: SK
Resources: SK
Software: HW, EFA, SK
Supervision: SK, SKS, JGS
Validation: HW, EFA, SK, SKS, JGS
Visualization: HW, EFA, SK
Writing – original draft: SK, HW
Writing – review & editing: SK, EFA, JGS, SKS

Conflicts of Interest

None declared.

Multimedia Appendix 1

SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) application and data.
[ZIP File (ZIP archive File), 357 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Shapley Additive Explanations analysis results for the additional machine learning classifiers.
[DOCX File (Microsoft Word File), 125 KB-Multimedia Appendix 2]

References

1. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. Sep 22, 2023;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
2. Choubey A, Choubey SB, K P, Daulatabad VS, John N. Healthcare transformation: artificial intelligence is the dire imperative of the day. Cureus. Jun 2024;16(6):e62652. [doi: 10.7759/cureus.62652] [Medline: 39036139]
3. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. Jul 2021;8(2):e188-e194. [doi: 10.7861/fhj.2021-0095] [Medline: 34286183]
4. Khalifa M, Albadawy M. AI in diagnostic imaging: revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update. 2024;5:100146. [doi: 10.1016/j.cmpbup.2024.100146]
5. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. Aug 2018;18(8):500-510. [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]

6. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med. Mar 2024;30(3):837-849. [doi: 10.1038/s41591-024-02850-w] [Medline: 38504016]

7. Shafi S, Parwani AV. Artificial intelligence in diagnostic pathology. Diagn Pathol. Oct 3, 2023;18(1):109. [doi: 10.1186/s13000-023-01375-z] [Medline: 37784122]

8. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med. 2019;2(1):28. [doi: 10.1038/s41746-019-0106-0] [Medline: 31304375]

9. Ghenciu LA, Dima M, Stoicescu ER, Iacob R, Boru C, Haţegan OA. Retinal imaging-based oculomics: artificial intelligence as a tool in the diagnosis of cardiovascular and metabolic diseases. Biomedicines. Sep 23, 2024;12(9):2150. [doi: 10.3390/biomedicines12092150] [Medline: 39335664]

10. Kong M, Song SJ. Artificial intelligence applications in diabetic retinopathy: what we have now and what to expect in the future. Endocrinol Metab (Seoul). Jun 2024;39(3):416-424. [doi: 10.3803/EnM.2023.1913] [Medline: 38853435]

11. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, et al. Artificial intelligence in cardiology: present and future. Mayo Clin Proc. May 2020;95(5):1015-1039. [doi: 10.1016/j.mayocp.2020.01.038] [Medline: 32370835]

12. Li Z, Yin S, Wang S, Wang Y, Qiang W, Jiang J. Transformative applications of oculomics-based AI approaches in the management of systemic diseases: a systematic review. J Adv Res. Sep 2025;75:679-696. [doi: 10.1016/j.jare.2024.11.018]

13. Barry T, Farina JM, Chao CJ, et al. The role of artificial intelligence in echocardiography. J Imaging. Feb 20, 2023;9(2):9. [doi: 10.3390/jimaging9020050] [Medline: 36826969]

14. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioengineering (Basel). Mar 29, 2024;11(4):337. [doi: 10.3390/bioengineering11040337] [Medline: 38671759]

15. Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J. Enhancing mental health with artificial intelligence: current trends and future prospects. Journal of Medicine, Surgery, and Public Health. Aug 2024;3:100099. [doi: 10.1016/j.glmedi.2024.100099]

16. Bekbolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. Healthcare (Basel). Jan 5, 2024;12(2):125. [doi: 10.3390/healthcare12020125] [Medline: 38255014]

17. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC state-of-the-art review. J Am Coll Cardiol. Jan 26, 2021;77(3):300-313. [doi: 10.1016/j.jacc.2020.11.030] [Medline: 33478654]

18. Drachman DA. Aging of the brain, entropy, and Alzheimer disease. Neurology (ECronicon). Oct 24, 2006;67(8):1340-1352. [doi: 10.1212/01.wnl.0000240127.89601.83] [Medline: 17060558]

19. Alzheimer's Association. 2018 Alzheimer's disease facts and figures. Alzheimer's & Dementia. Mar 2018;14(3):367-429. [doi: 10.1016/j.jalz.2018.02.001]

20. Breijyeh Z, Karaman R. Comprehensive review on Alzheimer's disease: causes and treatment. Molecules. Dec 8, 2020;25(24):5789. [doi: 10.3390/molecules25245789] [Medline: 33302541]

21. Kumar A, Sidhu J, Lui F, Tsao JW. Alzheimer Disease. StatPearls Publishing; 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK499922/ [Accessed 2025-11-29]

22. Jahn H. Memory loss in Alzheimer's disease. Dialogues Clin Neurosci. Dec 2013;15(4):445-454. [doi: 10.31887/DCNS.2013.15.4/hjahn] [Medline: 24459411]

23. Crecelius C. Diagnosis and treatment of non-Alzheimer's dementias. J Am Med Dir Assoc. 2003;4(4 Suppl):H25-9. [doi: 10.1097/01.JAM.0000078013.50275.FF] [Medline: 12857366]

24. Giau VV, Senanarong V, Bagyinszky E, An SSA, Kim S. Analysis of 50 neurodegenerative genes in clinically diagnosed early-onset Alzheimer's disease. Int J Mol Sci. Mar 26, 2019;20(6):1514. [doi: 10.3390/ijms20061514] [Medline: 30917570]

25. Kennison M, Long E. The long journey of Alzheimer's disease. J Christ Nurs. 2018;35(4):218-227. [doi: 10.1097/CNJ.0000000000000529] [Medline: 30198992]

26. Cummings JL. Current perspectives in Alzheimer's disease. Neurology (ECronicon). Jul 1998;51(1_suppl_1):S2-S17. [doi: 10.1212/WNL.51.1_Suppl_1.S1]

27. Mendez MF. Early-onset Alzheimer disease. Neurol Clin. May 2017;35(2):263-281. [doi: 10.1016/j.ncl.2017.01.005] [Medline: 28410659]

28. DiLoreto R, Murphy CT. The cell biology of aging. Mol Biol Cell. Dec 15, 2015;26(25):4524-4531. [doi: 10.1091/mbc.E14-06-1084] [Medline: 26668170]

29. Bouska M, Huang K, Kang P, Bai H. Organelle aging: lessons from model organisms. J Genet Genomics. Apr 20, 2019;46(4):171-185. [doi: 10.1016/j.jgg.2019.03.011] [Medline: 31080045]

30. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. Sep 11, 2012;22(17):R741-52. [doi: 10.1016/j.cub. 2012.07.024] [Medline: 22975005]

31. Morawe T, Hiebel C, Kern A, Behl C. Protein homeostasis, aging and Alzheimer's disease. Mol Neurobiol. Aug 2012;46(1):41-54. [doi: 10.1007/s12035-012-8246-0] [Medline: 22361852]

32. Cummings JL, Gonzalez MI, Pritchard MC, May PC, Toledo-Sherman LM, Harris GA. The therapeutic landscape of tauopathies: challenges and prospects. Alz Res Therapy. 2023;15(1):168. [doi: 10.1186/s13195-023-01321-7]

33. Pais M, Martinez L, Ribeiro O, et al. Early diagnosis and treatment of Alzheimer's disease: new definitions and challenges. Braz J Psychiatry. Aug 2020;42(4):431-441. [doi: 10.1590/1516-4446-2019-0735] [Medline: 31994640]

34. Nayab DE, Din FU, Ali H, et al. Nano biomaterials based strategies for enhanced brain targeting in the treatment of neurodegenerative diseases: an up-to-date perspective. J Nanobiotechnology. Dec 13, 2023;21(1):477. [doi: 10.1186/ s12951-023-02250-1] [Medline: 38087359]

35. Singh A, Dawson TM, Kulkarni S. Neurodegenerative disorders and gut-brain interactions. J Clin Invest. Jul 1, 2021;131(13):13. [doi: 10.1172/JCI143775] [Medline: 34196307]

36. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. Jan 2019;18(1):88-106. [doi: 10.1016/S1474-4422(18)30403-4] [Medline: 30497964]

37. Maresova P, Hruska J, Klimova B, Barakovic S, Krejcar O. Activities of daily living and associated costs in the most widespread neurodegenerative diseases: a systematic review. Clin Interv Aging. 2020;15:1841-1862. [doi: 10.2147/CIA. S264688] [Medline: 33061334]

38. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimer's & Dementia. Apr 2016;12(4):459-509. [doi: 10.1016/j.jalz.2016.03.001]

39. Rizzi L, Rosset I, Roriz-Cruz M. Global epidemiology of dementia: Alzheimer's and vascular types. Biomed Res Int. 2014;2014:908915. [doi: 10.1155/2014/908915] [Medline: 25089278]

40. Aisen PS, Cummings J, Jack CR Jr, et al. On the path to 2025: understanding the Alzheimer's disease continuum. Alzheimers Res Ther. Aug 9, 2017;9(1):60. [doi: 10.1186/s13195-017-0283-5] [Medline: 28793924]

41. Bowroju SK, Mainali N, Ayyadevara S, et al. Design and synthesis of novel hybrid 8-hydroxy quinoline-indole derivatives as inhibitors of Aβ self-aggregation and metal chelation-induced Aβ aggregation. Molecules. Aug 8, 2020;25(16):16. [doi: 10.3390/molecules25163610] [Medline: 32784464]

42. Kakraba S, Ayyadevara S, Penthala NR, et al. A novel microtubule-binding drug attenuates and reverses protein aggregation in animal models of Alzheimer's disease. Front Mol Neurosci. 2019;12:310. [doi: 10.3389/fnmol.2019. 00310] [Medline: 31920540]

43. Kakraba S, Ayyadevara S, Mainali N, et al. Thiadiazolidinone (TDZD) analogs inhibit aggregation-mediated pathology in diverse neurodegeneration models, and extend *C. elegans* life- and healthspan. Pharmaceuticals (Basel). Oct 20, 2023;16(10):10. [doi: 10.3390/ph16101498] [Medline: 37895969]

44. Ayyadevara S, Balasubramaniam M, Kakraba S, Alla R, Mehta JL, Shmookler Reis RJ. Aspirin-mediated acetylation protects against multiple neurodegenerative pathologies by impeding protein aggregation. Antioxid Redox Signal. Dec 10, 2017;27(17):1383-1396. [doi: 10.1089/ars.2016.6978] [Medline: 28537433]

45. Lee LY, Vaghari D, Burkhart MC, et al. Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings. EClinicalMedicine. Aug 2024;74:102725. [doi: 10.1016/j.eclinm.2024.102725] [Medline: 39764178]

46. Schöll M, Verberk IMW, Del Campo M, et al. Challenges in the practical implementation of blood biomarkers for Alzheimer's disease. Lancet Healthy Longev. Oct 2024;5(10):100630. [doi: 10.1016/j.lanhl.2024.07.013] [Medline: 39369727]

47. Agostinho D, Simões M, Castelo-Branco M. Predicting conversion from mild cognitive impairment to Alzheimer's disease: a multimodal approach. Brain Commun. 2024;6(4):fcae208. [doi: 10.1093/braincomms/fcae208] [Medline: 38961871]

48. Chudzik A, Śledzianowski A, Przybyszewski AW. Machine learning and digital biomarkers can detect early stages of neurodegenerative diseases. Sensors (Basel). Feb 29, 2024;24(5):1572. [doi: 10.3390/s24051572] [Medline: 38475108]

49. Olawade DB, David-Olawade AC, Wada OZ, Asaolu AJ, Adereni T, Ling J. Artificial intelligence in healthcare delivery: prospects and pitfalls. Journal of Medicine, Surgery, and Public Health. Aug 2024;3:100108. [doi: 10.1016/j. glmedi.2024.100108]

50. Chakraborty C, Bhattacharya M, Pal S, Lee SS. From machine learning to deep learning: advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology. 2024;7:100164. [doi: 10.1016/j. crbiot.2023.100164]

51. Jamarani A, Haddadi S, Sarvizadeh R, Haghi Kashani M, Akbari M, Moradi S. Big data and predictive analytics: a systematic review of applications. Artif Intell Rev. 2024;57(7):176. [doi: 10.1007/s10462-024-10811-5]

52. Soori M, Arezoo B, Dastres R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cognitive Robotics. 2023;3:54-70. [doi: 10.1016/j.cogr.2023.04.001]

53. Kakraba S, Yadem AC, Abraham KE. Unraveling protein secrets: machine learning unveils novel biologically significant associations among amino acids. Preprints. Preprint posted online on May 6, 2025. [doi: 10.20944/preprints202505.0139.v1]

54. Yang Z, Zhou H, Srivastav S, et al. Optimizing Parkinson's disease prediction: a comparative analysis of data aggregation methods using multiple voice recordings via an automated artificial intelligence pipeline. Preprints. Preprint posted online on Dec 5, 2024. [doi: 10.20944/preprints202412.0366.v1]

55. Parkinson-s-disease-prediction-code-and-data-repository-main. ResearchGate. 2025. URL: https://www.researchgate.net/publication/387796710_Parkinson-s-Disease-Prediction-Code-and-Data-Repository-main [Accessed 2025-11-29]

56. Netsey EK, Naandam SM, Asante Jnr J, et al. Structural and functional impacts of SARS-CoV-2 spike protein mutations: insights from predictive modeling and analytics (Preprint). JMIR Bioinform Biotechnol. Dec 8, 2025;6:e73637. [doi: 10.2196/73637] [Medline: 41359941]

57. Kakraba S, Ayyadevara S, Yadem Clement A, Abraham KE, Compadre CM, Shmookler Reis RJ. Machine Learning–Enhanced Quantitative Structure-Activity Relationship Modeling for DNA Polymerase Inhibitor Discovery: Algorithm Development and Validation. JMIR AI. Dec 3, 2025;4:e77890. [doi: 10.2196/77890] [Medline: 41340396]

58. Kakraba S, Agyemang EF, Shmookler Reis RJ. Accelerating Discovery of Leukemia Inhibitors with AI-Driven QSAR Modeling (forthcoming). JMIR AI. Dec 5, 2025;05/12/2025:81552. URL: https://preprints.jmir.org/preprint/81552 [doi: 10.2196/81552] [Medline: 41358925]

59. Mahesh B. Machine learning algorithms - a review. International Journal of Science and Research. Jan 5, 2020;9(1):381-386. [doi: 10.21275/ART20203995]

60. Netsey EK, Kakraba S, Naandam SM, Yadem AC. A mathematical graph-theoretic model of single point mutations associated with sickle cell anemia disease. JBT. 2021;9:1-14. [doi: 10.24297/jbt.v9i.9109]

61. Kakraba S, Knisley D. A graph-theoretic model of single point mutations in the cystic fibrosis transmembrane conductance regulator. JBT. 2016;6(1):780-786. [doi: 10.24297/jbt.v6i1.4013]

62. Kakraba S. A hierarchical graph for nucleotide binding domain 2. East Tennessee State University; 2015. URL: https://dc.etsu.edu/etd/2517/ [Accessed 2025-11-29]

63. Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: a comprehensive review. Healthcare (Basel). Mar 15, 2022;10(3):541. [doi: 10.3390/healthcare10030541] [Medline: 35327018]

64. Shivahare BD, Singh J, Ravi V, et al. Delving into machine learning's influence on disease diagnosis and prediction. The Open Public Health Journal. Apr 26, 2024;17(1). [doi: 10.2174/0118749445297804240401061128]

65. Xu D, Xu Z. Machine learning applications in preventive healthcare: a systematic literature review on predictive analytics of disease comorbidity from multiple perspectives. Artif Intell Med. Oct 2024;156:102950. [doi: 10.1016/j.artmed.2024.102950] [Medline: 39163727]

66. Park DJ, Park MW, Lee H, Kim YJ, Kim Y, Park YH. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep. Apr 7, 2021;11(1):7567. [doi: 10.1038/s41598-021-87171-5] [Medline: 33828178]

67. Dixon D, Sattar H, Moros N, et al. Unveiling the influence of AI predictive analytics on patient outcomes: a comprehensive narrative review. Cureus. May 2024;16(5):e59954. [doi: 10.7759/cureus.59954] [Medline: 38854327]

68. Jia L, Yao W, Jiang Y, et al. Development of interactive biological web applications with R/Shiny. Brief Bioinformatics. Jan 17, 2022;23(1). [doi: 10.1093/bib/bbab415]

69. Giorgi FM, Ceraolo C, Mercatelli D. The R language: an engine for bioinformatics and data science. Life (Basel). Apr 27, 2022;12(5):648. [doi: 10.3390/life12050648] [Medline: 35629316]

70. Perkins MA, Carrier JW. Using R Shiny to develop a dashboard using IPEDS, U.S. Census, and bureau of labor statistics data. PLOS ONE. 2023;18(1):e0278573. [doi: 10.1371/journal.pone.0278573] [Medline: 36630356]

71. Bastian G, Baker GH, Limon A. Bridging the divide between data scientists and clinicians. Intelligence-Based Medicine. 2022;6:100066. [doi: 10.1016/j.ibmed.2022.100066]

72. Singh P. Systematic review of data-centric approaches in artificial intelligence and machine learning. Data Science and Management. Sep 2023;6(3):144-157. [doi: 10.1016/j.dsm.2023.06.001]

73. Cilia ND, De Gregorio G, De Stefano C, Fontanella F, Marcelli A, Parziale A. Diagnosing Alzheimer's disease from on-line handwriting: a novel dataset and performance benchmarking. Eng Appl Artif Intell. May 2022;111:104822. [doi: 10.1016/j.engappai.2022.104822]

74. DARWIN. UCI Machine Learning Repository. URL: https://archive.ics.uci.edu/dataset/732/darwin [Accessed 2025-11-29]

75. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed. Feb 2022;214:106584. [doi: 10.1016/j.cmpb.2021.106584] [Medline: 34942412]

76. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clinical Translational Sci. Nov 2024;17(11):e70056. [doi: 10.1111/cts.70056]

77. Feng J, Liang J, Qiang Z, et al. A hybrid stacked ensemble and Kernel SHAP-based model for intelligent cardiotocography classification and interpretability. BMC Med Inform Decis Mak. Nov 28, 2023;23(1):273. [doi: 10.1186/s12911-023-02378-y] [Medline: 38017460]

78. The R Core Team. R: a language and environment for statistical computing. R Project; 2024. URL: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf [Accessed 2025-11-29]

79. Weihs C, Ligges U, Luebke K, Raabe N. KlaR analyzing German business cycles. In: Baier D, Decker R, Schmidt-Thieme L, editors. Data Analysis and Decision Support Studies in Classification, Data Analysis, and Knowledge Organization. Springer; 2005:335-343. [doi: 10.1007/3-540-28397-8_36]

80. Kuhn M. Building predictive models in R using the caret package. J Stat Soft. 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]

81. Therneau T, Atkinson B, Ripley B. rpart: recursive partitioning and regression trees. R Project; 2023. URL: https://cran.r-project.org/web/packages/rpart/rpart.pdf [Accessed 2025-11-29]

82. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18-22. URL: https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf [Accessed 2025-11-29]

83. Ridgeway G, Edwards D, Kriegler B, et al. gbm: generalized boosted regression models. R Project; 2024. URL: https://cran.r-project.org/web/packages/gbm/gbm.pdf [Accessed 2025-11-29]

84. Culp M, Johnson K, Michailidis G. The R package ada for stochastic boosting. R Project; 2016. URL: https://cran.r-project.org/web/packages/ada/ada.pdf [Accessed 2025-11-29]

85. Chen T, He T. xgboost: extreme gradient boosting. R Project; 2024. URL: https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf [Accessed 2025-11-29]

86. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. Springer; 2002. [doi: 10.1007/978-0-387-21706-2]

87. Uddin MF. Addressing accuracy paradox using enhanched weighted performance metric in machine learning. Presented at: Sixth HCT Information Technology Trends (ITT); Nov 20-21, 2019; Ras Al Khaimah, United Arab Emirates. [doi: 10.1109/ITT48889.2019.9075071]

88. Salihoğlu YS, Uslu Erdemir R, Aydur Püren B, et al. Diagnostic performance of machine learning models based on ^{18}F-FDG PET/CT radiomic features in the classification of solitary pulmonary nodules. Mol Imaging Radionucl Ther. Jun 27, 2022;31(2):82-88. [doi: 10.4274/mirt.galenos.2021.43760] [Medline: 35770958]

89. Ávila-Jiménez JL, Cantón-Habas V, Carrera-González MDP, Rich-Ruiz M, Ventura S. A deep learning model for Alzheimer's disease diagnosis based on patient clinical records. Comput Biol Med. Feb 2024;169:107814. [doi: 10.1016/j.compbiomed.2023.107814] [Medline: 38113682]

90. Sun Y, Zhang J, Zhang Y. Adaboost algorithm combined multiple random forest models (Adaboost-RF) is employed for fluid prediction using well logging data. Phys Fluids (1994). Jan 1, 2024;36(1). [doi: 10.1063/5.0179422]

91. Hossain S, Hasan MK, Faruk MO, Aktar N, Hossain R, Hossain K. Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. BMC Cardiovasc Disord. Apr 18, 2024;24(1):214. [doi: 10.1186/s12872-024-03883-2] [Medline: 38632519]

92. Ghasemieh A, Lloyed A, Bahrami P, Vajar P, Kashef R. A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. Decision Analytics Journal. Jun 2023;7:100242. [doi: 10.1016/j.dajour.2023.100242]

93. Hussain SS, Zaidi SSH. AdaBoost ensemble approach with weak classifiers for gear fault diagnosis and prognosis in DC motors. Appl Sci (Basel). 2024;14(7):3105. [doi: 10.3390/app14073105]

94. Dostmohammadi M, Pedram MZ, Hoseinzadeh S, Garcia DA. A GA-stacking ensemble approach for forecasting energy consumption in a smart household: a comparative study of ensemble methods. J Environ Manage. Jul 2024;364:121264. [doi: 10.1016/j.jenvman.2024.121264] [Medline: 38870783]

95. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci. 2017;9:329. [doi: 10.3389/fnagi.2017.00329] [Medline: 29056906]

96.   Moore PJ, Lyons TJ, Gallacher J, Alzheimer's Disease Neuroimaging Initiative. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. PLoS ONE. 2019;14(2):e0211558. [doi: 10.1371/journal.pone.0211558] [Medline: 30763336]

97.   Tripoliti EE, Fotiadis DI, Argyropoulou M, Manis G. A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. J Biomed Inform. Apr 2010;43(2):307-320. [doi: 10.1016/j.jbi.2009.10.004] [Medline: 19883796]

98.   Song M, Jung H, Lee S, Kim D, Ahn M. Diagnostic classification and biomarker identification of Alzheimer's disease with random forest algorithm. Brain Sci. Apr 2, 2021;11(4):453. [doi: 10.3390/brainsci11040453] [Medline: 33918453]

99.   Metcalf CJE, Tate AT, Graham AL. Demographically framing trade-offs between sensitivity and specificity illuminates selection on immunity. Nat Ecol Evol. Nov 2017;1(11):1766-1772. [doi: 10.1038/s41559-017-0315-3] [Medline: 28963489]

100.   Dimitriadis SI, Liparas D, Tsolaki MN. Random forest feature selection, fusion and ensemble strategy: combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and Alzheimer's disease patients: from the Alzheimer's disease neuroimaging initiative (ADNI) database. J Neurosci Methods. May 15, 2018;302:14-23. [doi: 10.1016/j.jneumeth.2017.12.010] [Medline: 29269320]

101.   Shree HV, Mouleeswaran SK. A comprehensive study of predictive models for mortality risk assessment in medical data. Presented at: 4th International Conference on Data Engineering and Communication Systems (ICDECS); Mar 22-23, 2024; Bangalore, India. [doi: 10.1109/ICDECS59733.2023.10502599]

102.   Borkan D, Dixon L, Sorensen J, Thain N, Vasserman L. Nuanced metrics for measuring unintended bias with real data for text classification. Presented at: WWW '19: Companion Proceedings of The 2019 World Wide Web Conference; May 13-17, 2019; San Francisco, CA, USA. [doi: 10.1145/3308560.3317593]

103.   Dumitrescu E, Hué S, Hurlin C, Tokpavi S. Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects. Eur J Oper Res. Mar 2022;297(3):1178-1192. [doi: 10.1016/j.ejor.2021.06.053]

104.   Zhang YP, Zhang XY, Cheng YT, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Military Med Res. 2023;10(1):22. [doi: 10.1186/s40779-023-00458-8]

105.   Zhang S, Pan M. An AutoML-powered analysis framework for forest fire forecasting: adapting to climate change dynamics. Atmosphere (Basel). 2024;15(12):1481. [doi: 10.3390/atmos15121481]

106.   Sui Q, Ghosh SK. Active learning for stacking and AdaBoost-related models. Stats. 2024;7(1):110-137. [doi: 10.3390/stats7010008]

107.   Ahmed SF, Alam M, Hassan M, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. Nov 2023;56(11):13521-13617. [doi: 10.1007/s10462-023-10466-8] [Medline: 37362885]

108.   Khalil M, AlSayed A, Liu Y, Vanrolleghem PA. An integrated feature selection and hyperparameter optimization algorithm for balanced machine learning models predicting N2O emissions from wastewater treatment plants. Journal of Water Process Engineering. Jun 2024;63:105512. [doi: 10.1016/j.jwpe.2024.105512]

109.   Xu J, Luo Y, Peng K, et al. Supplementary motor area driving changes of structural brain network in blepharospasm. Brain (Bacau). Apr 19, 2023;146(4):1542-1553. [doi: 10.1093/brain/awac341] [Medline: 36130317]

110.   Wang L, Zhao J. Artificial intelligence. In: Strategic Blueprint for Enterprise Analytics Studies in Big Data, Vol 150. Springer; 2024:111-138. [doi: 10.1007/978-3-031-55885-6_6]

111.   Yadav S, Singh A, Singhal R, Yadav JP. Revolutionizing drug discovery: the impact of artificial intelligence on advancements in pharmacology and the pharmaceutical industry. Intelligent Pharmacy. Jun 2024;2(3):367-380. [doi: 10.1016/j.ipha.2024.02.009]

112.   ZainEldin H, Gamel SA, Talaat FM, et al. Silent no more: a comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. Artif Intell Rev. 2024;57(7):188. [doi: 10.1007/s10462-024-10816-0]

113.   Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering (Basel). Dec 18, 2023;10(12):1435. [doi: 10.3390/bioengineering10121435] [Medline: 38136026]

114.   Moore I, Magnante C, Embry E, et al. Doctor AI? A pilot study examining responses of artificial intelligence to common questions asked by geriatric patients. Front Artif Intell. 2024;7:1438012. [doi: 10.3389/frai.2024.1438012] [Medline: 39118788]

115.   Li R, Wang X, Lawler K, Garg S, Bai Q, Alty J. Applications of artificial intelligence to aid early detection of dementia: a scoping review on current capabilities and future directions. J Biomed Inform. Mar 2022;127:104030. [doi: 10.1016/j.jbi.2022.104030] [Medline: 35183766]

116.   Ahuja S, Zaheer S. Advancements in pathology: digital transformation, precision medicine, and beyond. J Pathol Inform. Jan 2025;16:100408. [doi: 10.1016/j.jpi.2024.100408] [Medline: 40094037]

117.   Bohr A, Memarzadeh K. Chapter 2 - the rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare. Academic Press; 2020:25-60. [doi: 10.1016/B978-0-12-818438-7.00002-2]

118. Hirani R, Noruzi K, Khuram H, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. Life (Basel). Apr 26, 2024;14(5):557. [doi: 10.3390/life14050557] [Medline: 38792579]

119. Iqbal J, Cortés Jaimes DC, Makineni P, et al. Reimagining healthcare: unleashing the power of artificial intelligence in medicine. Cureus. Sep 2023;15(9):e44658. [doi: 10.7759/cureus.44658] [Medline: 37799217]

120. Arya SS, Dias SB, Jelinek HF, Hadjileontiadis LJ, Pappa AM. The convergence of traditional and digital biomarkers through AI-assisted biosensing: a new era in translational diagnostics? Biosens Bioelectron. Sep 1, 2023;235:115387. [doi: 10.1016/j.bios.2023.115387] [Medline: 37229842]

121. Singh M, Kumar A, Khanna NN, et al. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. EClinicalMedicine. Jul 2024;73:102660. [doi: 10.1016/j.eclinm.2024.102660] [Medline: 38846068]

122. DeGroat W, Abdelhalim H, Peker E, et al. Multimodal AI/ML for discovering novel biomarkers and predicting disease using multi-omics profiles of patients with cardiovascular diseases. Sci Rep. Nov 3, 2024;14(1):26503. [doi: 10.1038/s41598-024-78553-6] [Medline: 39489837]

123. Kapustina O, Burmakina P, Gubina N, Serov N, Vinogradov V. User-friendly and industry-integrated AI for medicinal chemists and pharmaceuticals. Artificial Intelligence Chemistry. Dec 2024;2(2):100072. [doi: 10.1016/j.aichem.2024.100072]

124. Olawade DB, Wada OZ, Ige AO, Egbewole BI, Olojo A, Oladapo BI. Artificial intelligence in environmental monitoring: advancements, challenges, and future directions. Hygiene and Environmental Health Advances. Dec 2024;12:100114. [doi: 10.1016/j.heha.2024.100114]

125. Yelne S, Chaudhary M, Dod K, Sayyad A, Sharma R. Harnessing the power of AI: a comprehensive review of its impact and challenges in nursing science and healthcare. Cureus. Nov 2023;15(11):e49252. [doi: 10.7759/cureus.49252] [Medline: 38143615]

126. Blanco-González A, Cabezón A, Seco-González A, et al. The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals (Basel). Jun 18, 2023;16(6):891. [doi: 10.3390/ph16060891] [Medline: 37375838]

127. Kakraba S, Ayyadevara S, Yadem Clement A, Abraham KE, Compadre CM, Reis RJS. Advanced QSAR modeling with machine learning for drug discovery: targeting DNA polymerase inhibitors (Preprint). JMIR AI. Preprint posted online on Aug 16, 2025. [doi: 10.2196/preprints.77890]

128. Ho NTN, Gonzalez P, Gogovi GK. Writing the signs: an explainable machine learning approach for Alzheimer's disease classification from handwriting. Healthc Technol Lett. 2025;12(1):e70006. [doi: 10.1049/htl2.70006] [Medline: 39949642]

129. Onofri E, Mercuri M, Donato G, Ricci S. Cognitive fluctuations in connection to dysgraphia: a comparison of Alzheimer's disease with dementia Lewy bodies. Clin Interv Aging. 2015;10:625-633. [doi: 10.2147/CIA.S79679] [Medline: 25848239]

130. Fernandes CP, Montalvo G, Caligiuri M, Pertsinakis M, Guimarães J. Handwriting changes in Alzheimer's disease: a systematic review. J Alzheimers Dis. 2023;96(1):1-11. [doi: 10.3233/JAD-230438] [Medline: 37718808]

131. Schröter A, Mergl R, Bürger K, Hampel H, Möller HJ, Hegerl U. Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. Dement Geriatr Cogn Disord. 2003;15(3):132-142. [doi: 10.1159/000068484] [Medline: 12584428]

132. Werner P, Rosenblum S, Bar-On G, Heinik J, Korczyn A. Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. J Gerontol B Psychol Sci Soc Sci. Jul 2006;61(4):228-236. [doi: 10.1093/geronb/61.4.p228] [Medline: 16855035]

133. Lanctôt KL, Boada M, Tariot PN, et al. Association between clinical dementia rating and clinical outcomes in Alzheimer's disease. Alzheimers Dement (Amst). 2024;16(1):e12522. [doi: 10.1002/dad2.12522] [Medline: 38239329]

134. Aggarwal NT, Wilson RS, Beck TL, Bienias JL, Bennett DA. Motor dysfunction in mild cognitive impairment and the risk of incident Alzheimer disease. Arch Neurol. Dec 2006;63(12):1763-1769. [doi: 10.1001/archneur.63.12.1763] [Medline: 17172617]

135. Garre-Olmo J, Faúndez-Zanuy M, López-de-Ipiña K, Calvó-Perxas L, Turró-Garriga O. Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. Curr Alzheimer Res. 2017;14(9):960-968. [doi: 10.2174/1567205014666170309120708] [Medline: 28290244]

136. Qi H, Zhang R, Wei Z, et al. A study of auxiliary screening for Alzheimer's disease based on handwriting characteristics. Front Aging Neurosci. 2023;15:1117250. [doi: 10.3389/fnagi.2023.1117250] [Medline: 37009455]

137. Kahindo C, El-Yacoubi MA, Garcia-Salicetti S, Rigaud AS, Cristancho-Lacroix V. Characterizing early-stage Alzheimer through spatiotemporal dynamics of handwriting. IEEE Signal Process Lett. ;25(8):1136-1140. [doi: 10.1109/LSP.2018.2794500]

138. Zheng R, Zhang Y, Yang Z, Han S, Cheng J. Reduced brain gray matter volume in patients with first-episode major depressive disorder: a quantitative meta-analysis. Front Psychiatry. 2021;12:671348. [doi: 10.3389/fpsyt.2021.671348] [Medline: 34276443]

139. Peng X, Srivastava S, Sutton F, Zhang Y, Badran BW, Kautz SA. Compensatory increase in ipsilesional supplementary motor area and premotor connectivity is associated with greater gait impairments: a personalized fMRI analysis in chronic stroke. Front Hum Neurosci. 2024;18:1340374. [doi: 10.3389/fnhum.2024.1340374] [Medline: 38487103]

140. Koppelmans V, Silvester B, Duff K. Neural mechanisms of motor dysfunction in mild cognitive impairment and Alzheimer's disease: a systematic review. J Alzheimers Dis Rep. 2022;6(1):307-344. [doi: 10.3233/ADR-210065] [Medline: 35891638]

141. Yang J, Yuan C, Chai Y, et al. Dynamic handwriting features for cognitive assessment in inflammatory demyelinating diseases: a machine learning study. Appl Sci (Basel). ;15(11):6257. [doi: 10.3390/app15116257]

142. Yuan J, Au R, Karjadi C, et al. Associations between the digital clock drawing test and brain volume: large community-based prospective cohort (Framingham Heart Study). J Med Internet Res. Apr 15, 2022;24(4):e34513. [doi: 10.2196/34513] [Medline: 35436225]

143. Mattsson N, Zetterberg H, Hansson O, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. JAMA. Jul 22, 2009;302(4):385-393. [doi: 10.1001/jama.2009.1064] [Medline: 19622817]

144. Papaliagkas V, Kalinderi K, Vareltzis P, Moraitou D, Papamitsou T, Chatzidimitriou M. CSF biomarkers in the early diagnosis of mild cognitive impairment and Alzheimer's disease. Int J Mol Sci. May 19, 2023;24(10):8976. [doi: 10.3390/ijms24108976] [Medline: 37240322]

145. Morbelli S, Garibotto V, Van De Giessen E, et al. A Cochrane review on brain [18F]FDG PET in dementia: limitations and future perspectives. Eur J Nucl Med Mol Imaging. Sep 2015;42(10):1487-1491. [doi: 10.1007/s00259-015-3098-2] [Medline: 26067090]

146. Wittenberg R, Knapp M, Karagiannidou M, Dickson J, Schott JM. Economic impacts of introducing diagnostics for mild cognitive impairment Alzheimer's disease patients. A&D Transl Res & Clin Interv. Jan 2019;5(1):382-387. [doi: 10.1016/j.trci.2019.06.001]

147. Müller S, Preische O, Heymann P, Elbing U, Laske C. Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. Front Aging Neurosci. 2017;9:101. [doi: 10.3389/fnagi.2017.00101] [Medline: 28443019]

148. Palmqvist S, Zetterberg H, Mattsson N, et al. Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. Neurology (ECronicon). Oct 6, 2015;85(14):1240-1249. [doi: 10.1212/WNL.0000000000001991] [Medline: 26354982]

149. Andreasen N, Minthon L, Davidsson P, et al. Evaluation of CSF-tau and CSF-Abeta42 as diagnostic markers for Alzheimer disease in clinical practice. Arch Neurol. Mar 2001;58(3):373-379. [doi: 10.1001/archneur.58.3.373] [Medline: 11255440]

150. Krishnan K, Rossetti H, Hynan LS, et al. Changes in montreal cognitive assessment scores over time. Assessment. Sep 2017;24(6):772-777. [doi: 10.1177/1073191116654217] [Medline: 27318033]

151. Speiser JL. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. J Biomed Inform. May 2021;117:103763. [doi: 10.1016/j.jbi.2021.103763] [Medline: 33781921]

152. Huang BFF, Boutros PC. The parameter sensitivity of random forests. BMC Bioinformatics. Sep 1, 2016;17(1):331. [doi: 10.1186/s12859-016-1228-x] [Medline: 27586051]

153. Islam KI, Elias E, Carroll KC, Brown C. Exploring random forest machine learning and remote sensing data for streamflow prediction: an alternative approach to a process-based hydrologic modeling in a snowmelt-driven watershed. Remote Sens (Basel). 2023;15(16):3999. [doi: 10.3390/rs15163999]

154. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV. A random forest based biomarker discovery and power analysis framework for diagnostics research. BMC Med Genomics. Nov 23, 2020;13(1):178. [doi: 10.1186/s12920-020-00826-6] [Medline: 33228632]

155. Chen C, Liang J, Sun W, Yang G, Meng X. An automatically recursive feature elimination method based on threshold decision in random forest classification. Geo-spatial Information Science. Jul 4, 2025;28(4):1494-1519. [doi: 10.1080/10095020.2024.2387457]

156. Shao Z, Ahmad MN, Javed A. Comparison of random forest and XGBoost classifiers using integrated optical and SAR features for mapping urban impervious surface. Remote Sens (Basel). 2024;16(4):665. [doi: 10.3390/rs16040665]

157. Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: a systematic literature review. Comput Biol Med. Nov 2023;166:107555. [doi: 10.1016/j.compbiomed.2023.107555] [Medline: 37806061]

158. Sadeghi Z, Alizadehsani R, Cifci MA, et al. A review of explainable artificial intelligence in healthcare. Computers and Electrical Engineering. Aug 2024;118:109370. [doi: 10.1016/j.compeleceng.2024.109370]

159. Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. Discov Artif Intell. 2024;4(1):15. [doi: 10.1007/s44163-024-00114-7]

160. Farah L, Murris JM, Borget I, Guilloux A, Martelli NM, Katsahian SIM. Assessment of performance, interpretability, and explainability in artificial intelligence-based health technologies: what healthcare stakeholders need to know. Mayo Clin Proc Digit Health. Jun 2023;1(2):120-138. [doi: 10.1016/j.mcpdig.2023.02.004] [Medline: 40206724]

161. Abgrall G, Holder AL, Chelly Dagdia Z, Zeitouni K, Monnet X. Should AI models be explainable to clinicians? Crit Care. Sep 12, 2024;28(1):301. [doi: 10.1186/s13054-024-05005-y] [Medline: 39267172]

162. Ahmad A, Imran M, Ahsan H. Biomarkers as biomedical bioindicators: approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases. Pharmaceutics. May 31, 2023;15(6):1630. [doi: 10.3390/pharmaceutics15061630] [Medline: 37376078]

163. Zhou Y, Tao L, Qiu J, et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. Signal Transduct Target Ther. May 20, 2024;9(1):132. [doi: 10.1038/s41392-024-01823-2] [Medline: 38763973]

164. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: exploring key domains and essential functions. Computer Methods and Programs in Biomedicine Update. 2024;5:100148. [doi: 10.1016/j.cmpbup.2024.100148]

165. Lawler M, Keeling P, Kholmanskikh O, et al. Empowering effective biomarker-driven precision oncology: a call to action. Eur J Cancer. Sep 2024;209:114225. [doi: 10.1016/j.ejca.2024.114225] [Medline: 39053288]

166. Passaro A, Al Bakir M, Hamilton EG, et al. Cancer biomarkers: emerging trends and clinical implications for personalized treatment. Cell. Mar 28, 2024;187(7):1617-1635. [doi: 10.1016/j.cell.2024.02.041] [Medline: 38552610]

167. Llitjos JF, Carrol ED, Osuchowski MF, et al. Enhancing sepsis biomarker development: key considerations from public and private perspectives. Crit Care. Jul 13, 2024;28(1):238. [doi: 10.1186/s13054-024-05032-9] [Medline: 39003476]

168. Marques L, Costa B, Pereira M, et al. Advancing precision medicine: a review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare. Pharmaceutics. Feb 27, 2024;16(3):332. [doi: 10.3390/pharmaceutics16030332] [Medline: 38543226]

169. Nadkarni SB, Vijay GS, Kamath RC. Comparative study of random forest and gradient boosting algorithms to predict airfoil self-noise. Presented at: RAiSE-2023 International Conference on Recent Advances in Science and Engineering; Oct 4-5, 2023; Dubai, United Arab Emirates. [doi: 10.3390/engproc2023059024]

170. Wei Y, Ding J, Yang S, Yang X, Wang F. Comparisons of random forest and stochastic gradient treeboost algorithms for mapping soil electrical conductivity with multiple subsets using Landsat OLI and DEM/GIS-based data at a type oasis in Xinjiang, China. European Journal of Remote Sensing. Jan 1, 2021;54(1):158-181. [doi: 10.1080/22797254.2021.1888657]

171. Patri A, Patnaik Y. Random forest and stochastic gradient tree boosting based approach for the prediction of airfoil self-noise. Procedia Comput Sci. 2015;46:109-121. [doi: 10.1016/j.procs.2015.02.001]

172. Sahin EK. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Appl Sci. Jul 2020;2(7):1308. [doi: 10.1007/s42452-020-3060-1]

173. Naderi Yaghouti AR, Zamanian H, Shalbaf A. Machine learning approaches for early detection of non-alcoholic steatohepatitis based on clinical and blood parameters. Sci Rep. Jan 30, 2024;14(1):2442. [doi: 10.1038/s41598-024-51741-0] [Medline: 38287043]

174. Perera LRD, Ganegoda GU. Alzheimer's disease prediction using clinical data approach. Presented at: 8th International Conference on Information Technology Research (ICITR); Dec 7-8, 2023; Colombo, Sri Lanka. [doi: 10.1109/ICITR61062.2023.10382737]

175. Asadi B, Hajj R. Prediction of asphalt binder elastic recovery using tree-based ensemble bagging and boosting models. Constr Build Mater. Jan 2024;410:134154. [doi: 10.1016/j.conbuildmat.2023.134154]

176. Nguyen KL, Trinh HT, Pham TM. Prediction of punching shear strength in flat slabs: ensemble learning models and practical implementation. Neural Comput & Applic. Mar 2024;36(8):4207-4228. [doi: 10.1007/s00521-023-09296-0]

177. Nazat S, Abdallah M. XAI-based feature ensemble for enhanced anomaly detection in autonomous driving systems. ArXiv. Preprint posted online on Oct 20, 2024. URL: https://arxiv.org/abs/2410.15405 [Accessed 2025-11-29]

178. Mahdizadeh Gharakhanlou N, Perez L. From data to harvest: leveraging ensemble machine learning for enhanced crop yield predictions across Canada amidst climate change. Sci Total Environ. Nov 15, 2024;951:175764. [doi: 10.1016/j.scitotenv.2024.175764] [Medline: 39182775]

179. Ascher S, Wang X, Watson I, Sloan W, You S. Interpretable machine learning to model biomass and waste gasification. Bioresour Technol. Nov 2022;364:128062. [doi: 10.1016/j.biortech.2022.128062] [Medline: 36202285]

180. Chattopadhyay T, Ozarkar SS, Buwa K, et al. Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer's disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI. Front Neurosci. 2024;18:1387196. [doi: 10.3389/fnins.2024.1387196] [Medline: 39015378]

181. Jenber Belay A, Walle YM, Haile MB. Deep ensemble learning and quantum machine learning approach for Alzheimer's disease detection. Sci Rep. Jun 20, 2024;14(1):14196. [doi: 10.1038/s41598-024-61452-1] [Medline: 38902368]

182. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. JMIR AI. Oct 30, 2024;3:e53207. [doi: 10.2196/53207] [Medline: 39476365]

183. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. Jan 2021;113:103655. [doi: 10.1016/j.jbi.2020.103655] [Medline: 33309898]

184. Gaube S, Suresh H, Raue M, et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci Rep. Jan 25, 2023;13(1):1383. [doi: 10.1038/s41598-023-28633-w] [Medline: 36697450]

185. Erdogmus P, Kabakus AT. The promise of convolutional neural networks for the early diagnosis of the Alzheimer's disease. Eng Appl Artif Intell. Aug 2023;123:106254. [doi: 10.1016/j.engappai.2023.106254]

186. Subha R, Nayana B, Selvadass M. Hybrid machine learning model using particle swarm optimization for effectual diagnosis of Alzheimer's disease from handwriting. Presented at: 4th International Conference on Circuits, Control, Communication and Computing; Dec 21-23, 2022; Bangalore, India. [doi: 10.1109/I4C57141.2022.10057948]

187. Gattulli V, Impedovo D, Pirlo G, Semeraro G. Handwriting task-selection based on the analysis of patterns in classification results on Alzheimer dataset. Presented at: IEEESDS'23: Data Science Techniques for Datasets on Mental and Neurodegenerative Disorders; Jun 22, 2023; Zürich, Switzerland. URL: https://ceur-ws.org/Vol-3521/paper2.pdf [Accessed 2025-11-29]

188. Önder M, Şentürk Ü, Polat K, Paulraj D. Diagnosis of Alzheimer's disease using boosting classification algorithms. Presented at: 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE); Nov 1-2, 2023; Chennai, India. [doi: 10.1109/RMKMATE59243.2023.10369418]

189. Anacur CA, Yilmaz AG, Aykut M. Handwriting analysis for Alzheimer's diagnosis: feature enrichment and machine learning classification. Presented at: 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP); Sep 21-22, 2024; Malatya, Turkiye. [doi: 10.1109/IDAP64064.2024.10711121]

190. Humphreys D, Kupresanin A, Boyer MD, et al. Advancing fusion with machine learning research needs workshop report. J Fusion Energ. Aug 2020;39(4):123-155. [doi: 10.1007/s10894-020-00258-1]

191. Olorunnimbe K, Viktor H. Deep learning in the stock market-a systematic survey of practice, backtesting, and applications. Artif Intell Rev. 2023;56(3):2057-2109. [doi: 10.1007/s10462-022-10226-0] [Medline: 35791405]

192. Scheidt F vom, Medinová H, Ludwig N, Richter B, Staudt P, Weinhardt C. Data analytics in the electricity sector – a quantitative and qualitative literature review. Energy and AI. Aug 2020;1:100009. [doi: 10.1016/j.egyai.2020.100009]

193. Sedlakova J, Daniore P, Horn Wintsch A, et al. Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review. PLOS Digit Health. Oct 2023;2(10):e0000347. [doi: 10.1371/journal.pdig.0000347] [Medline: 37819910]

194. Kazemi P, Ghisi A, Entezami A. Artificial intelligence-powered computational strategies in selecting and augmenting data for early design of tall buildings with outer diagrids. Buildings. 2024;14(4):1118. [doi: 10.3390/buildings14041118]

195. Moreira A, Ferreira A, Leite N. Prediction of Alzheimer disease on the DARWIN dataset with dimensionality reduction and explainability techniques. Presented at: 1st International Conference on Explainable AI for Neural and Symbolic Methods; Nov 20-22, 2024; Porto, Portugal. [doi: 10.5220/0013017400003886]

196. Singh SK, Chaturvedi A. Leveraging handwriting dynamics, explainable AI and machine learning for Alzheimer prediction. In: Singh JP, Singh MP, Singh AK, Mukhopadhyay S, Mandal JK, Dutta P, editors. Computational Intelligence in Communications and Business Analytics CICBA 2024 Communications in Computer and Information Science, Vol 2366. Springer; 2025:338-351. [doi: 10.1007/978-3-031-81342-9_27]

197. Samrat N, Islam M, Sharmi S, et al. Explainable AI-driven handwriting analysis for early Alzheimer's disease detection using machine learning and ensemble models. Presented at: 2024 27th International Conference on Computer and Information Technology (ICCIT); Dec 20-22, 2024; Cox's Bazar, Bangladesh. [doi: 10.1109/ICCIT64611.2024.11021728]

198. SMART-Pred. GitHub. URL: https://github.com/Agyemang1z/SMART-Pred-Shiny-Multi-Algorithm-R-Tool-for-Predictive-Modeling-Manuscript [Accessed 2025-12-10]

Abbreviations

AD: Alzheimer disease
AdaBoost: adaptive boosting
ADNI: Alzheimer's Disease Neuroimaging Initiative
ADRD: Alzheimer disease and related dementias
AI: artificial intelligence
AUC: area under the curve
DARWIN: Diagnosis Alzheimer With Handwriting
KNN: k-nearest neighbors
PCA: principal component analysis
ROC: receiver operating characteristic
SHAP: Shapley Additive Explanations
SMART-Pred: Shiny Multi-Algorithm R Tool for Predictive Modeling
SVM: support vector machine
XGBoost: extreme gradient boosting