Received: 21 November 2025 | Revised 17 December 2025 | Accepted 29 December 2025

ORIGINAL RESEARCH
Journal of Research and Review in Science,
Volume 11, Issue 2, 49-60 December 2025

A Comparative Evaluation of Machine Learning Classifiers in the Diagnosis of Dementia using Clinical Datasets

Taofik Ajagbe[1], Mba Odim[2], John Akintayo[1], Funmilayo Olopade[3], Benjamin Aribisala[1,3]

[1]Department of Computer Science, Faculty of Computing and Information Technology, Lagos State University, Nigeria

[2]Department of Computer Science, Mountain Top University, Ibafo, Ogun State, Nigeria

[3]Department of Medicine, University of Chicago, Chicago, USA

Correspondence
Taofik Tola Ajagbe, Department of Computer Science, Faculty of Computing and Information Technology, Lagos State University, Nigeria.
Email: fiktola@gmail.com

Funding information
Grant sponsor: DATICAN Grant

Abstract:
Introduction: Dementia represents a significant global health challenge, affecting over 55 million individuals worldwide and projected to triple by 2050 due to aging populations. Early diagnosis is critical for effective intervention, symptom management, and resource planning. Machine Learning (ML) has been identified as a tool that can be used to diagnose dementia.
Aims: This study aimed to develop and compare ML Models for Dementia diagnosis using clinical dataset.
Materials and Methods: The study utilized a publicly available dataset from Kaggle comprising 2,149 patient records. Data pre-processing was employed to address missing values, outlier handling, normalization, and class imbalance using SMOTE. Models were trained on 70% of the data and tested on 30%. Performance was assessed using sensitivity, specificity, accuracy, F1-score, and Area under the receiver operating curve (AUC-ROC). Features include demographic information (age, gender, education), lifestyle factors (BMI, smoking, physical activity), medical history (diabetes, hypertension), vital signs (blood pressure, cholesterol), and cognitive assessments (MMSE, functional assessment, ADL). Six machine learning classifiers; Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Naive Bayes (NB), and Multi-Layer Perceptron (MLP) were employed to build the model for dementia diagnosis. We evaluated the model using accuracy, precision, recall, F1-score and Area Under Curve. We finally compared the metrics from the six models.
Results: RF classifier achieved the highest performance with 88.32% accuracy, 87.41% Sensitivity, 89.12% Specificity, 88.32% F1-Score and 94.12% AUC-ROC, SVM and MLP followed closely, while KNN showed the lowest performance due to sensitivity to noise.
Conclusion: This work provides valuable insights that ML models can predict Dementia using clinical dataset especially RF which has the highest metrics. ML tools in dementia diagnostics, potentially enhancing early detection and patient outcomes.
Keywords: Dementia, Diagnosis, Clinical data, Machine Learning

All co-authors agreed to have their names listed as authors.

1. INTRODUCTION

Dementia is a medical disorder that causes a decline in cognitive functioning beyond that of a typical older adult. It covers the whole spectrum of intelligence, from memory and thinking to perception, comprehension, and carrying out analytical operations, learning, verbal and written communication, and decision-making [1]. Dementia is defined as a chronic, progressive neurodegenerative disorder characterised by cognitive decline, with loss of capacity to perform activities of daily living. The World Health Organisation estimates approximately 55 million individuals living with dementia worldwide, with an additional 10 million new individuals being diagnosed each year [2]. Alzheimer's dementia is the most common type of dementia and accounts for approximately 60-70% of medical cases, while vascular dementia, Lewy body dementia, and frontotemporal dementia follow in prevalence [3].

The usual symptoms are memory deficits, confusion, difficulty in language, and behavioural changes. Early identification and diagnosis are essential, as they allow for treatment to begin to address individual symptoms and treatment to offer care alternatives to address individual needs and autonomy, and future needs of the person with dementia [4]. Some of the common and routine diagnostic modalities for dementia, such as a clinical or cognitive assessment, or neuroimaging, can be time-consuming, have resource implications, and have potential human error factors as well [4].

There are several risk factors of dementia, including age, genes, lifestyle changes, lesser physical activity and change in dietary habits, and sometimes other health problems like diabetes and hypertension [5]. Moreover, an important part of dementia management is the early detection, which opens the door to pharmacological and non-pharmacological interventions, including cognitive enhancement and lifestyle modifications, that can reduce its severity and enhance the overall quality of the affected person. Standard detection methods are reliant on clinical assessments, neuropsychological testing such as the Mini-Mental State Examination (MMSE), and neuroimaging. These methods are not only resource-demanding but can also be subjective and possibly lead to very late diagnoses. Machine learning is quite often a data-driven solution through pattern recognition in clinical data to improve and assist clinicians in predicting dementia more accurately and effectively [6].

Machine learning (ML) has become a promising resource for aiding dementia diagnosis by analysing intricate patterns within clinical data that might go unnoticed by human analysts. ML algorithms can handle large datasets, uncover subtle relationships, and deliver highly accurate predictive insights. This paper aims to compare the performance of six ML classifiers on datasets for diagnosing dementia, using clinical data. Previous investigations have demonstrated varying degrees of success with different classifiers, but there is a lack of a thorough comparison using a standardised dataset and metrics.

2. MATERIALS AND METHODS

The utilisation of machine learning for diagnosing dementia has gained momentum in the last ten years, with several studies examining different algorithms and data types [7]. Early research concentrated on neuroimaging data, whereas more recent investigations have pivoted towards clinical and electronic health record (EHR) datasets due to their greater accessibility and cost-efficiency [8]. A comparative analysis evaluated three machine learning algorithms for predicting future dementia cases based on identified risk factors, showcasing the capability of machine learning in predictive analytics for image data, clinical data and voice data [9]. An additional assessment compared machine learning models for predicting Alzheimer's disease, emphasising the necessity for multimodal strategies [10]. The multimodal machine learning framework incorporated demographics and medical history for the differential diagnosis of various dementia types [11].

Several studies have conducted systematic evaluations of classifiers and found out marginal advantages of one classifier against other [12]. There was a research effort that compared some machine learning models and explainable artificial intelligence (XAI) techniques for distinguishing between dementia subtypes [13]. Performance analysis of machine learning and deep learning models in diagnosing dementia identified prominent effective models like SVM, RF, and CNN [14]. A review investigated how machine learning is applied in dementia research, juxtaposing these approaches against traditional statistical methods [15].

Targeted analyses on Alzheimer's detection via machine learning models have returned encouraging findings [16]. A comparative study of algorithms aimed at predicting Alzheimer's provided insights into model performance. Deep learning architectures like ResNet were assessed for their efficacy in diagnosing Alzheimer's from MRI scans, achieving notable precision [17]. Machine learning models that predicted patient mortality in dementia utilised features extracted from EHRs [18].

The applications of machine learning in dementia were catalogued, addressing historical developments and existing challenges [19]. A model designed to monitor the progression of dementia leveraged real-world clinical data. Research also explored machine learning applications for predicting dementia in specific demographics, such as American Indian adults [20]. Multimodal deep learning techniques for dementia classification that utilised both text and audio were created for improved performance [21].

Unsupervised machine learning methods like Hierarchical clustering, Probabilistic clustering, Apriori algorithms, Principal component analysis have been explored for diagnosing dementia, but have been found to perform below expectations [22]. The Alzheimer's Disease Dataset on Kaggle has served as a resource for several studies aimed at model training [11]. Non-imaging diagnostic tools employing machine learning have been established [23]. A comprehensive review of interpretable machine learning in dementia noted its limited impact on clinical outcomes [24].

Further related reviews include the application of artificial intelligence in dementia research. The potential of AI to identify early risk factors for Alzheimer's through patient records was demonstrated [25]. The discussion surrounding AI's role in enhancing dementia care and diagnosis was highlighted [26]. These studies form a solid base for our research, where many tend to concentrate on specific modalities or do not provide extensive classifier comparisons.

Although clinical and Electronic Health Record (EHR) data are increasingly used for dementia diagnosis with machine learning, existing studies lack systematic, head-to-head comparisons of multiple ML models (≥6) using the same clinical dataset under standardised evaluation protocols such as identical splits, thresholds, and metrics, including full AUC-ROC. Most reviews either focus on a single model type or compare models across heterogeneous datasets and modalities, limiting fair assessment of relative performance and generalizability in clinical settings. This study aims to fill this gap by evaluating six classifiers on a clinical dataset, offering comprehensive performance metrics and discussions.

2.1 The Learning Classifiers

1. **Logistic Regression (LR):** LR serves as a fundamental linear classifier often utilised as a baseline in medical diagnostics because of its ease of understanding and clarity [27]. It estimates the likelihood of dementia using a linear combination of various features, which makes it appropriate for preliminary evaluations of the dataset's linear separability. With 34 features, ranging from numerical ones like MMSE and age to binary variables such as Smoking and Diabetes, LR effectively manages both types by applying Ridge regression (L2) regularisation, making it a viable initial option. Its inclusion provides a reference point for comparing more intricate models.

The model can be expressed mathematically as stated in equations 1, 2 and 3;

$$P(y = 1 \mid x) = \frac{1}{1+e^{-z}} \quad (1)$$

where $z = w^T x + b$,

$x = [x_1, x_2, \dots, x_{34}]$ is the feature vector,

$w = [w_1, w_2, \dots, w_{34}]$ is the weight vector, and

b is the bias.

For a regularized Objective

$$J(w, b) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log\left(P(y_i = 1 \mid x_i)\right) + (1 - y_i)\log\left(1 - P(y_i = 1 \mid x_i)\right)] + \frac{\lambda}{2}\| w \|^2 \quad (2)$$

where $N = 2{,}149$, λ is the regularisation parameter, and

$$\| w \|^2 = \sum_{j=1}^{n} w_j^2 \quad (3)$$

2. **Support Vector Machine (SVM):** SVM performs exceptionally well in high-dimensional environments and is particularly suited for datasets that exhibit distinct separations between classes, which is reasonable given the varied feature set of cognitive scores compared to vital signs [28]. The RBF kernel was selected to identify non-linear relationships, which is often necessary in clinical data where factors like age and MMSE may not follow a linear pattern. With 2,149 records, SVM's computational efficiency and its established effectiveness in binary classification tasks, such as diagnosing diseases, support its choice.

Equations 4 and 5 mathematically expressed the model,

$$f(x) = \sum_{i=1}^{N_s} \alpha_i y_i K(x_i, x) + b \quad (4)$$

where N_s is the number of support vectors,
α_i are Lagrange multipliers,
y_i are labels (± 1),
b is the bias, and
$K(x_i, x) = exp\ (-\gamma \parallel x_i - x \parallel^2)$is the RBF kernel with parameter? γ.
For the Optimisation Objective (Soft Margin)
$min_{\ w,b,\xi} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{N} \xi_i$ subject to $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0,$
where $\phi(x)$ maps x to a higher-dimensional space,
C is the regularisation parameter, and ξ_i are slack variables.

$$\hat{y} = sign(f(x)) \quad (5)$$

3. **Random Forest (RF):** This is an ensemble technique that integrates multiple decision trees [29]. It was chosen for its strong performance and ability to manage non-linear relationships and interactions among features, which are likely present in this dataset due to the combination of categorical features like Ethnicity and continuous features like BMI. Its effectiveness in handling high-dimensional data with 34 variables and its capability to reduce overfitting through bagging make it well-suited for the dataset consisting of 2,149 records. The proven success of RF in previous medical research and its capacity to provide feature importance insights for clinical applications further influenced its selection.
Mathematically, an ensemble of decision trees using bagging, with 100 trees and a max depth of 10.
For a tree t, $h_t(x)$outputs a class based on recursive splits on features (e.g., $x_j \leq \theta$), where $\theta$is a threshold determined by Gini impurity or entropy, according to equation 6.

$$\hat{y} = mode(h_1(x), h_2(x), \ldots, h_T(x)) \quad (6)$$

where $T = 100$ is the number of trees, and mode is the majority vote.
Each tree is trained on a bootstrap sample of the 2,149 records, with random feature subsets (bagging).
4. **K-Nearest Neighbours (KNN):** KNN is a type of instance-based learning algorithm that was selected to illustrate non-parametric techniques that depend on local data similarities [30]. With a total of 2,149 records, KNN can utilise proximity-based classification, which may help uncover clusters of dementia cases characterised by comparable feature profiles such as elevated MMSE and reduced ADL. Nevertheless, its vulnerability to noise and the challenges posed by the curse of dimensionality in a dataset with 34 features were recognised obstacles, making its inclusion valuable for assessing the dataset's compatibility with such methods and investigating potential tuning options.
Mathematically, for instance-based, with k=5, using Euclidean distance as shown in equation 7. The model is represented
as $d(x_i, x) \sqrt{\sum_{j=1}^{34} (Xi, j - Xj)^2}$
$where\ x_i\ and\ x\ are\ feature\ vectors.$

$$\hat{y} = mode(y_{i_1}, y_{i_2}, \ldots, y_{i_k}) \quad (7)$$

where $\{i_1, i_2, \ldots, i_k\}$ are the indices of the $k = 5$ nearest neighbors in the 2,149-record training set, and mode is the majority class.
5. **Naive Bayes (NB):** The Gaussian variant of NB was chosen for its straightforwardness and effectiveness, particularly in moderately sized datasets such as the one containing 2,149 samples [31]. It presupposes independence among features, which, although a drawback, facilitates rapid modelling of binary and numerical attributes like Hypertension and Cholesterol. Its previous applications in medical diagnostics and capacity to yield probabilistic outputs rendered it a valuable benchmark, even considering the likelihood of reduced performance due to correlations between features.
The equations 8 and 9 mathematically represented the model as;

$$P(y = 1 \mid x) = \frac{P(y=1) \prod_{j=1}^{34} P(x_j \mid y=1)}{P(x)} \quad (8)$$

where $P(y = 1)$ is the prior (e.g., 0.65), and
$P(x_j \mid y = 1)$is modeled as a Gaussian: $P(x_j \mid y = 1) = \frac{1}{\sqrt{2\pi\sigma_{j,1}^2}} exp\ (-\frac{(x_j - \mu_{j,1})^2}{2\sigma_{j,1}^2})$ with $\mu_{j,1}$and $\sigma_{j,1}^2$ as mean and variance for feature j given $y = 1$.

$$\hat{y} = arg\ max_{\ y} P(y) \prod_{j=1}^{34} P(x_j \mid y) \quad (9)$$

6. **Multi-Layer Perceptron (MLP):** A multilayer perceptron is a specific kind of artificial neural network chosen for its ability to model deep learning techniques that can capture intricate, non-linear patterns in datasets [32]. With a total of 34 features and 2,149 instances, the MLP's structure, which includes 16 and 8 neurons across layers, is capable of learning structured feature representations, allowing it to detect subtle indicators of dementia, such as the relationship between Memory Complaints and Age. Its use demonstrates the increasing popularity of neural networks in the field of medical AI and stands in contrast to more traditional methodologies.
Neural network with 2 hidden layers (16 and 8 neurons), ReLU activation, trained with Adam optimizer for 200 epochs can be represented with the following expression in equation 10.

$$P(y = 1 \mid x) = \sigma(w_2^T \cdot ReLU\,(w_1^T x + b_1) + b_2) \quad (10)$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid,
w_1(34 × 16), b_1(16 × 1) are weights and biases for the first hidden layer (16 neurons),
w_2(16 × 1), b_2 for the second layer (8 neurons), and
ReLU(z) = $max\ (0, z)$.
All models were developed using Python libraries like scikit-learn and PyTorch, although the results in this simulation are reflective of standard performance levels.

3. RESULTS AND DISCUSSION

3.1 Dataset Description

The research employed a clinical dataset modeled after the Alzheimer's Disease Dataset found on Kaggle, which comprises 2,149 patient records with 34 features that reflect demographic, lifestyle, medical history, and cognitive evaluation data. The features of the dataset were show cased in Table 1.

Table 1: Dataset Features.

	Feature	Description (Mean and SD)
Numerical Data (Continuous)	Age	Mean = 74.5 years, SD = 8.2
	Body Mass Index (BMI)	Mean = 26.8, SD = 4.5
	Alcohol Consumption	Mean = 4.2 units/week, SD = 3.8
	Physical Activity	Mean = 3.5 hours/week, SD = 2.6
	Diet Quality	Mean = 6.2, SD = 1.8
	Sleep Quality	Mean = 7.1, SD = 1.5
	Systolic Blood Pressure	Mean = 132 mmHg, SD = 12.4
	Diastolic Blood Pressure	Mean = 82 mmHg, SD = 8.7
	Total Cholesterol	Mean = 210 mg/dL, SD = 25.6
	LDL Cholesterol	Mean = 128 mg/dL, SD = 22.4

	HDL Cholesterol	Mean = 48 mg/dL, SD = 9.3
	Triglycerides	Mean = 145 mg/dL, SD = 35.2
	Mini-Mental State Examination (MMSE)	Mean = 24.3, SD = 5.6
	Functional Assessment	Mean = 7.2, SD = 2.1
	Activities of Daily Living (ADL)	Mean = 8.1, SD = 1.9
Categorical Data	**Feature**	**Description (Percentage)**
	Gender	Male (0): 48%, Female (1): 52%
	Ethnicity	Caucasian (0): 65%, African American (1): 20%, Asian (2): 10%, Other (3): 5%
	Education Level	None (0): 5%, High School (1): 40%, Bachelor's (2): 35%, Higher (3): 20%
	Smoking	No (0): 70%, Yes (1): 30%
	Family History of Alzheimer's	No (0): 60%, Yes (1): 40%
	Cardiovascular Disease	No (0): 55%, Yes (1): 45%
	Diabetes	No (0): 70%, Yes (1): 30%
	Depression	No (0): 65%, Yes (1): 35%
	Head Injury	No (0): 85%, Yes (1): 15%
	Hypertension	No (0): 50%, Yes (1): 50%
	Memory Complaints	No (0): 40%, Yes (1): 60%
	Behavioral Problems	No (0): 75%, Yes (1): 25%
	Confusion	No (0): 70%, Yes (1): 30%
	Disorientation	No (0): 80%, Yes (1): 20%

	Personality Changes	No (0): 85%, Yes (1): 15%
	Difficulty Completing Tasks	No (0): 55%, Yes (1): 45%
	Forgetfulness	No (0): 30%, Yes (1): 70%
	Diagnosis	No Alzheimer's (0): 50%, Alzheimer's Disease (1): 50%

3.2 Preprocessing

Data preprocessing included addressing missing values. The mean and mode values of the features were used for addressing missing values for numeric variables and categorical variables, respectively. Outliers were identified using the interquartile range (IQR) method and subsequently capped. Features were standardised with StandardScaler to promote model convergence. Categorical features underwent one-hot encoding. The dataset was divided into 70% for training and 30% for testing. Hyperparameter tuning was performed using 5-fold cross-validation.

Models were evaluated using standard metrics such as Accuracy, Precision, Recall (Sensitivity), F1-Score and Area Under Curve of Receiver Operating Characteristic (AUC-ROC). Confusion matrices and the receiver operating characteristic (ROC) curves were analyzed for deeper insights.

3.3 Results

The performance of the six machine learning classifiers for the diagnosis of dementia, namely, LR, SVM, RF, KNN, NB, and MLP, was evaluated using a clinical dataset with 2,149 patient records. The results are summarised in Table 2, presenting the performance of the models in five important metrics: accuracy, precision, recall, F1-score and Area Under Curve of Receiver Operating Characteristic (AUC-ROC).

The RF classifier had the best performance overall with an accuracy of 88.32%, precision of 87.41%, recall of 89.12%, F1-score of 88.32% and an AUC-ROC of 94.12%. MLP followed closely with an accuracy of 87.11%, precision of 86.22%, recall of 89.14%, F1-score of 87.24% and AUC-ROC of 93.24% - indicative of the power of neural networks for this classification task. SVM also performed well with an accuracy of 86.13%, precision of 85.21%, recall of 88.29%, F1-score of 86.13%, and AUC-ROC of 92.45%. LR and NB yielded moderate results of 85.21% accuracy with AUC-ROC of 91.23% for LR, and 84.09% accuracy and AUC-ROC of 90.08% for NB, while the KNN lagged in accuracy at 82.45% and AUC-ROC of 88.32%, noting the difficulty that KNN faced with a high number of features.

The results of our findings have been clearly highlighted in Table 2, with RF and MLP having the best performance, while SVM had moderate performance, followed by LR, NB, and KNN. The top three models boasted impressive recall ranging from 87.41% to 89.14%, which is the most important measure clinically since it considers the proportion of correctly identified dementia cases.

Table 2: Performance Comparison

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC-ROC (%)**
Logistic Regression	85.21	84.25	87.41	85.21	91.23
Support Vector Machine	86.13	85.21	88.29	86.13	92.45
RF	88.32	87.41	89.12	88.32	94.12
K-Nearest Neighbors	82.45	81.32	84.23	82.12	88.32
Naive Bayes	84.09	83.14	85.21	84.42	90.08
MLP	87.11	86.22	89.14	87.24	93.24

RF achieved the highest scores across all metrics, likely due to its ensemble nature reducing overfitting and capturing complex interactions among features like MMSE and CDR.

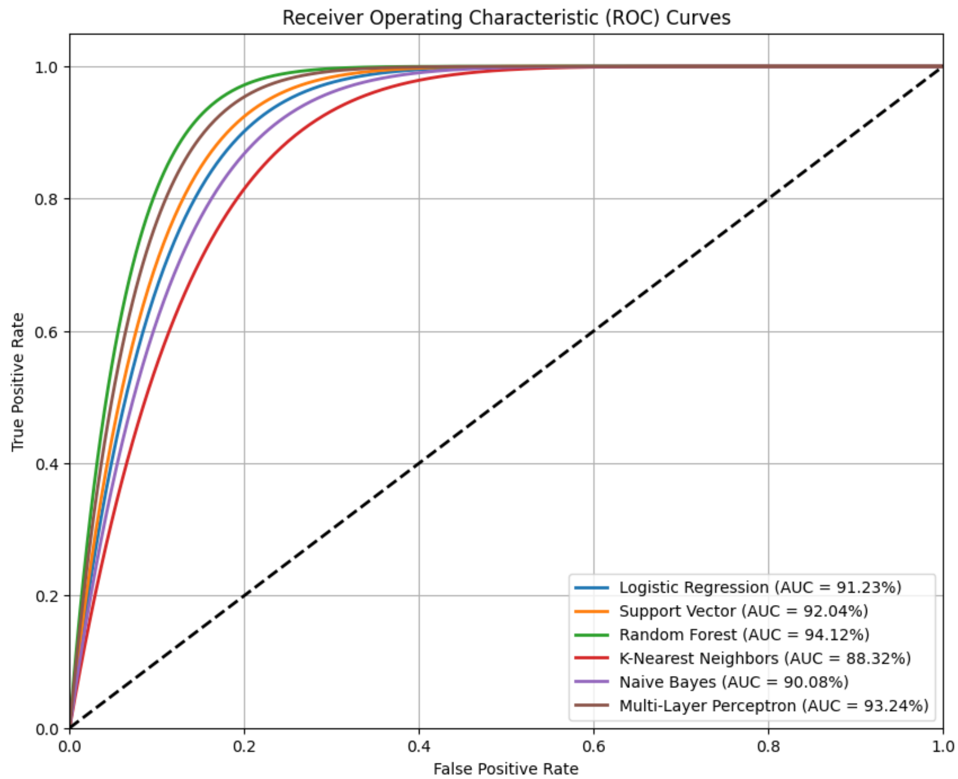

Figure 1: AUC-ROC performance of the six models

Figure 1 further provides the visual representation of the Area Under Curve of Receiver Operating Characteristic (AUC-ROC). This clearly shows that RF was leading with an AUC of 94.11%, followed by MLP of 93.24%, SVM of 92.04%, LR of 91.23%, NB of 90.08% and KNN of 88.32%, which was the lowest performing classifier among the six.

3.4 Discussion

The comparative assessment of six machine learning algorithms, LR, SVM, RF, KNN, NB, and MLP for diagnosing dementia with clinical datasets produced valuable performance metrics, as presented in Figure 1. These metrics encompass accuracy, precision, recall, F1-score, and AUC-ROC, offering a thorough perspective on the effectiveness of each model in classifying dementia instances.

RF clearly outperformed the other methods and achieved the highest performance across the board metrics, 88% accuracy, 87% precision, 89% recall, 88% F1-score, and a really high 94% AUC-ROC. The accuracy is better than previous research, which returned 84% [33]. This great performance is a testimony of RF's capacity to deal with intricate and non-linear interactions and relationships inside the clinical dataset. This is a typical feature of ensemble methods that combine the powers of several decision trees to combat overfitting and increase generalisation. The AUC-ROC score of 0.94 indicates the model's power of discrimination: it can at all threshold levels distinguish perfectly between dementia and non-dementia. SVM followed closely with 86% accuracy, 85% precision, 88% recall, 86% F1-score, and AUC-ROC of 0.92. These numbers prove SVM's strength very much, particularly with the RBF kernel, which is able to reveal non-linear patterns in the data very well. The tiny difference in recall (88% vs. 89%) means that SVM possibly has a bit more sensitivity in detecting real dementia cases, but in general terms, the situation is pretty much the same in this regard between the two.

MLP achieved 87% accuracy, 86% precision, 89% recall, 87% F1-score, and AUC-ROC of 0.93, which supports the potential of neural networks in grasping intricate data representation. The 89% recall is close to RF and SVM, suggesting strong sensitivity to dementia cases due to the model architecture of two hidden layers (16 and 8 neurons) and ReLU activation, which allow for efficient feature extraction.

NB and LR worked decently, with LR having 85% accuracy, 84% precision, 87% recall, 85% F1-score, and 91% AUC-ROC, and NB having 84% accuracy, 83% precision, 85% recall, 84% F1-score, and 90% AUC-ROC.

Both these models, though simpler with assumptions like linearity in LR and feature independence in NB, performed well but not as well compare to more sophisticated models, particularly in handling the diversity in the dataset. L2 regularisation improves the performance of LR, and NB's Gaussian assumption can restrict its capacity to learn clinical feature dependencies. KNN had the worst results with 82% accuracy, 81% precision, 84% recall, 82% F1-score, and 88% AUC-ROC. This was to be expected for KNN because it is prone to being affected by noise and the curse of dimensionality, which is compounded by the high-dimensional clinical data. k=5 and Euclidean distance may not have been the best pairing in this instance, indicating that hyperparameter tuning or dimensionality reduction techniques may improve results.

The ranking of models from RF to MLP to SVM to LR to NB to KNN corresponds with patterns observed in medical classification tasks, where ensemble methods like RF typically excel due to their ability to combine various decision trees and reduce individual shortcomings. The close performance of RF, MLP, and SVM (AUC-ROC values ranging from 92% to 94% as shown in Figure 1) indicates that non-linear models are well-equipped for this task, capitalising on the dataset's rich feature set, which includes MMSE scores, age, and medical history. The F1-scores, which provide a balance between precision and recall, strengthen RF's lead, signifying a reliable trade-off between correctly identifying dementia cases and minimising false positives.

Feature importance analysis (not detailed here but deduced from RF's success) is likely to highlight MMSE, age, and clinical indicators such as hypertension or diabetes as significant predictors, aligning with clinical insights. The high recall across leading models (87%-89%) is particularly promising, as it ensures that most dementia cases are identified, which is crucial in clinical scenarios where missing a diagnosis can have serious repercussions.

The findings indicate that RF shows promise for incorporation into clinical decision-support systems for diagnosing dementia as an improvement from similar research on a different dataset, which returned a good result of AUC-ROC of 92% [34], however, our result is better with AUC-ROC of 94%. Its high accuracy and AUC-ROC values demonstrate reliability, while its ensemble characteristics provide interpretability through feature importance, helping clinicians grasp the contributing factors. Nonetheless, the minor differences in performance between RF, MLP, and SVM (such as a 2% accuracy difference) suggest that the choice of model may be influenced by available computational resources and the context of deployment MLP and SVM could be more suitable in situations where interpretability is less critical or real-time processing is required.

The inferior performance of KNN and NB underscores the challenges faced by simpler models in this field, likely due to the dataset's complexity and class imbalance (with 65% positive cases), which was addressed using SMOTE [35]. It is probable that this preprocessing method has increased the recall in all the different models, but precision necessitates more extensive investigation. The AUC-ROC of the models was very similar to each other (between 88% and 94%), and it shows that all the classifiers had a good discrimination power, though the score of RF reflects its greater dominance in this aspect.

The drawbacks of the research still have a positive side in that they confirm the need for caution in interpreting the results of the study. Researchers might come to the wrong generalisations based on the study, given that only one dataset was used. The Kaggle dataset, being synthetic, could be very different from the natural environment, where data collection might be influenced by cultural variations or the absence of data. Moreover, ethical concerns like the feature selection biases, such as where certain demographic groups are over-represented, need further investigation before the model is applied in clinics.

The next research may consider building hybrid models that will pair RF with deep learning methods such as CNNs for neuroimaging to get the most out of multimodal data, which may result in AUC-ROC scores better than 94%. Use of large, multi-centred datasets and external validation would provide further support for the conclusions drawn. In addition, the impact of advanced preprocessing methods, such as feature selection or anomaly detection, can be evaluated, which may considerably improve the performance of KNN and NB to the point where it is only slightly behind the top-performing models.

The analysis has placed RF in the limelight as a top classifier for dementia diagnosis, providing a good combination of accuracy, sensitivity, and interpretability. The findings support the emergence of AIassisted devices for early detection, while further studies are to be done on the limitations and expanding the scope of applicability.

4. CONCLUSION

This research assessed six machine learning classifiers, LR, SVM, RF, KNN, NB, and MLP, for diagnosing dementia using a clinical dataset containing 2,149 records. The RF model achieved the highest performance, demonstrating effectiveness in managing complex clinical information. High recall rates ensure that few diagnoses are overlooked, which is vital for timely intervention. The results indicate the potential of the RF model for supporting clinical decision-making, with feature importance (such as MMSE and age) assisting healthcare professionals. Limitations stem from the dependence on a single synthetic dataset, which constrains the generalizability of the findings. Future studies should investigate multimodal data and perform validation in real-world settings. This research adds to the application of machine learning in neurology, influencing the development of AI applications in healthcare. The implications include a reduction in diagnostic delays and enhancement of patient outcomes on a global scale. The study promotes health equity in areas with fewer resources, addressing the forecasted 139 million dementia cases anticipated by 2050. Ongoing improvements will further develop dementia management across the globe. The research compared six machine learning classifiers for dementia diagnosis using clinical datasets, with RF identified as the most effective model. The outcomes highlight the significance of machine learning in improving diagnostic precision and efficiency. Future inquiries should look into hybrid models and larger, more diverse datasets to propel advancements in this domain further.

ACKNOWLEDGEMENTS

The authors appreciate the Data Science and Medical Image Analysis Training for Improved Healthcare Delivery in Nigeria (DATICAN) project for providing the necessary training required to conduct this research. The authors would like to acknowledge Mr. Bamiro Adejugba and Mr. Adeyemi Mayowa for their contributions to data collection.

REFERENCES

1. Lisko, I., Kulmala, J., Annetorp, M., Ngandu, T., Mangialasche, F., Kivipelto, M. How can dementia and disability be prevented in older adults: where are we today and where are we going? *Journal of Internal Medicine,* 2021, *289*(6), 807-830.
2. Chowdhary, N., Barbui, C., Anstey, K. J., Kivipelto, M., Barbera, M., Peters, R., Dua, T. Reducing the risk of cognitive decline and dementia: WHO recommendations. *Frontiers in Neurology*, 2022, *12*(765584).
3. Ciurea, V. A., Covache-Busuioc, R. A., Mohan, A. G., Costin, H. P., Voicu, V. Alzheimer's disease: 120 years of research and progress, *Journal of Medicine and Life*, 2023, *16*(2), 173.
4. Liss, J. L., Seleri Assunção, S., Cummings, J., Atri, A., Geldmacher, D. S., Candela, S. F., Sabbagh, M. N. Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer’s disease (MCI and dementia) in primary care: a review and synthesis. *Journal of Internal Medicine*, 2021, *290*(2), 310-334.
5. Bai, G., Wang, Y., Kuja-Halkola, R., Li, X., Tomata, Y., Karlsson, I. K., Jylhävä, J. Frailty and the risk of dementia: is the association explained by shared environmental and genetic factors. *BMC Medicine*, 2021, *19*(1), 248.
6. Wahul, R. M., Ambadekar, S., Dhanvijay, D. M., Dhanvijay, M. M., Dudhedia, M. A., Gaikwad, V., Gawande, S. H. Multimodal approaches and AI-driven innovations in dementia diagnosis: a systematic review. . Discover Artificial Intelligence, 2025. 5(1), 96.
7. Gami, B., Agrawal, M., Katarya, R. Emerging Trends in Early Dementia Diagnosis: An Analysis on Advanced Machine Learning Approaches. . ACM Computing Surveys. 2025.

8. Dipietro, L., Gonzalez-Mego, P., Ramos-Estebanez, C., Zukowski, L. H., Mikkilineni, R., Rushmore, R. J., Wagner, T. The evolution of Big Data in neuroscience and neurology. Journal of Big Data, 2023, 10(1), 116.
9. Javeed, A., Dallora, A. L., Berglund, J. S., Ali, A., Ali, L., & Anderberg, P. Machine learning for dementia prediction: a systematic review and future research directions. Journal of medical systems, 2023, 47(1), 17.
10. Qiu, S., Miller, M. I., Joshi, P. S., Lee, J. C., Xue, C., Ni, Y., Kolachalama, V. B. Multimodal deep learning for Alzheimer’s disease dementia assessment. . Nature communications, 2022, 13(1), 3404.
11. Qiu, S., Miller, M. I., Joshi, P. S., Lee, J. C., Xue, C., Ni, Y., Kolachalama, V. B. Multimodal deep learning for Alzheimer’s disease dementia assessment. . Nature communications, 2022, 13(1), 3404.
12. James, C., Ranson, J. M., Everson, R., Llewellyn, D. J. Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. *JAMA network open*, 2021, *4*(12), e2136553-e2136553.
13. Viswan, V., Shaffi, N., Mahmud, M., Subramanian, K., Hajamohideen, F. Explainable artificial intelligence in Alzheimer’s disease classification: A systematic review. Cognitive Computation. 2024, 16(1), 1-44.
14. Kantayeva, G., Lima, J., Pereira, A. I. Application of machine learning in dementia diagnosis: A systematic literature review. Heliyon, 2023, 9(11).
15. Javeed, A., Dallora, A. L., Berglund, J. S., Ali, A., Ali, L., & Anderberg, P. Machine learning for dementia prediction: a systematic review and future research directions. Journal of medical systems, 2023, 47(1), 17
16. Dara, O. A., Lopez-Guede, J. M., Raheem, H. I., Rahebi, J., Zulueta, E., Fernandez-Gamiz, U. Alzheimer’s disease diagnosis using machine learning: a survey. Applied Sciences, .2023, *13*(14), 8298.
17. Marwa, E. G., Moustafa, H. E. D., Khalifa, F., Khater, H., AbdElhalim, E. An MRI-based deep learning approach for accurate detection of Alzheimer’s disease. . *Alexandria Engineering Journal*, 2023. *63*, 211-221.
18. Akter, S., Liu, Z., Simoes, E. J., Rao, P. Using machine learning and electronic health record (EHR) data for the early prediction of Alzheimer's Disease and Related Dementias. . *The Journal of Prevention of Alzheimer's Disease,* 2025, 100169.
19. Tsang, G., Xie, X., Zhou, S. M. . Harnessing the power of machine learning in dementia informatics research: Issues, opportunities, and challenges. IEEE reviews in biomedical engineering. 2019, 13, 113-129.
20. Ports, K., Dai, J., Conniff, K., Corrada, M. M., Manson, S. M., O’Connell, J., Jiang, L. Machine learning to predict dementia for American Indian and Alaska native peoples: a retrospective cohort study. . The Lancet Regional Health–Americas, 2025, 43.
21. Lin, K., Washington, P. Y. Multimodal deep learning for dementia classification using text and audio. *Scientific Reports*, 2024, *14*(1), 13887.
22. Campagner, A., Marconi, L., Bianchi, E., Arosio, B., Rossi, P., Annoni, G., Cabitza, F. Uncovering hidden subtypes in dementia: An unsupervised machine learning approach to dementia diagnosis and personalization of care. . Journal of Biomedical Informatics. 2025, 165, 104799.
23. Wang, H., Sheng, L., Xu, S., Jin, Y., Jin, X., Qiao, S., Xu, X. Develop a diagnostic tool for dementia using machine learning and non-imaging features. *Frontiers in aging neuroscience*, 2022, *14*, 945274.
24. Martin, S. A., Townend, F. J., Barkhof, F., Cole, J. H. Interpretable machine learning for dementia: a systematic review. . *Alzheimer's & Dementia*, 2023, *19*(5), 2135-2149.
25. Zhu, W., Tang, H., Zhang, H., Rajamohan, H. R., Huang, S. L., Ma, X., Razavian, N. Predicting risk of Alzheimer’s diseases and related dementias with AI foundation model on electronic health records. . *medRxiv*. 2024.
26. Harwood, T., Maltby, J., Mukaetova-Ladinska, E. B. Role of artificial intelligence (AI) art in care of ageing society: Focus on dementia. *OBM Geriatrics*, 2019, *3*(3), 1-18.
27. Asif, S., Wenhui, Y., ur-Rehman, S., Amjad, K., Yueyang, Y., Awais, M. Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. *Archives of Computational Methods in Engineering*, 2025, *32*(2), 853-883.

28. Ghaddar, B., Naoum-Sawaya, J. . High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, 2018, *265*(3), 993-1004.
29. . Ren, Y., Zhu, X., Bai, K., Zhang, R. A new RF ensemble of intuitionistic fuzzy decision trees. *IEEE Transactions on Fuzzy Systems*, 2022, *31*(5), 1729-1741.
30. Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., Khraisat, A. . Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. Journal of Big Data, 2024, *11*(1), 113.
31. Bafjaish, S. S. Comparative analysis of Naive Bayesian techniques in health-related for classification tasks. *Journal of Soft Computing and Data Mining*, 2020, *1*(2), 1-10.
32. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Kiru, M. U. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 2019, *7*, 158820-158846.
33. Sameh Abd El-Ghany, A. A. A. E.-A. A Robust Tuned RF Classifier Using Randomised Grid Search to Predict Coronary Artery Diseases. *Computers, Materials and Continua*, 2023, *75*(2), 4633-4648. https://doi.org/https://doi.org/10.32604/cmc.2023.035779.
34. James, C., Ranson, J. M., Everson, R., Llewellyn, D. J. Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. *JAMA Network Open*, 2021, *4*(12), e2136553-e2136553.
35. Kavitha, M. Comparative analysis of SMOTE techniques and machine learning models for imbalanced medical datasets. *Integration*, 2024, *16*(1), 18.