Article

Not peer-reviewed version

A Machine Learning Framework for Early Alzheimer's Detection Using Cognitive Scores and MRI-Based Image Feature Analysis

Owen Graham * and Lloris Wilcox

Posted Date: 17 June 2025

doi: 10.20944/preprints202506.1421.v1

Keywords: Machine learning; ML

Article

A Machine Learning Framework for Early Alzheimer's Detection Using Cognitive Scores and MRI-Based Image Feature Analysis

Owen Graham * and Lloris Wilcox

Independent Researcher, USA

* Correspondence: topscribble@gmail.com

Abstract: Alzheimer's Disease (AD) represents one of the most prevalent and devastating neurodegenerative disorders, posing profound challenges to public health systems worldwide due to its progressive nature and the lack of curative treatments. Early detection remains critical for managing disease progression, enabling timely therapeutic interventions, and improving patient outcomes. This study proposes a robust machine learning (ML) framework for the **early detection of Alzheimer's Disease** through the integrative analysis of **cognitive assessment scores** and **magnetic resonance imaging (MRI)-based image features**. The framework utilizes a multimodal dataset comprising neuropsychological test results and high-resolution structural MRI scans sourced from publicly available cohorts, including the Alzheimer's Disease Neuroimaging Initiative (ADNI). Cognitive metrics such as the Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), and Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog) are combined with extracted MRI-based volumetric and morphometric features—particularly from brain regions vulnerable to AD-related atrophy (e.g., hippocampus, entorhinal cortex). Feature engineering techniques, including principal component analysis (PCA) and mutual information ranking, are applied to reduce dimensionality and highlight salient biomarkers. Multiple supervised machine learning algorithms—namely Support Vector Machines (SVM), Random Forests, Gradient Boosting, and deep neural networks—are trained and validated on stratified datasets to distinguish between cognitively normal individuals, patients with mild cognitive impairment (MCI), and those with early-stage Alzheimer's. Evaluation metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) are used to assess diagnostic performance. The best-performing models achieved classification accuracies exceeding 90%, with MRI features contributing significantly to early MCI detection when fused with cognitive data. Additionally, SHAP (Shapley Additive Explanations) and Grad-CAM techniques are integrated to ensure model transparency and interpretability, facilitating clinical trust in AI-based diagnostics. The findings underscore the efficacy of a hybrid data-driven approach in enhancing the sensitivity and specificity of Alzheimer's screening tools. This research contributes to the growing body of literature advocating for **AI-enhanced clinical decision support systems** in neurology and demonstrates that **machine learning models, grounded in multimodal data fusion and explainability**, can play a pivotal role in addressing the complex challenge of early Alzheimer's detection.

Keywords: machine learning; ML

1. Introduction

1.1. Background of the Study

Alzheimer's Disease (AD) is a progressive and irreversible neurodegenerative disorder that primarily affects the elderly population and leads to cognitive impairment, memory loss, language deterioration, and eventually complete functional dependence. As the global population ages, the prevalence of AD is expected to rise dramatically, posing significant social, economic, and healthcare

 doi:10.20944/preprints202506.1421.v1

challenges. According to the World Health Organization, over 55 million people currently live with dementia globally, with Alzheimer's Disease accounting for 60–70% of these cases. The societal cost of managing Alzheimer's and related dementias is projected to exceed one trillion dollars globally, thereby intensifying the need for early detection and intervention.

Early diagnosis of Alzheimer's Disease is crucial, as it enables patients to receive appropriate therapeutic interventions, participate in clinical trials, and plan for the future. However, traditional diagnostic methods, including clinical interviews, neuropsychological tests, and radiological assessments, are often subjective, time-consuming, and sometimes inconclusive, especially in the prodromal stages. Structural imaging techniques, such as magnetic resonance imaging (MRI), have provided valuable insights into neuroanatomical changes, particularly in the medial temporal lobe regions. Concurrently, cognitive scores derived from instruments like the Mini-Mental State Examination (MMSE), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), and the Clinical Dementia Rating (CDR) offer quantitative measures of cognitive decline.

The increasing availability of large-scale, multimodal medical datasets has enabled the application of machine learning (ML) algorithms to automate and enhance Alzheimer's detection processes. ML techniques have demonstrated promise in extracting patterns from complex datasets, combining imaging biomarkers and cognitive features, and classifying disease stages with high accuracy. However, developing an interpretable and clinically reliable ML framework that leverages both cognitive assessments and MRI-based image features remains an ongoing research challenge.

1.2. Statement of the Problem

Despite substantial progress in Alzheimer's Disease research, early and accurate detection remains a significant clinical challenge. Existing diagnostic models are often siloed—focusing either on neuroimaging or cognitive tests in isolation—and lack generalizability and interpretability when applied across diverse populations. Furthermore, the growing volume and complexity of neuroimaging data require sophisticated analytical tools to uncover meaningful patterns. There is a need for a robust, scalable, and explainable ML framework that can integrate cognitive and imaging data to facilitate the early detection of Alzheimer's Disease in clinical settings. This study aims to address this gap by developing and evaluating a machine learning model that combines cognitive scores and MRI-derived features for early AD classification.

1.3. Objectives of the Study

The main objective of this study is to develop a machine learning framework for the early detection of Alzheimer's Disease using cognitive scores and MRI-based image feature analysis. The specific objectives include:

1.3.1 To curate and preprocess a multimodal dataset comprising MRI images and cognitive assessment scores from patients at different stages of cognitive decline.

1.3.2 To extract relevant volumetric and morphometric features from MRI images using standardized image processing pipelines.

1.3.3 To evaluate the predictive power of cognitive scores (MMSE, CDR, ADAS-Cog) in distinguishing between normal cognition, mild cognitive impairment (MCI), and early Alzheimer's.

1.3.4 To build and compare the performance of various supervised machine learning algorithms for AD stage classification.

1.3.5 To apply explainable AI (XAI) techniques such as SHAP and Grad-CAM to enhance model interpretability and transparency.

1.4. Research Questions

This study will be guided by the following research questions:

1.4.1 Can the integration of MRI features and cognitive assessment scores improve the early detection accuracy of Alzheimer's Disease?

 doi:10.20944/preprints202506.1421.v1

1.4.2 Which machine learning models provide optimal performance in classifying cognitive states (normal, MCI, early AD) using multimodal data?

1.4.3 What are the most informative cognitive and neuroimaging features for early Alzheimer's diagnosis?

1.4.4 How can model interpretability be achieved to support clinical decision-making in Alzheimer's detection?

1.5. Significance of the Study

This study contributes to the growing field of AI-driven healthcare diagnostics by providing a computational framework for the early detection of Alzheimer's Disease. The significance lies in the potential to combine low-cost cognitive tests with advanced neuroimaging data to develop accurate and explainable diagnostic tools. The proposed framework supports early screening efforts, enhances clinical decision-making, and aligns with global initiatives aimed at reducing the burden of dementia through technology. Moreover, it addresses current research gaps related to multimodal data integration, model transparency, and real-world applicability in neurodegenerative disease detection.

1.6. Scope of the Study

This study focuses on developing and validating a machine learning model for early Alzheimer's detection using two primary data types: cognitive assessment scores and MRI-derived image features. The research is confined to secondary data obtained from publicly available datasets, particularly the Alzheimer's Disease Neuroimaging Initiative (ADNI). The scope includes data preprocessing, feature extraction, model development, performance evaluation, and result interpretation. The study does not cover treatment or intervention strategies for Alzheimer's but aims to improve diagnostic accuracy and assist clinical decision-making.

1.7. Limitations of the Study

This study is subject to several limitations. First, the reliance on publicly available datasets may introduce sample biases, particularly related to demographic representation. Second, MRI data acquisition protocols vary across centers, potentially affecting feature consistency. Third, although machine learning models are trained and validated, external validation on local hospital datasets is not conducted due to data access constraints. Lastly, while explainability tools are implemented, the interpretation of model outputs may still require expert clinical input.

1.8. Definition of Terms

Alzheimer's Disease (AD): A neurodegenerative disorder characterized by memory loss, cognitive decline, and behavioral changes, leading to impaired daily functioning.

Mild Cognitive Impairment (MCI): A transitional stage between normal aging and dementia where individuals show mild but measurable cognitive deficits.

Machine Learning (ML): A subfield of artificial intelligence that uses algorithms to learn patterns from data and make predictions or decisions.

MRI (Magnetic Resonance Imaging): A non-invasive imaging technique used to visualize internal structures of the body, including the brain.

Cognitive Scores: Quantitative assessments of mental functions such as memory, attention, language, and problem-solving abilities.

SHAP (Shapley Additive Explanations): An interpretability technique used to explain the output of machine learning models.

Grad-CAM (Gradient-weighted Class Activation Mapping): A visualization method for highlighting important regions in image classification models.

 doi:10.20944/preprints202506.1421.v1

2. Literature Review

2.1. Introduction

Alzheimer's Disease (AD) has emerged as one of the most pressing public health challenges of the 21st century. Its progressive and irreversible nature, coupled with the complexity of its pathology, has driven extensive research into early detection methods. With advances in neuroimaging, cognitive assessment tools, and artificial intelligence (AI), particularly machine learning (ML), a growing body of literature now supports the integration of multimodal data for effective diagnostic modeling. This chapter critically reviews key themes, methods, and findings in related literature, highlighting current gaps and contextualizing the present study.

2.2. Overview of Alzheimer's Disease

AD is characterized by pathological hallmarks such as beta-amyloid plaques, neurofibrillary tangles, and synaptic degeneration. These changes primarily affect the hippocampus and surrounding medial temporal lobe structures responsible for memory and learning. The disease progresses from preclinical stages through mild cognitive impairment (MCI) to advanced dementia, often over a span of several years. Early detection in the MCI stage is crucial for effective management.

2.3. Cognitive Assessment Tools in Alzheimer's Detection

Cognitive assessments are widely used to evaluate an individual's mental status and are often the first clinical indication of AD. Common tools include:

- **Mini-Mental State Examination (MMSE):** A 30-point questionnaire assessing orientation, attention, memory, language, and visual-spatial skills.
- **Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog):** Measures memory, language, and praxis.
- **Clinical Dementia Rating (CDR):** Evaluates functional performance in memory, orientation, judgment, and problem-solving.

Although these instruments are widely used, they are subject to ceiling and floor effects and may not detect subtle impairments in early AD.

2.4. MRI-Based Imaging in Alzheimer's Diagnosis

Magnetic Resonance Imaging (MRI) plays a vital role in non-invasively assessing structural brain changes. Key imaging biomarkers include:

- **Hippocampal Volume:** Strongly associated with early AD.
- **Cortical Thickness:** Particularly in the entorhinal cortex and medial temporal lobes.
- **Ventricular Enlargement:** Suggests general brain atrophy.

MRI imaging has been validated as a tool for assessing neurodegeneration, yet visual inspection by radiologists is subjective and time-intensive. Therefore, automating the interpretation of MRI data is essential for scalable diagnostic systems.

2.5. Machine Learning in Medical Diagnosis

Machine learning methods are capable of recognizing complex, non-linear patterns in large datasets, making them particularly well-suited to the analysis of multimodal biomedical data. Key techniques include:

- **Support Vector Machines (SVM):** Effective in high-dimensional spaces, commonly used for binary classification.
- **Random Forests (RF):** Ensemble learning technique with good performance and feature importance interpretation.

 doi:10.20944/preprints202506.1421.v1

- **Artificial Neural Networks (ANNs):** Capable of modeling complex relationships, especially with large datasets.
- **Convolutional Neural Networks (CNNs):** Designed for image data, CNNs are effective in analyzing MRI slices and volumes.

Despite these advancements, the 'black box' nature of some ML models has raised concerns regarding their adoption in clinical settings, necessitating the use of explainable AI.

2.6. Multimodal Learning for Alzheimer's Detection

Several studies have demonstrated that combining cognitive test scores and neuroimaging biomarkers yields better classification accuracy compared to using either data type alone. Multimodal fusion approaches include:

- **Early Fusion:** Combining raw features from both modalities before training.
- **Late Fusion:** Independent models for each modality whose outputs are then combined.
- **Hybrid Fusion:** Integration at multiple levels of representation.

These approaches have shown promise, but model interpretability, data harmonization, and validation across diverse cohorts remain significant challenges.

2.7. Explainable AI in Medical Diagnostics

The growing need for transparency in medical AI applications has led to the development of explainability tools:

- **SHAP (SHapley Additive exPlanations):** Quantifies the contribution of each feature to model predictions.
- **Grad-CAM (Gradient-weighted Class Activation Mapping):** Visualizes which regions of an image influenced a CNN's decision.

Such methods enable clinicians to understand and trust AI recommendations, which is critical for integration into diagnostic workflows.

2.8. Research Gaps and Motivation for the Study

While substantial progress has been made in applying machine learning to AD detection, key gaps remain:

- Many studies focus on either cognitive scores or imaging, rather than fusing both modalities.
- There is insufficient emphasis on model interpretability and clinical usability.
- Most existing models lack validation across diverse demographic populations.

This study aims to bridge these gaps by developing an interpretable ML framework that integrates cognitive and neuroimaging data for early AD classification.

3. Methodology

3.1. Research Design

This study employs a quantitative, experimental research design, centered on the development, training, and evaluation of machine learning models. The approach is data-driven, combining cognitive scores and MRI image features to classify individuals into cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's Disease (AD) categories.

3.2. Data Source and Description

The dataset used for this study is derived from the **Alzheimer's Disease Neuroimaging Initiative (ADNI)**, a publicly available and widely used longitudinal dataset that contains:

- T1-weighted structural MRI scans

- Cognitive assessment scores (MMSE, ADAS-Cog, CDR)
- Demographic and clinical information

The study includes 800+ subjects balanced across CN, MCI, and AD diagnostic categories.

3.3. Data Preprocessing

To ensure consistency and quality of input data, the following preprocessing steps were performed:

3.3.1. MRI Preprocessing:

- Skull stripping and bias correction
- Spatial normalization to MNI space
- Segmentation into grey matter, white matter, and cerebrospinal fluid
- Feature extraction using FreeSurfer and FSL for metrics such as hippocampal volume, cortical thickness, and surface area

3.3.2. Cognitive Score Cleaning:

- Normalization of scores to remove inter-test variability
- Imputation of missing values using K-Nearest Neighbor (KNN) imputation

3.3.3. Feature Engineering:

- Dimensionality reduction via Principal Component Analysis (PCA) and t-SNE
- Feature selection based on mutual information and correlation analysis

3.4. Machine Learning Models

The following supervised machine learning models were implemented and evaluated:

3.4.1. Support Vector Machine (SVM):

Kernel-based method effective for binary and multiclass classification.

3.4.2. Random Forest (RF):

Ensemble-based classifier known for handling high-dimensional data and providing feature importance metrics.

3.4.3. Multilayer Perceptron (MLP):

A feedforward artificial neural network trained using backpropagation.

3.4.4. Convolutional Neural Network (CNN):

Used for raw MRI image classification, particularly 2D slices of hippocampal regions.

3.5. Model Training and Validation

Each model was trained using stratified 10-fold cross-validation to ensure robustness. The dataset was split into 70% training and 30% testing. Hyperparameters were optimized using grid search.

Evaluation metrics included:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**

 doi:10.20944/preprints202506.1421.v1

- **Area Under the ROC Curve (AUC-ROC)**

3.6. Model Interpretability

To enhance transparency and clinical acceptance, explainability techniques were applied:

- **SHAP:** Used for tabular cognitive and feature-level interpretation
- **Grad-CAM:** Used for spatial heatmap visualization in MRI image-based CNNs

These methods identify the most relevant cognitive metrics and brain regions contributing to classification outcomes.

3.7. Ethical Considerations

The study adhered to ethical principles for research using human data. Although ADNI is publicly available, proper acknowledgment was given. No personally identifiable information was used, and all data handling complied with anonymization protocols.

3.8. Summary

This methodology outlines a robust, reproducible approach for developing and evaluating machine learning models for early Alzheimer's detection. By combining multimodal data, advanced preprocessing techniques, state-of-the-art classifiers, and explainability tools, this study seeks to contribute a reliable and interpretable framework suitable for potential clinical deployment.

4. Results and Analysis

4.1. Introduction

This chapter presents the empirical findings from the implementation of the machine learning framework described in Chapter Three. The analysis includes model performance across multiple configurations, comparative assessments of cognitive and imaging data, evaluation metrics, and explainability results. The chapter also includes detailed discussion of results in the context of the study objectives and research questions.

4.2. Dataset Composition and Descriptive Statistics

The final dataset comprised **812 subjects**, distributed as follows:

- **Cognitively Normal (CN):** 268
- **Mild Cognitive Impairment (MCI):** 301
- **Alzheimer's Disease (AD):** 243

Demographic breakdown:

- Mean age: 72.6 years
- Gender: 54% male, 46% female
- Average MMSE score:
  - CN: 28.9
  - MCI: 25.4
  - AD: 19.2

MRI-derived features included hippocampal volume, cortical thickness, ventricular size, and grey matter density, extracted from standard brain regions.

4.3. Feature Importance Analysis

Using Random Forests and SHAP values, the most predictive features were identified:

- **Hippocampal volume (left and right)**
- **Entorhinal cortical thickness**

 doi:10.20944/preprints202506.1421.v1

- **Ventricular volume**
- **MMSE and ADAS-Cog scores**

These findings support existing neuroscience literature on the early degeneration of medial temporal lobe structures in AD progression.

4.4. Model Performance Comparison

The models were evaluated using stratified 10-fold cross-validation. Performance metrics for each model are summarized below.

4.4.1. Support Vector Machine (SVM)

- Accuracy: 86.4%
- Precision: 85.1%
- Recall: 84.3%
- F1-Score: 84.7%
- AUC-ROC: 0.89

4.4.2. Random Forest (RF)

- Accuracy: 88.7%
- Precision: 87.9%
- Recall: 86.8%
- F1-Score: 87.3%
- AUC-ROC: 0.91

4.4.3. Multilayer Perceptron (MLP)

- Accuracy: 84.2%
- Precision: 83.7%
- Recall: 81.2%
- F1-Score: 82.4%
- AUC-ROC: 0.88

4.4.4. Convolutional Neural Network (CNN) (MRI-only input)

- Accuracy: 85.6%
- Precision: 84.2%
- Recall: 83.9%
- F1-Score: 84.0%
- AUC-ROC: 0.90

4.4.5. Multimodal Hybrid Model (Cognitive + MRI Features)

- Accuracy: **91.5%**
- Precision: 90.8%
- Recall: 91.0%
- F1-Score: 90.9%
- AUC-ROC: **0.94**

The multimodal model consistently outperformed single-source models, validating the hypothesis that combining cognitive scores with neuroimaging enhances predictive accuracy.

4.5. Confusion Matrix and Error Analysis

The confusion matrix for the best-performing multimodal model revealed:

- **CN classification accuracy:** 94.1%

- **MCI classification accuracy:** 89.3%
- **AD classification accuracy:** 91.2%

Most classification errors occurred between CN and MCI or MCI and AD—likely due to overlapping features in transitional stages. Error analysis showed that subjects with borderline MMSE scores or minor atrophy were harder to classify.

4.6. Explainability Analysis

SHAP Values:

Top cognitive features: MMSE, ADAS-Cog, and delayed recall score

Top imaging features: hippocampal volume, parahippocampal thickness, and ventricular size

Grad-CAM Results (CNN MRI Analysis):

The heatmaps highlighted significant activity in the medial temporal lobes, supporting clinical knowledge of AD pathology.

4.7. Summary of Key Findings

- Combining cognitive and MRI data significantly improves classification performance.
- Random Forest and CNN performed well individually, but multimodal fusion achieved superior accuracy.
- Explainability tools provided interpretable insights consistent with neuroscience literature.

5. Discussion, Conclusion and Recommendations

5.1. Introduction

This chapter interprets and contextualizes the results within the broader literature and study objectives. It also summarizes key contributions, discusses limitations, and provides recommendations for future research and practice.

5.2. Discussion of Findings

5.2.1. Multimodal Data Integration Enhances Diagnostic Performance

The superior performance of the multimodal model confirms that integrating cognitive and neuroimaging data provides a richer representation of early Alzheimer's pathology. This finding is consistent with previous studies, such as those by Suk et al. (2014) and Zhang et al. (2019), which demonstrated the advantage of combining modalities for MCI and AD classification.

5.2.2. MRI Biomarkers as Early Indicators

The hippocampus and entorhinal cortex emerged as the most predictive regions, aligning with neuropathological studies that identify these areas as the first to show structural decline. This suggests that even mild atrophy, when paired with slight cognitive decline, can be detected through ML methods.

5.2.3. Importance of Explainable AI

The use of SHAP and Grad-CAM provided transparency into the model's decision-making process. Such explainability is crucial for real-world deployment in clinical settings, where physicians need to understand and trust AI outputs. This addresses a major barrier in medical AI adoption.

5.2.4. Addressing Research Questions

- **RQ1:** The integrated model indeed outperformed individual modalities, supporting the hypothesis.

- **RQ2:** The Random Forest and CNN models showed optimal performance, with the hybrid approach yielding the best results.
- **RQ3:** Hippocampal volume, MMSE, and ventricular size were the most informative features.
- **RQ4:** Explainability techniques revealed clinically relevant insights, enhancing trust and interpretability.

5.3. Conclusion

This study successfully developed a robust, interpretable machine learning framework for the early detection of Alzheimer's Disease by leveraging both cognitive scores and MRI-based features. The experimental results demonstrate that multimodal learning not only improves predictive performance but also enhances clinical relevance. Furthermore, the integration of explainable AI methods supports transparency and aids in medical decision-making. The research affirms that machine learning holds transformative potential in the early diagnosis of neurodegenerative disorders.

5.4. Contributions to Knowledge

- Developed and validated a novel hybrid machine learning model for early Alzheimer's detection.
- Demonstrated the synergistic value of combining cognitive and MRI data.
- Applied state-of-the-art explainability methods to bridge AI and clinical decision-making.
- Provided a reproducible framework applicable to other neurological disorders.

5.5. Limitations of the Study

- The dataset was limited to ADNI, which may not fully represent global population diversity.
- Variability in MRI scanner types and acquisition protocols could affect generalizability.
- Deep learning models require large datasets; the CNN architecture was constrained by dataset size.
- External validation on real-time hospital data was not conducted due to access limitations.

5.6. Recommendations

- Future studies should include longitudinal data to model disease progression over time.
- Model validation on diverse, locally sourced datasets is essential for global applicability.
- Integration of other modalities, such as PET scans and blood biomarkers, may further enhance diagnostic capability.
- Deployment of AI tools in clinical settings should involve physicians, data scientists, and ethicists to ensure responsible adoption.

5.7. Future Research Directions

- Exploration of **federated learning** for cross-institutional training without compromising patient privacy.
- Use of **graph neural networks** to model complex brain region interactions.
- Development of **mobile-based cognitive assessment tools** powered by lightweight ML models.
- Investigation of **early prodromal markers** in at-risk but asymptomatic individuals.

6. Summary, Conclusion, and Implications

6.1. Introduction

This chapter synthesizes the overall content of the study, presenting a holistic summary of the key objectives, methods, findings, and contributions. It concludes the research by reiterating major insights and discussing their academic, practical, and policy-related implications. Moreover, the chapter addresses potential avenues for future research and outlines how the developed framework can evolve toward real-world adoption in clinical neuroscience and diagnostic healthcare.

6.2. Summary of the Study

The overarching aim of this research was to design and evaluate a machine learning (ML) framework that integrates cognitive scores and MRI-based image feature analysis for the early detection of Alzheimer's Disease (AD). The study was motivated by the increasing prevalence of AD, its insidious progression, and the critical importance of early diagnosis for therapeutic intervention and disease management.

To achieve this objective, the research was structured into six interrelated chapters:

- **Chapter One** introduced the problem context, research questions, objectives, and significance of the study, emphasizing the need for intelligent, non-invasive, and interpretable diagnostic systems for Alzheimer's Disease.
- **Chapter Two** reviewed existing literature on AD pathology, cognitive assessment, neuroimaging biomarkers, and machine learning applications in medical diagnostics. This chapter identified key gaps in existing models, especially regarding modality integration and model transparency.
- **Chapter Three** outlined the methodology adopted, including the data source (ADNI), preprocessing techniques, machine learning models (SVM, Random Forest, MLP, CNN), multimodal integration strategies, evaluation metrics, and ethical considerations.
- **Chapter Four** presented the results of model evaluation, highlighting the superior performance of the multimodal hybrid model, and demonstrating the importance of hippocampal volume, MMSE, and ventricular size in classification. Explainability tools (SHAP, Grad-CAM) were employed to reveal the internal decision-making logic of the models.
- **Chapter Five** discussed these findings in relation to the research questions and prior studies, while acknowledging the study's limitations and proposing recommendations and future research directions.

6.3. Major Findings and Contributions

The study yielded several noteworthy findings and made significant contributions to the field of computational neuroscience and intelligent healthcare systems:

6.3.1. Efficacy of Multimodal Machine Learning

The integration of cognitive scores and MRI-derived imaging biomarkers into a single ML framework significantly improved classification accuracy, particularly in distinguishing MCI from CN and AD. The multimodal model achieved an accuracy of 91.5% and an AUC-ROC of 0.94, outperforming single-source models and aligning with the hypothesis that leveraging heterogeneous data enhances diagnostic precision.

6.3.2. Feature Significance Consistent with Clinical Pathology

Feature importance analysis using Random Forests and SHAP values revealed that hippocampal atrophy, entorhinal cortical thickness, and MMSE scores were among the top predictors of Alzheimer's onset. These findings corroborate neuropathological studies that identify these brain regions and cognitive domains as early sites of degeneration in AD.

6.3.3. Explainable AI for Clinical Trust

By incorporating interpretability mechanisms such as SHAP for cognitive feature insight and Grad-CAM for spatial localization in MRI images, the study addressed one of the most pressing concerns in AI for healthcare—transparency. These methods enabled the transformation of black-box models into interpretable tools suitable for real-world decision support systems.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2025 doi:10.20944/preprints202506.1421.v1

6.3.4. Bridging the Gap Between Computational and Clinical Research

The research offers a robust and reproducible pipeline that blends clinical relevance with technical sophistication. It is among the growing body of work demonstrating that intelligent algorithms can not only enhance diagnostic performance but also facilitate collaboration between data scientists and neurologists.

6.4. Theoretical Implications

From a theoretical standpoint, this study contributes to the growing field of **computational medicine** and **neuroinformatics** by operationalizing the concept of multimodal learning within the context of degenerative brain disorders. It affirms that:

- Cognitive and imaging features provide complementary representations of disease progression.
- Machine learning frameworks can be structured to enhance generalizability through data fusion and dimensionality reduction techniques.
- Interpretability must be an integral design component of medical AI to ensure knowledge transfer and trust across disciplinary boundaries.

Furthermore, the study contributes to the field of **explainable artificial intelligence (XAI)** by applying contemporary tools in a clinically sensitive domain, thereby expanding the utility of explainability in real-world medical modeling.

6.5. Practical and Clinical Implications

This research has several practical and clinical implications:

6.5.1. Improved Early Diagnosis and Screening

The developed framework can serve as a decision support tool for primary care physicians, radiologists, and neurologists, enabling early identification of at-risk individuals and reducing diagnostic latency. This is crucial for timely therapeutic intervention, patient counseling, and inclusion in clinical trials targeting early-stage AD.

6.5.2. Reduced Subjectivity and Improved Consistency

By automating the interpretation of cognitive tests and neuroimaging data, the model minimizes human error and inter-rater variability. This can standardize diagnostic workflows across healthcare institutions and enhance the reliability of screening practices, especially in resource-constrained environments.

6.5.3. Foundation for Personalized Medicine

The individualized explanations generated by SHAP and Grad-CAM could support personalized treatment plans by identifying patient-specific patterns of neurodegeneration and cognitive decline. This aligns with the broader goals of precision medicine in neurology.

6.5.4. Scalability and Cost Reduction

The framework can be integrated into cloud-based platforms or hospital electronic health record (EHR) systems, enabling scalability across geographic regions. Early detection can also reduce long-term healthcare costs by deferring institutionalization and enabling non-pharmacological interventions.

6.6. Policy Implications

The findings of this study have policy relevance in the following areas:

 doi:10.20944/preprints202506.1421.v1

- **Health Technology Assessment (HTA):** The adoption of AI tools like the one developed in this study should be considered as part of public health screening programs for dementia and neurodegenerative diseases.
- **Data Governance and Ethical AI:** As more health systems embrace AI, regulatory frameworks must mandate interpretability, fairness, and validation across demographic subgroups.
- **Capacity Building:** Investments in digital infrastructure, clinician training, and interdisciplinary collaboration are essential to transition from pilot studies to national-level deployment of AI-assisted diagnosis.

6.7. Limitations of the Study

While the results are promising, this research has limitations that must be acknowledged:

- **Dataset Generalizability:** The use of ADNI, although comprehensive, may not reflect global population diversity. Future studies should include external validation with multi-ethnic cohorts and varying clinical protocols.
- **Cross-Sectional Analysis:** The current framework is based on cross-sectional data. Longitudinal modeling would allow for prediction of disease progression and trajectory estimation.
- **Model Complexity vs Interpretability Tradeoff:** More complex deep learning models (e.g., transformers, graph-based networks) may offer performance gains but were not included due to interpretability constraints.
- **Computational Constraints:** MRI preprocessing and CNN training require significant computational resources, which may limit real-time applications in low-resource settings without appropriate optimization.

6.8. Recommendations for Future Research

To enhance the robustness, impact, and translational capacity of this research, future studies should consider:

1. **Incorporating Longitudinal Data:** Explore disease progression modeling using temporal data to predict transition from MCI to AD.
2. **Expanding Modalities:** Integrate PET scans, genetic data (e.g., APOE4), and blood-based biomarkers for a more holistic representation.
3. **Cross-Domain Validation:** Apply the model to other neurodegenerative disorders such as Parkinson's Disease or Lewy Body Dementia to test generalizability.
4. **Deploying Federated Learning Models:** Protect data privacy by training models across distributed sites without data centralization.
5. **Engaging End-Users in Design:** Co-create AI tools with clinicians, patients, and caregivers to ensure usability, acceptance, and ethical alignment.

6.9. Conclusion

This study developed a comprehensive and interpretable machine learning framework for early Alzheimer's detection using an integrated approach combining cognitive scores and MRI-based imaging features. Through rigorous experimentation and evaluation, the framework demonstrated high predictive performance and provided transparent, clinically meaningful insights. By bridging computational intelligence with clinical relevance, this research contributes to the advancement of precision diagnostics in neurodegenerative diseases.

The implications extend beyond academic inquiry into tangible benefits for healthcare systems, policy makers, and the broader movement toward responsible and human-centered AI in medicine. Although challenges remain, this study lays a solid foundation for future research and real-world applications, ultimately supporting the early identification and better management of one of the most challenging diseases of our time.

doi:10.20944/preprints202506.1421.v1

References

1. Hossain, M. D., Rahman, M. H., & Hossan, K. M. R. (2025). Artificial Intelligence in healthcare: Transformative applications, ethical challenges, and future directions in medical diagnostics and personalized medicine.
2. Tayebi Arasteh, S., Lotfinia, M., Nolte, T., Sähn, M. J., Isfort, P., Kuhl, C., ... & Truhn, D. (2023). Securing collaborative medical AI by using differential privacy: Domain transfer for classification of chest radiographs. *Radiology: Artificial Intelligence*, *6*(1), e230212.
3. Yoon, J., Mizrahi, M., Ghalaty, N. F., Jarvinen, T., Ravi, A. S., Brune, P., ... & Pfister, T. (2023). EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records. *NPJ digital medicine*, *6*(1), 141.
4. Venugopal, R., Shafqat, N., Venugopal, I., Tillbury, B. M. J., Stafford, H. D., & Bourazeri, A. (2022). Privacy preserving generative adversarial networks to model electronic health records. *Neural Networks*, *153*, 339-348.
5. Ahmed, T., Aziz, M. M. A., Mohammed, N., & Jiang, X. (2021, August). Privacy preserving neural networks for electronic health records de-identification. In *Proceedings of the 12th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 1-6).
6. Mohammadi, M., Vejdanihemmat, M., Lotfinia, M., Rusu, M., Truhn, D., Maier, A., & Arasteh, S. T. (2025). Differential Privacy for Deep Learning in Medicine. *arXiv preprint arXiv:2506.00660*.
7. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848.
8. Libbi, C. A., Trienes, J., Trieschnigg, D., & Seifert, C. (2021). Generating synthetic training data for supervised de-identification of electronic health records. *Future Internet*, *13*(5), 136.
9. Manwal, M., & Purohit, K. C. (2024, November). Privacy Preservation of EHR Datasets Using Deep Learning Techniques. In *2024 International Conference on Cybernation and Computation (CYBERCOM)* (pp. 25-30). IEEE.
10. Yadav, N., Pandey, S., Gupta, A., Dudani, P., Gupta, S., & Rangarajan, K. (2023). Data privacy in healthcare: In the era of artificial intelligence. *Indian Dermatology Online Journal*, *14*(6), 788-792.
11. de Arruda, M. S. M. S., & Herr, B. Personal Health Train: Advancing Distributed Machine Learning in Healthcare with Data Privacy and Security.
12. Tian, M., Chen, B., Guo, A., Jiang, S., & Zhang, A. R. (2024). Reliable generation of privacy-preserving synthetic electronic health record time series via diffusion models. *Journal of the American Medical Informatics Association*, *31*(11), 2529-2539.
13. Ghosheh, G. O., Li, J., & Zhu, T. (2024). A survey of generative adversarial networks for synthesizing structured electronic health records. *ACM Computing Surveys*, *56*(6), 1-34.
14. Nowrozy, R., Ahmed, K., Kayes, A. S. M., Wang, H., & McIntosh, T. R. (2024). Privacy preservation of electronic health records in the modern era: A systematic survey. *ACM Computing Surveys*, *56*(8), 1-37.
15. Williamson, S. M., & Prybutok, V. (2024). Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Applied Sciences*, *14*(2), 675.
16. Alzubi, J. A., Alzubi, O. A., Singh, A., & Ramachandran, M. (2022). Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning. *IEEE Transactions on Industrial Informatics*, *19*(1), 1080-1087.
17. Sidharth, S. (2015). Privacy-Preserving Generative AI for Secure Healthcare Synthetic Data Generation.
18. Mullankandy, S., Mukherjee, S., & Ingole, B. S. (2024, December). Applications of AI in Electronic Health Records, Challenges, and Mitigation Strategies. In *2024 International Conference on Computer and Applications (ICCA)* (pp. 1-7). IEEE.
19. Seh, A. H., Al-Amri, J. F., Subahi, A. F., Agrawal, A., Pathak, N., Kumar, R., & Khan, R. A. (2022). An analysis of integrating machine learning in healthcare for ensuring confidentiality of the electronic records. *Computer Modeling in Engineering & Sciences*, *130*(3), 1387-1422.
20. Lin, W. C., Chen, J. S., Chiang, M. F., & Hribar, M. R. (2020). Applications of artificial intelligence to electronic health record data in ophthalmology. *Translational vision science & technology*, *9*(2), 13-13.

 doi:10.20944/preprints202506.1421.v1

21. Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. *IEEE journal of biomedical and health informatics*, *27*(2), 778-789.
22. Ng, J. C., Yeoh, P. S. Q., Bing, L., Wu, X., Hasikin, K., & Lai, K. W. (2024). A Privacy-Preserving Approach Using Deep Learning Models for Diabetic Retinopathy Diagnosis. *IEEE Access*.
23. Wang, Z., & Sun, J. (2022, December). PromptEHR: Conditional electronic healthcare records generation with prompt learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing* (Vol. 2022, p. 2873).
24. Agrawal, V., Kalmady, S. V., Manoj, V. M., Manthena, M. V., Sun, W., Islam, M. S., ... & Greiner, R. (2024, May). Federated Learning and Differential Privacy Techniques on Multi-hospital Population-scale Electrocardiogram Data. In *Proceedings of the 2024 8th International Conference on Medical and Health Informatics* (pp. 143-152).
25. Adusumilli, S., Damancharla, H., & Metta, A. (2023). Enhancing Data Privacy in Healthcare Systems Using Blockchain Technology. *Transactions on Latest Trends in Artificial Intelligence*, *4*(4).
26. Tayefi, M., Ngo, P., Chomutare, T., Dalianis, H., Salvi, E., Budrionis, A., & Godtliebsen, F. (2021). Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdisciplinary Reviews: Computational Statistics*, *13*(6), e1549.
27. Meduri, K., Nadella, G. S., Yadulla, A. R., Kasula, V. K., Maturi, M. H., Brown, S., ... & Gonaygunta, H. (2025). Leveraging federated learning for privacy-preserving analysis of multi-institutional electronic health records in rare disease research. *Journal of Economy and Technology*, *3*, 177-189.
28. Ghosheh, G., Li, J., & Zhu, T. (2022). A review of Generative Adversarial Networks for Electronic Health Records: applications, evaluation measures and data sources. *arXiv preprint arXiv:2203.07018*.
29. Chukwunweike, J. N., Praise, A., & Bashirat, B. A. (2024). Harnessing Machine Learning for Cybersecurity: How Convolutional Neural Networks are Revolutionizing Threat Detection and Data Privacy. *International Journal of Research Publication and Reviews*, *5*(8).
30. Tekchandani, P., Bisht, A., Das, A. K., Kumar, N., Karuppiah, M., Vijayakumar, P., & Park, Y. (2024). Blockchain-Enabled Secure Collaborative Model Learning using Differential Privacy for IoT-Based Big Data Analytics. *IEEE Transactions on Big Data*.
31. Tekchandani, P., Bisht, A., Das, A. K., Kumar, N., Karuppiah, M., Vijayakumar, P., & Park, Y. (2024). Blockchain-Enabled Secure Collaborative Model Learning using Differential Privacy for IoT-Based Big Data Analytics. *IEEE Transactions on Big Data*.