International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 481-494
https://doi.org/10.70917/ijcisim-2026-2090
Article

EXPLAINABLE AI IN THE CATEGORIZATION OF ALZHEIMER'S DISEASE

Swapnaja Patwardhan[1], Minakshi More[2], Mukul Kulkarni[3], Darshana Yadav[4]

[1] Department of MCA, MES IMCC, Pune, India. sp.imcc@mespune.in
[2] Department of MCA, MES IMCC, Pune, India. mst.imcc@mespune.in
[3] Department of MCA, MES IMCC, Pune, India. mlk.imcc@mespune.in
[4] Department of MCA, MES IMCC, Pune, India. dpy.imcc@mespune.in
Corresponding Author: Swapnaja Patwardhan (sp.imcc@mespune.in)

Abstract: Recent years have seen an unparalleled expansion in computing power, which has made it possible to create Artificial Intelligence (AI) models for medical applications with impressive outcomes. The usual blackbox nature of many AI models, however, has hindered the acceptability and implementation of many AI-powered Computer Aided Diagnosis (CAD) techniques in the medical field. Thus, in order to encourage medical professionals to use these AI models, the algorithms' predictions need to be comprehensible and interpretable. The goal of the new discipline of explainable AI (XAI) is to demonstrate why the predictions made by these models are reliable. The literature on Alzheimer's disease (AD) detection with XAI that has been published in the past ten years is systematically reviewed in this paper. In order to classify AI models into various conceptual approaches (such as Post-hoc, Ante-hoc, Model-Agnostic, Model-Specific, Global, Local, etc.) and frameworks (such as Layer-wise Relevance Propagation, or LRP, Gradient-weighted Class Activation Mapping, or GradCAM, the Local Interpretable Model-Agnostic Explanation, or LIME, and SHapley Additive exPlanations, or SHAP), research questions were carefully formulated. This classification offers a wide range of interpretations, from intrinsic to global, by extending local explanations. Additionally, many interpretations that offer a thorough understanding of the elements supporting the clinical diagnosis of AD are also covered. Finally, XAI research's needs, limitations, and unresolved issues are described, along with potential applications in AD detection

Keywords Explainable Artificial Intelligence, Blaxk Box Models, Alzimer Disease Classification, Intrepretable Machine Learning

1. INTRODUCTION

Elderly people are susceptible to Alzheimer's Disease (AD), a neurological condition that is incurable and can change a person's life [1]. About 55 million people worldwide have been clinically diagnosed with AD, and by 2050, that number is expected to climb to 139 million, according to the most

Identify applicable funding agency here. If none, delete this

recent World Alzheimer Report [2]. According to the report, a startling 75 of problems remain misdiagnosed for a variety of reasons. AD sufferers will experience a variety of challenges, including cognitive decline, behavioral abnormalities, vision problems, and mobility issues that disrupt everyday activi-ties [3] [4]. A person's capacity to conduct an independent personal and social life will be hindered by these sufferings, which will also cause a great deal of hardship for the family members offering care [5].

In recent times, a variety of application domains have ben-efited from Artificial Intelligence (AI) techniques involving machine learning (ML) and deep learning (DL) algorithms. These include anomaly detection, biosignal and image analy-sis, neurodevelopmental disorder assessment and classification with an emphasis on autism, neurological disorder detection and management, elderly monitoring and care, various dis-ease diagnosis, smart healthcare service delivery, personalized learning, so on [6]- [16].

Using compound medical data, several of these techniques have greatly improved the clinical diagnosis of AD in a very precise, quick, and effective way [16] [17]. This achievement is due to many different kinds of things, including the devel-opment of some algorithms and the availability of powerful GPUs that are already loaded with a variety of open-source computing tools [16]. In order to make the best decisions with the least amount of human involvement, AI-driven AD prediction is predicated on the idea that systems can recognize dementia phases by identifying patterns in the input data. Modern ML and DL algorithms for AD identification have produced incredibly impressive outcomes across a range of criteria [17] [21].

Nonetheless, medical professionals primarily view these AI models as ”blackbox” models since they are unable to provide

valid explanations (explainability) for the predictions they make, which results in uncertainty [22] [23]. Even highly qualified medical professionals frequently find it difficult to understand the answers due to the high level of opacity of these contemporary AI techniques. Because of this, decision-makers will have to choose accuracy over reliability [24]. Because of this, stakeholders and policymakers frequently favor trustworthy and accountable decision-making over pre-cise decision-making. Even though AI-driven computer-aided diagnosis (CAD) has been shown to be accurate in recent research, the medical community is still hesitant to use it because of this lack of explainability [25].

Over the past ten years, a number of machine learning and deep learning algorithms have produced ground-breaking outcomes in a variety of AI-based decisionmaking fields, including machine fault analysis , drug discovery and develop-ment , solid-state material research, and disease prognosis and prediction. Additionally, DL is used in biology, biomedicine, and audio processing, voice recognition, and synthesis [26]-[31]. These performances occasionally outperformed humans in precision.

Uncertain situations like ”Why didn’t you classify that as class Y, as you had predicted?” ”When will you be successful or unsuccessful?” ”How can an incorrect feature selection be fixed?” ”To train the model, which prominent feature are you looking for?” and ”Can I trust the forecast you provided?” are frequently the result of using such blackbox models [22]. Conversely, Explainable AI (XAI) models can provide users with comforting results like ”I know that the wrong feature selection needs to be fixed,” ”I understand why you are putting that in class x and not class y,” ”I can trust your prediction,” and similar statements. Therefore, XAI is essential to the reliable deployment of AI-based CAD.

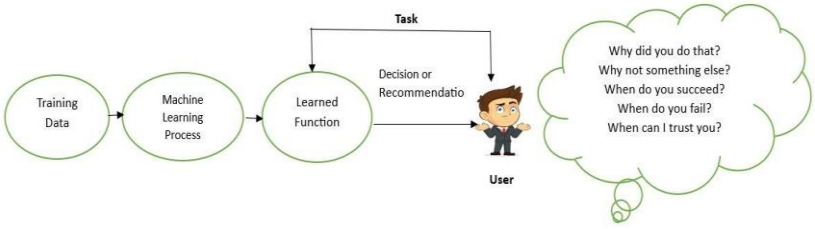

Fig. 1. AI.

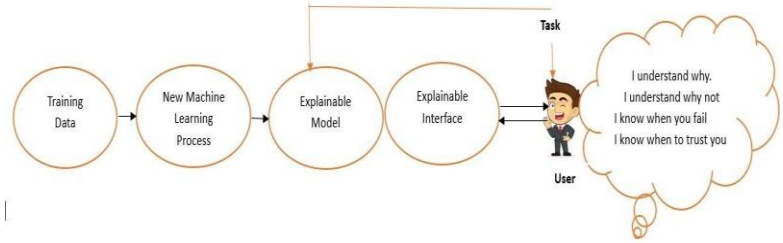

Fig. 2. Explainable AI.

The phrases ”XAI” and ”interpretable AI” can be used inter-changeably to describe a new area of artificial intelligence. A set of capabilities known as XAI can interpret the construction of a blackbox model, make predictions, and win people’s trust so they may effectively use the system [32] - [34]. A

model must be explainable in order to support its output, make the operation of blackbox models clear, acquire new information for more informed decision-making to enhance model performance, and boost user confidence in the model’s output [33]. It seeks to develop techniques and resources that help decision-makers comprehend the choices, advice, or direction made by AI systems. For clinical applications to be accurate, trustworthy, and usable, the model must be interpreted and explained. To fully utilize modern AI solutions, the trust of all stakeholders is required, and XAI is the only way to achieve this. Apart from offering sophisticated insights into AI solutions, XAI has the potential to present novel prospects.AI solutions and human knowledge work together to address difficult issues when neither can produce a suitable solution, and including a human in the decision-making process is a common medical scenario [35]. Fig. 3 provides a broad perspective on converting the blackbox concept into an explainable model. There is a frequent trade-

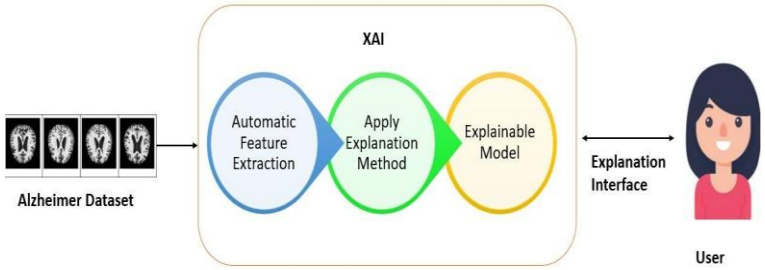

Fig. 3. Explanable AI Outlook.

off among model accuracy and it’s associated explainability. Decision tree (DT) or linear regression models are intuitive, naturally interpretable, and simple for even a non-expert in AI to validate and comprehend [36]. This makes such models more trustworthy. However, machine learning algorithms may provide a non-linear model to handle a complex problem, which might produce good results but compromise explain-ability. Convolutional Neural Networks (CNNs), for example, frequently exhibit the best performance but are the least explicable [37].

XAI has been increasingly significant in the AI world in recent years, not only because it is utilized in high-stakes decisions but also because regulators hold businesses responsible for the choices their AI models make. The explainable aspect of AI has been adopted by a number of various fields, which place a higher value on reliability than accuracy. XAI has been used in recommender systems, neurological illnesses, industrial applications, gaming, and drug development. In recent years, the healthcare industry has seen a number of XAI-based papers as a result of this amazing rise [38] - [43]. This papers aims to creation of fundamental research ques-tions (RQ) that encompass the full range of XAI for the classification of AD, GitHub links to a collection of vari-ous XAI methods for deciphering blackbox models used for AD detection, An examination of XAI techniques for AD classification that has been published in the past 10 years, together with a critical evaluation of the results, discoveries, capabilities, and limits, determining the XAI models’ strengths for AD detection in order to guarantee their dependability and

credibility for clinician adoption and a thorough examination of the advantages, drawbacks, difficulties, and potential future paths of present XAI research. These important discoveries will close a number of research gaps and spur the development of new models that help physicians better understand how an AI system is perceived.

2. CONCEPT AND CONTEXTT

A brief overview of the various XAI techniques in general is given in this section (XAI Methods). Additionally, it offers a succinct synopsis of the most widely used XAI frameworks for AD prediction. This section’s main goal is to give a thorough background that will be useful for discussions in subsequent sections.

XAI Methods

As seen in Fig. 4, the XAI techniques can be roughly divided into four groups [44] according to: i) explanation scope, ii) implementation stages, iii) model applicability, and

iv) explanation formats. The range of explanations produced

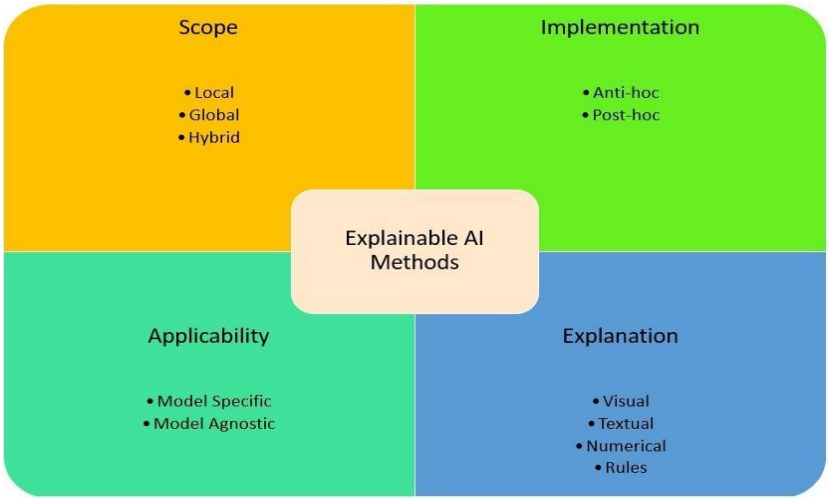

Fig. 4. XAI Methods

by the XAI approach is known as the explainability scope. Depending on the input test data, it can interpret a single instance of the model or the complete model. As a result, a model's explainability can be classified as either local or global. A global approach takes into account the full set of inferential data in order to explain the entire model. It provides a broad overview of how the model interacts with each input instance. Since the decision-making process for every instance of input data can be clearly explained by visualizing in tree form, the well-known DT algorithm may be inherently global in nature. However, a local technique aims to provide the user with an explanation of only a few examples of test data. In particular cases, local explanations can boost user confidence and provide insight into the reasoning behind particular decisions. For DTs, a single branch in the tree may represent a local explanation. It is noteworthy that the model can gain global insights by integrating local inferences from several input cases.

A concept in XAI known as "applicability of models" de-scribes how explainable techniques can be applied to any model as a post-process or limited to certain models. The former is referred to as model-specific, whereas the latter is known as model-agnostic, which is a post-hoc method. Explainability is incorporated into the model architecture in a model-specific method, which is intrinsic and cannot be applied to other model architectures. Interpreting the weight or activation values of a neural network model, for example, is unique to that neural network learning methodology. Deep neural network model-specific methods emphasize particular areas of the input image that significantly influenced the choice by following the CNN's route in reverse order. Two examples of model-specific techniques are LRP [45] and guided backpropagation [46].

In XAI, the term "applicability of models" refers to the idea that explainable techniques can be applied to any model as a post-process or limited to certain models. The second technique is referred to as model-agnostic, while the former is known as model-specific. When explainability is incorpo-rated into the model architecture and cannot be transferred to another model architecture, this is known as a model-specific method. Interpreting the weight or activation values of a neural network model, for example, is unique to that neural network learning methodology. Deep neural network model-specific techniques highlight particular areas of the input image that significantly influenced the choice by navigating the CNN's path in reverse order. Two examples of model-specific techniques are LRP [45] and guided backpropagation [46].

While model agnostic approaches can be applied to any learn-ing strategy, they do not take into account internal factors such as weight or activation values. They do this by altering and perturbing the input data, then comparing the performance's sensitivity to the original data to uncover reasons. To put it another way, we can gauge the extent to which changing the input or weights of significant features has affected the model's performance. This will offer important information about a specific area of the input data that experienced disturbance. Other techniques include Feature Importance, GradCAM, and Occlusion Sensitivy analysis. LIME and SHapley Additive exPlanations (SHAP) are two well-liked model-agnostic tech-niques [47] - [50].

Images can be classified using a very different approach than those that use text, categories, or temporal data, like voice. For this reason, a model's input formats—numerical, visual, textual, or temporal—can be crucial in defining several ways to explain XAI techniques. The interpretations of predictions can be in many forms and depend on the end users' needs and concerns. There are four forms of explanations commonly used to interpret a prediction: numerical, visual, rule-based, and textual. Depending on the needs and concerns of the end users, forecasts might be interpreted in a variety of ways. Four types of explanations are frequently used to interpret a prediction: textual, rule-based, visual, and numerical [44]. Models typi-cally produce numerical explanations that are a measure of the

input variables that influence the model's output. They depict numerical representations such as matrices, values, or vectors of numbers. A neural network layer's probability measure can also serve as a numerical explanation. Graphical descriptions of a model's operation are most frequently provided using visual aids. Novice AI model end users can quickly under-stand a visual explanation. For individual predictions, textual explanations are typically used because they are accurate and detailed. The significant computational cost of this explanation method, which requires natural language processing (NLP), makes it uncommon. Nonetheless, they are mostly created for a local audience and are readily comprehensible to humans. Compared to textual and visual explanations, rule-based ex-planations are more ordered and straightforward. They are easy for people to understand and can be used to explain how models with IF-THEN rules or trees with AND/OR operators anticipate outcomes [44].

XAI Frameworks

Table II categorizes XAI frameworks based on scope, ap-plication, implementation, and interpretable forms for various popular XAI tools in the literature.

SHapley Additive exPlanations SHAP is a XAI technique that applies a weight, or Shapley value, to each feature in a trained model. All potential weighted input combinations are observed to have these qualities with a given weight. Each Shapley value-added feature's contribution is evaluated for all conceivable weighed input combinations based on efficiency, symmetry, features with no zero contributions, and cumulative contribution of a feature with sub-parts. SHAP consistently performs well and predicts accurately within its small scope [53].

Saliency Map (SM) Another key idea in deep learning is the SM, which Simonyan et al. initially proposed [51]. In SM, every pixel of an AD-classified image is eliminated and then processed, in contrast to an occlusion map, which hides parts of the input image with a black patch and produces a heatmap. The resulting heatmap is examined for differences in probability; a low probability suggests that the deleted pixel is crucial to the classification of AD [52].

Gradient-weighted Class Activation Mapping GradCAM is an approach for increasing the transparency of CNN models by selecting the most essential portions of an input image for prediction. GradCAM uses gradient information from the CNN model's output layer to create a localization map that highlights key places in an image. This is accomplished by providing an important value to each neuron for making certain decisions. GradCAM produces a heatmap that identifies key zones for prediction and explanation [54].

Local Interpretable Model-Agnostic Explanations LIME is a freebie that generates explanations for a single instance rather than the complete dataset, thus the word local. LIME gives explanations by manipulating the model's input data, building a surrogate model, analyzing the changes in pre-diction, and picking the most significant aspects. The LIME model is agnostic and may be applied to any blackbox model

after prediction training. LIME can analyze picture classifi-cations and explain text-based models and tabular datasets in textual, numeric, or visual formats for blackbox explanations [55].

textbfLayer-wise Relevance Propagation (LRP) LRP, like GradCAM, provides a heatmap with highlighted sections of an image. LRP is utilized in CNN, where the inputs might be photos or videos. LRP assigns relevance scores to all neurons of a certain output in the last layer of a CNN. LRP generates a heat map based on the final relevance score to indicate influential places for prediction [56].

Occlusion Sensitivity Analysis To accurately forecast Alzheimer's disease, it's important to pinpoint portions of the image that contribute to the classification. Zeiler and Fergus first proposed the OSA approach [57]. This approach creates a heatmap by obscuring or hiding areas of the input image with a gray or black patch. Variations in the output likelihood of an obscured image are detected [58]. If the most crucial region is occluded, it will have a significant impact with a low chance. As a result, an occlusion sensitivity map is utilized to identify key areas of the image that are responsible for AD.

SEARCH APPROACH

The following section describes the general stages required in doing a systematic review, including searching for and finding relevant papers. Fig. 5 shows the approach considered for the study. This summary will look into research articles that employ XAI in diagnosis/early detection and interpret the reasons for classification. To uncover contributions and summarize the findings, published articles on artificial intel-ligence and related disciplines are reviewed. This study aims to highlight research gaps and encourage XAI-based research for AD detection.

Fig. 5. Research Strategy

Research Question: The primary objective of formulating research questions is to establish a clear strategy for obtaining papers only from the targeted areas of study. In this manner, the reader will be able to comprehend the information more thoroughly.

Which XAI-incorporating AI systems are available for AD research?

For blackbox interpretability, which XAI techniques are employed to identify AD?

Which XAI frameworks are available in the literature for the detection of AD in healthcare as a whole?

What are the established advantages of applying XAI to AD?

What are the constraints, difficulties, requirements, and future potential of XAI in the detection of AD and healthcare in general?

Choosing the right search terms is one of the difficult aspects for an inclusive and thorough systematic review. The search terms used for this study were chosen with care to ensure that they are neither too general to weed out irrelevant publications nor too specific to miss pertinent articles [59]. Articles were chosen from the popular databases IEEE Explore, ACM Digi-tal Library, SpringerLink, ScienceDirect, and PubMed. Below is the inclusion and exclusion criteria used innthis article - **Inclusion Criteria:** Research on the use of AI methods for AD diagnosis, Explainable AI research for AD prediction and research on the performance outcomes of ML/DL models for AD.

Exclusion Criteria: Magazines, proceedings, editorials, and pilot papers, Articles unrelated to the diagnosis of AD and AI-based AD and Detecting AD is not covered in the article.

OUTCOMES AND DISCUSSIONS

This section presents the results of our thorough review of the articles using the RQs set.

AI Systems for AD Research Using XAI ”Which XAI-incorporating AI systems are available for AD research?” is the first research question that this section attempts to answer.The use of AI in the diagnosis, treatment, and prognosis of disease began in the early 1970s and has grown significantly over time. XAI was not used in research on AI-based AD prediction until the last ten years [60]. A growing desire for explainability and openness in healthcare and medical practice has led to the recent integration of XAI into AI-based AD prediction, even though the time has not yet come for computers to replace doctors. Numerous research that use XAI in AI-based AD detections have been found. Numerical datasets have been utilized in numerous studies to train AI models, with results that can be explained. Sappagh El et al [61]. AD and mild cognitive impairment (MCI), and use XAI techniques to interpret the predictions. A trustworthy multi-class classification model backed by XAI techniques is put out by Xu and Yan to precisely explain the predictions [62]. Sha et al. suggest a computer method called Systems Metabolomics utilizing Interpretable Learning and Evolution (SMILE) [63]. In order to comprehend and diagnose the onset and progression

of disease, this paper employs supervised metabolomics data analysis and the XAI approach to learn and identify the most informative compounds. In order to classify AD, Hammond et al. examine beta-amyloid, tau, and neurodegenerative biomark-ers. The author also employs XAI techniques to determine which biomarker has the greatest impact on AD detection [64]. According to Bloch and Friedrich, the various causes of AD can result in inconsistent disease patterns, protocols for obtaining scans, and MRI scan preprocessing problems that contribute to incorrect machine learning classification [65]. This study examines whether employing an automated and equitable data valuation strategy based on XAI techniques can improve ML classification by choosing the most informative people from the ADNI and Australian Imaging Biomarker and Lifestyle (AIBL) cohorts. The top three models from ”The Alzheimer’s Disease Prediction of Longitudinal Evolution” (TADPOLE) competition are compared by Hernandez et al. in terms of prediction and interpretability using a common XAI framework [66]. Using an interpretable machine learning approach, Chun et al. [126] attempt to increase the predicting power of the development from amnestic MCI to AD [48]. Nu-merical input values from neuropsychological and apolipopro-tein test datasets are used in this investigation. In order to give an early diagnosis of AD, Sidulova et al. suggest a unique method for categorizing Electroencephalogram (EEG) data. Using EEG recordings from people with probable AD, MCI, and HC, the study’s XAI technique offers quantitative elements that aid in making the prediction [67]. In order to train AI-

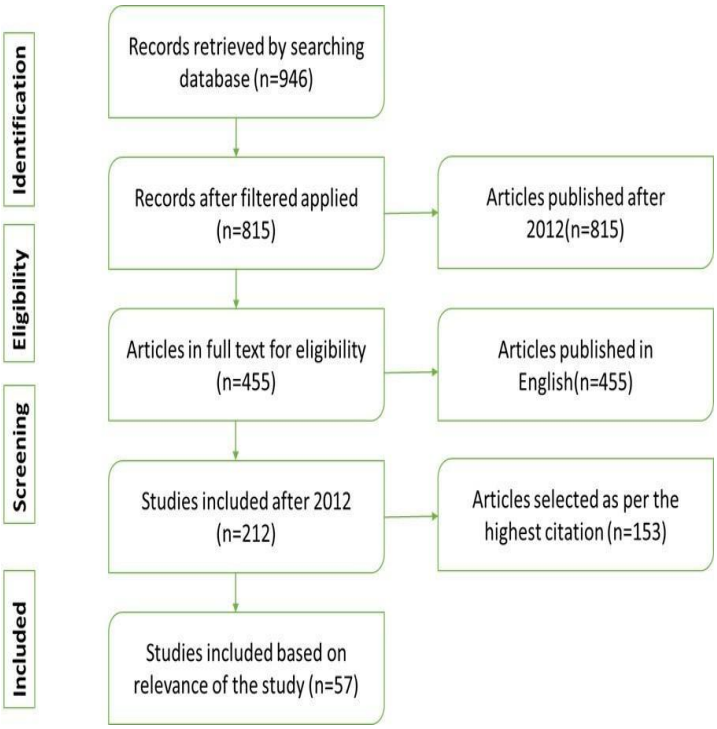

Fig. 6. PRISMA Chart

based AD detection models using MRI as input data, numerous research articles have used datasets such as ADNI, OASIS, and Kaggle data. The following table suggests deep neural network classifiers for HC, MCI, and AD prediction and classification. Each of these articles selects MRI as the input and uses ADNI datasets. To identify and classify dementia into many categories, Jain et al. provide a DCGAN-based Augmentation and Classification (D-BAC) model technique based on the prevalence and severity of dementia in the available MRI [68].

The datasets of MRI scans used for this purpose are gathered from Kaggle.

XAI Techniques for Deciphering Blackbox Models to Dis-cover AD: This section discusses the research question, For blackbox interpretability, which XAI techniques are employed to identify AD?

Finding the quantity and kind of XAI techniques now accessi-ble for the blackbox interpretability in AD detection is the goal of this research inquiry. In the XAI context of AD detection research, it offers crucial information such comprehending the main actions made to be local/global, posthoc/ante-hoc, and model agnostic/model-specific.

Why do certain explanations have a local AND global scope, a local limit, or a global limit?

Why do certain blackbox models—like Random Forest (RF)—appear in many XAI approach categories?

What makes CNN a model-specific as well as model-aginostic?

What makes a XAI approach both posthoc and model-agnostic at the same time?

For what reason would a XAI technique be regarded as both Antehoc and Model-Specific at the same time?

In order to address the research question with greater clarity, we respond to these questions.

Certain XAI techniques exhibit either global or locally inter-pretable behavior. However, by combining the local explana-tions, the researcher has the right to utilize those methodolo-gies to interpret internationally. Consequently, there is no hard and fast rule stating that a model can either be local or global. For example, local interpretation is the main application for the XAI framework SHAP. But a global population can also be interpreted using SHAP. Likewise, by combining local explanations, the LIME approach, a local explainer, may also be applied to global understanding.

Blackbox models Both XGBoost and LGBM are tree-based models; XGBoost classifies data level-wise, whereas LGBM classifies data branch-wise. As a branch-wise classifier, LGBM's explainability can therefore be local. Global expla-nations can then be established by aggregating local results. The ultimate path to the last level can be seen as a global explanation, and XGBoost can achieve a local description at each tree level due to its level-wise classification. Therefore, while being blackbox models, LGBM and XGBoost can be aggregated both locally and internationally. A CNN model's prediction can be explained using a model-agnostic method that doesn't alter the internal layers (like kernel SHAP).

These blackbox models, in which the inner workings of the model are left unaltered, are the major applications for post-hoc models. The resulting prediction needs to be subjected to a XAI process in order to generate explainability.This idea might be referred to as model-agnostic or post-hoc.

The key components of a training model are naturally available in the ante-hoc model. One can use either a model-specific or model-agnostic XAI approach to extract explainability from an ante-hoc model. An ante-hoc model's training requires explainability to be included, while a model-specific XAI approach requires the inner working details.

Frameworks for AD Detection with XAI: This section answers the third research question: Which XAI frameworks are accessible in the literature and are applied to the detection of AD?

Finding the XAI frameworks and methods employed in the research to unravel AI-based AD categorization is the goal of this Research Question. Researchers, developers, and sub-ject matter experts will be inspired to understand the inner workings of a machine-learning model by the discussions in this section. Explainable embedded machines, especially in healthcare, can significantly minimize the time medical per-sonnel spend on repeated patient examinations and spend time concentrating on deciphering disease diagnoses. To address the issue of blackbox models—predictions that are extremely accurate but the inner workings are concealed—there are numerous XAI frameworks available. Among the numerous well-known XAI frameworks that are widely utilized in AD and relevant to RQ3 are LIME, SHAP, and GradCAM. A well-liked technique for straightforward human readings of predictive models is LIME. The research employ a variety of ML/DL classifiers, such as CNN, SpinalNet, kNN, XG-Boost,

SVM, and transfer-based model BERT, to interpret the predictions using LIME. These kinds of datasets, including as MRI, gene expressions, EEG signals, and linguistic or textual data, have been used by the classifiers in the investigations. Kamal et. al [77] suggested a study that uses MRI scans and gene expression to classify people into four categories: mild dementia, moderate dementia, no dementia, and very mild dementia. Using MRI with CNN and gene expressions with kNN and XGBoost, the author employs LIME to provide local explanations for AD classifications. LIME was helpful in determining and prioritizing the important feature sets as-sociated with an AD patient based on likelihood values. Users can learn which features have a good and negative impact on the prediction thanks to LIME. The probability values can be used to describe trust, even though trust cannot be defined. In order to distinguish between control and dementia patients, Illias and Askounis [78] perform a comprehensive linguistic analysis using a dataset of medical transcripts using the transfer learning model, BERT, with the co-attention mechanism using the co-attention mechanism to distinguish between people with dementia and those in control.

Game theory is used by SHAP, which is model-agnostic, to explain the results of any machine learning model. The SHAP framework is another popular XAI framework, according to this review. MRI and PET scan volumetric measurements, Apolipoprotein measurements, Mini-Mental state tests, Clin-ical Dementia scores, and demographic data are among the types of datasets used in the majority of studies. Finding the Shapley values for each sample characteristic that requires comprehension is the basic idea behind SHAP.

3. ADVANTAGES OF DETECTING AD USING XAI TECHNIQUES

The fourth research question, ”What are the established advantages of applying XAI to AD?” is addressed in this section. According to the review, there are a number of advantages to using the idea of XAI to AI-based AD detection. The majority of research has attempted to report model transparency, fairness, and correctness. When employing AI models for prediction, they have emphasized the value of XAI in building trust and confidence, especially in the medical sector. Benefits from independent research have indicated a responsible approach to XAI’s AI development. The advan-tages of the chosen research are categorized in this section according to the four types of explanation: textual, rule-based, visual, and numerical. Researchers will find this classification useful in determining which theories, given the available data modalities, should be pursued.

Textual: Promising classification findings are obtained in the field of dementia detection using transcripts and the transformer-based network, BERT, developed by Ilias et al. [78]. The authors use LIME to demonstrate how transcripts clarify the distinction between people with dementia and those without.Different colors are applied to the textual forms or tokens in transcripts to denote which ones belong to a control group. The significance of these markers for the final transcript classification is shown by the tokens’ color intensity.

Rule-based: In our review, we discovered two articles that provide rules as explanations. One project uses AI agents and the Internet of Things (IoT) to remotely check on the health of senior citizens. In order to help with the early detection of cognitive decline, Khodabandehloo et al. [75] provide a novel HealthXAI system that uses a DT regression algorithm. The system also provides caregivers with high-level numerical scores for reporting inappropriate behaviors and natural language explanations of the forecasts. The decision rule forecasts the target variable’s value and describes it in natural language as either HC or AD. The diagnostic tool presented by Garc´ıa-Gutierrez et al. [79] uses a DT that gives doctors a clear and straightforward set of choice rules to help them understand the pathophysiology of AD and behavioral frontotemporal dementia (bvFTD).

Visual: CNN and 3D CNN are data models used in studies that employ LRP as an AI explanation. When recognizing brain atrophy, LRP offers visual explanations in the form of heat maps of important brain regions. The amygdala, entorhinal cortex, and hippocampal regions are among the important elements identified by LRP for interpretation. The use of LRP with guided backpropagation to find heat maps with pertinent, important features is covered by Bo¨hle et al. [80]. Similar noteworthy features have been found by Pohl et al. [81] utilizing composite LRP with several propagation rules. According to the author, verbal semantic memory and visual memory issues are caused by injury to the left and right temporal lobes, respectively. A DL model’s predictions are visually explained in a number of experiments using

the GradCAM XAI tool. A VGG16 deep learning model’s predictions are visually explained by Ruengchaijatuporn et al.

[72] using GradCAM.

Numeric: By selecting the appropriate XAI approach, classi-fier, and available data, Salih et al. [82] attempt to create a proxy that will verify the explanation’s stability. By quanti-fying the informative predictors, the authors

have employed Principal Component Analysis (PCA) to confirm the stability of the identified predictors with the selected parameters. The method used in this study to associate predictors with SHAP and the proxy PCA yields uncorrelated variables that provide stable ranking for the majority of classifiers. The results are useful to the medical community because XAI is widely used in sensitive areas, such as the prognosis of long-term mortality, admission to critical care units, and ejection failure.

4. CHALLENGES, LIMITS, REQUIREMENTS, AND FUTURE OF XAI IN AD DETECTION

This part tackles research question ”What are the con-straints, difficulties, requirements, and future potential of XAI in the detection of AD and healthcare in general?”

The XAI idea has been developed in a number of research in recent years to better explain the decisions made by AI systems. The advent of high-performance computers and easy access to a number of XAI frameworks with easily accessible source code have made it possible to integrate these explainers into stand-alone AI systems with ease. Despite the encouraging findings of independent research, it should come as no surprise that these efforts have a number of limitations. In order to stimulate more study in this area, we have listed a number of XAI-based AD detection limitations and research gaps.

Without consulting a medical expert, XAI researchers frequently rely on their own instincts to evaluate what constitutes a sound explanation [83]. In order to maximize the benefits to stakeholders, medical and AI profession-als must work together to determine the interpretability developed by the XAI framework.

The lack of ground truth data is one of the major problems with XAI-based AD diagnosis [84]. There are a number of clinical biomarker and neuroimaging datasets for AD, but none offer ground truth to support the explainability that XAI models elicit.

Additionally, when given to individuals with different degrees of topic competence, the effects of XAI expla-nations differ significantly [80]. People become confused and start to doubt the XAI systems’ paradoxical relation-ship when they see explanations that go against their own intuition.

In computer-aided diagnosis, where an incorrect pre-diction is nearly always fatal, confidence measures are essential. If the system is unable to provide a reliable prediction, a manual intervention must be justified in order to reach a suitable conclusion. Therefore, before offering explanations, XAI techniques must additionally include a confidence score to detect instances in which

the classifier is wrong. If not, the end user could instill erroneous confidence in the system [85].

Several XAI frameworks were utilized in some articles to improve explainability. From an intellectual perspec-tive, it might be beneficial, but in practice, it adds to the already existing ambiguity. For example, one study combined the use of the SHAP and LIME frameworks [69].

The sparse use of medical datasets or the absence of a thorough benchmark dataset with variations that reflect real-world situations is another serious flaw in practically all of the research we looked at [86].

Even while some research used multimodal data to pre-dict AD [66] [77], only one modality’s explanations were drawn.

MRI volumetry, cortical thickness, and other disease indicators that are highly correlated with dementia were not taken into account by certain research, despite the fact that they employed XAI methods to predict AD [72] [47].

The majority of research has not identified variables that impact model performance and the resulting explainabil-ity, such as hyperparameter settings, the split percentage of train-test data, data preparation, etc. [74].

Medical professionals’ hesitation to trust AI solutions is further exacerbated by the technology’s incapacity to take into account the history of abnormalities that led to cognitive deterioration.A fundamental drawback for any medical area, not alone AI-based AD diagnosis, is the absence of real-world labelled data sets of people gathered over an extended period of time [75].

The AI forecast and the related brain region did not always correlate, despite the fact that researchers used one or more XAI frameworks in the AD prediction [67].

5. CONCLUSION

Over the past few years, explainable AI has become increasingly important due to scientific needs and regulatory compliance. Different XAI frameworks that describe the model's accuracy, rationality, and clarity in AI-assisted decision-making—all of which are crucial in the healthcare industry—are being investigated by researchers. Extubation failures and long-term mortality are two examples of ex-pectations that XAI helps to solve effectively by fostering synergistic ecosystems. Therefore, it is essential to encourage a broader distribution of XAI concepts, backgrounds, and approaches across the research community.

This paper provides a thorough overview of the use of XAI models and frameworks on multimodal AD data in order to achieve this goal and act as a reference. Over the past ten years, we have examined papers that use XAI to diagnose AD. Re-search publications that were carefully examined using well-crafted research questions were included in the study. Several ML and DL models that have incorporated XAI frameworks to integrate transparency and integrity in AI predictions were revealed by the RQs, which also highlighted numerous XAI-

based studies used for AD diagnosis.In addition, the study identifies a number of advantages, drawbacks, and potential directions for clinical diagnosis. We acknowledge that it is too soon to make any comments about closing the gap between the medical and AI domains to zero. However, these assessments will highlight the advantages and drawbacks for the research community, allowing the trade-off between explainability and accuracy in AI solutions to be resolved with a satisfactory degree of precision.In order to fully utilize AI's potential in promoting accuracy in clinical decision support systems, this review will assist in investigating numerous healthcare domains.

Author Contribution To finish this project, all authors collaborated closely. S.Patwardhan came up with the idea that was given. Finding the pertinent literature for this study included all of the writers. The findings of this work were monitored by M.M. More. The authors of the manuscript were M.L.Kulkarni, D.P.Yadav, R.S. Zirmite and M.P.Shirurkar. Every author has made edits to the manuscript. Every author has reviewed and approved the manuscript's final draft.

References:

1. Shaffi, Noushath, et al. "Triplet-loss based Siamese convolutional neural network for 4-way classification of Alzheimer's disease." International Conference on Brain Informatics. Cham: Springer International Publish-ing, 2022.
2. Webster, Claire, et al. "Better and Earlier Detection of Dementia in a Changing Landscape; November 2–4, 2023; Toronto, Ontario." (2024).
3. Dubois, Bruno, Gaetane Picard, and Marie Sarazin. "Early detection of Alzheimer's disease: new diagnostic criteria." Dialogues in clinical neuroscience 11.2 (2009): 135-139.
4. Tatulian, Suren A. "Challenges and hopes for Alzheimer's disease." Drug discovery today 27.4 (2022): 1027-1043.
5. Knopman, David S., et al. "Alzheimer disease." Nature reviews Disease primers 7.1 (2021): 33.
6. Yahaya, Salisu Wada, Ahmad Lotfi, and Mufti Mahmud. "Towards the development of an adaptive system for detecting anomaly in human activities." 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2020.
7. Lalotra, Gotam Singh, et al. "iReTADS: An Intelligent Real-Time Anomaly Detection System for Cloud Communications Using Temporal Data Summarization and Neural Network." Security and Communication Networks 2022.1 (2022): 9149164.
8. Fabietti, Marcos, et al. "Neural network-based artifact detection in local field potentials recorded from chronically implanted neural probes." 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.
9. Rahman, Shelia, Tanusree Sharma, and Mufti Mahmud. "Improving al-coholism diagnosis: comparing instance-based classifiers against neural networks for classifying EEG signal." International Conference on Brain Informatics. Cham: Springer International Publishing, 2020.
10. Wadhera, Tanu, and Mufti Mahmud. "Computing hierarchical complex-ity of the brain from electroencephalogram signals: a graph convolu-tional network-based approach." 2022 international joint conference on neural networks (IJCNN). IEEE, 2022.
11. Ahmed, Sabbir, et al. "Toward machine learning-based psychological assessment of autism spectrum disorders in school and community." Proceedings of trends in electronics and health informatics: Tehi 2021. Singapore: Springer Nature Singapore, 2022. 139-149.
12. Das, Sahana, et al. "A machine learning pipeline to classify foetal heart rate deceleration with optimal feature set." Scientific Reports 13.1 (2023): 2495.
13. Akter, Tania, et al. "Towards autism subtype detection through identi-fication of discriminatory factors using machine learning." International Conference on Brain Informatics. Cham: Springer International Publish-ing, 2021
14. Ghosh, Tapotosh, et al. "Artificial intelligence and internet of things in screening and management of autism spectrum disorder." Sustainable Cities and Society 74 (2021): 103189.

15. Niamat Ullah Akhund, Tajim Md, et al. ”Adeptness: Alzheimer’s disease patient management system using pervasive sensors-early prototype and preliminary results.” Brain Informatics: International Conference, BI 2018, Arlington, TX, USA, December 7–9, 2018, Proceedings 11. Springer International Publishing, 2018.
16. Jesmin, Sabrina, M. Shamim Kaiser, and Mufti Mahmud. ”Towards artificial intelligence driven stress monitoring for mental wellbeing tracking during COVID-19.” 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). IEEE, 2020.
17. Mahmud, Mufti, et al. ”Deep learning in mining biological data.” Cognitive computation 13.1 (2021): 1-33.
18. Davenport, Thomas, and Ravi Kalakota. ”The potential for artificial intelligence in healthcare.” Future healthcare journal 6.2 (2019): 94-98.
19. Mahmud, Mufti, et al. ”Applications of deep learning and reinforcement learning to biological data.” IEEE transactions on neural networks and learning systems 29.6 (2018): 2063-2079.
20. Fabrizio, C., et al. ”Artificial Intelligence for Alzheimer’s Disease: Promise or Challenge? Diagnostics 2021, 11, 1473.” 2021,
21. Noor, Manan Binth Taj, et al. ”Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer’s disease, Parkinson’s disease and schizophre-nia.” Brain informatics 7 (2020): 1-21.
22. Yang, Guang, Qinghao Ye, and Jun Xia. ”Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond.” Information Fusion 77 (2022): 29-52.
23. Rai, Arun. ”Explainable AI: From black box to glass box.” Journal of the academy of marketing science 48 (2020): 137-141.
24. Kaur, Davinder, et al. ”Trustworthy artificial intelligence: a review.” ACM computing surveys (CSUR) 55.2 (2022): 1-38.
25. Nazar, Mobeen, et al. ”A systematic review of human–computer inter-action and explainable artificial intelligence in healthcare with artificial intelligence techniques.” IEEE Access 9 (2021): 153316-153348.
26. Nazar, Mobeen, et al. ”A systematic review of human–computer inter-action and explainable artificial intelligence in healthcare with artificial intelligence techniques.” IEEE Access 9 (2021): 153316-153348.
27. Schmidt, Jonathan, et al. ”Recent advances and applications of machine learning in solid-state materials science.” npj computational materials

    5.1 (2019): 83.

28. Mamoshina, Polina, et al. ”Applications of deep learning in biomedicine.” Molecular pharmaceutics 13.5 (2016): 1445-1454.
29. Vamathevan, Jessica, et al. ”Applications of machine learning in drug discovery and development.” Nature reviews Drug discovery 18.6 (2019): 463-477.
30. Lei, Yaguo, et al. ”Applications of machine learning to machine fault diagnosis: A review and roadmap.” Mechanical systems and signal processing 138 (2020): 106587.
31. Deng, Li, and Dong Yu. ”Deep learning: methods and applications.” Foundations and trends® in signal processing 7.3–4 (2014): 197-387.
32. Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. ”” Why should i trust you?” Explaining the predictions of any classifier.” Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016.
33. Kumar, Dheeraj, and Mayuri A. Mehta. ”An overview of explainable AI methods, forms and frameworks.” Explainable AI: Foundations, Methodologies and Applications (2022): 43-59.
34. Loh, Hui Wen, et al. ”Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022).” Computer methods and programs in biomedicine 226 (2022): 107161.
35. Pawar, Urja, et al. ”Explainable AI in healthcare.” 2020 international conference on cyber situational awareness, data analytics and assessment (CyberSA). IEEE, 2020.
36. Wanner J, Herm LV, Heinrich K, Janiesch C. Stop ordering machine learning algorithms by their explainability! An empirical investigation of the tradeoff between performance and explainability. In: Conference on e-Business, e-Services and e-Society. Springer; 2021;245–58
37. Jung, Yeon-Jee, Seung-Ho Han, and Ho-Jin Choi. ”Explaining CNN and RNN using selective layer-wise relevance propagation.” IEEE Access 9 (2021): 18670-18681
38. Gade K, Geyik S, Kenthapadi K, Mithal V, Taly A. Explainable AI in industry: Practical challenges and lessons learned. In: Companion Proceedings of the Web Conference 2020;303–4
39. Tao J, Xiong Y, Zhao S, Wu R, Shen X, Lyu T, et al. Explainable AI for Cheating Detection and Churn Prediction in Online Games. IEEE Transactions on Games. 2022
40. Fulton LB, Lee JY, Wang Q, Yuan Z, Hammer J, Perer A. Getting playful with explainable AI: games with a purpose to improve human understanding of AI. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems; 2020;1–8
41. Fellous JM, Sapiro G, Rossi A, Mayberg H, Ferrante M. Explain-able artificial intelligence for neuroscience: behavioral neurostimulation.

    Frontiers in neuroscience. 2019;13:1346

42. Chen K, Hwu T, Kashyap HJ, Krichmar JL, Stewart K, Xing J, et al. Neurorobots as a means toward neuroethology and explainable AI.

Frontiers in Neurorobotics. 2020;14

43. Vultureanu-Albis¸i A, Ba˘dica˘ C. Recommender systems: an explainable AI perspective. In: 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA). IEEE; 2021. p. 1-6
44. Kumar D, Mehta MA. 3. In: An Overview of Explainable AI Meth-ods, Forms and Frameworks. Cham: Springer International Publishing; 2023;43–59. Available from: https://doi.org/10.1007/978-3-031-12807-3-3.
45. Montavon G, Binder A, Lapuschkin S, Samek W, Mu¨ller KR. Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning. 2019:193-209
46. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striv-ing for simplicity: The all convolutional net. arXiv preprint http:// arxiv.org/abs/1412.6806. 2014.
47. Rieke J, Eitel F, Weygandt M, Haynes JD, Ritter K. Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. In: Understanding and Interpreting Machine Learning in Medical Image Computing Applications. Springer; 2018. p. 24-31
48. Chun MY, Park CJ, Kim J, Jeong JH, Jang H, Kim K, et al. Prediction of conversion to dementia using interpretable machine learning in patients with amnestic mild cognitive impairment. Frontiers in Aging Neuroscience. 2022;14.
49. Ribeiro MT, Singh S, Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016;1135–44
50. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fer-gus R, Vishwanathan S, et al. editors. Advances in Neural Informa-tion Processing Systems 30. Curran Associates, Inc. 2017;4765–74. Available from: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.
51. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint http://arxiv.org/abs/1312.6034. 2013.
52. Petsiuk V, Jain R, Manjunatha V, Morariu VI, Mehra A, Ordonez V, et al. Black-box explanation of object detectors via saliency maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021;11443–52.
53. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fer-gus R, Vishwanathan S, et al. editors. Advances in Neural Informa-tion Processing Systems 30. Curran Associates, Inc. 2017;4765–74. Available from: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.
54. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision. 2019.
55. Ribeiro MT, Singh S, Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016;1135–44.
56. Montavon G, Binder A, Lapuschkin S, Samek W, Mu¨ller KR. Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning. 2019:193-209
57. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer; 2014;818–33.
58. Folego G, Weiler M, Casseb RF, Pires R, Rocha A. Alzheimer's disease detection through whole-brain 3D-CNN MRI. Frontiers in bioengineer-ing and biotechnology. 2020;8
59. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic reviews. 2021;10(1):1–11
60. Yang C, Rangarajan A, Ranka S. Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. In: AMIA annual symposium proceedings. vol. 2018. American Medical Informatics Association. 2018;1571.
61. El-Sappagh S, Alonso JM, Islam SMR, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explain-able artificial intelligence for Alzheimer's disease. Scientific Reports. 2021;11(1):2660
62. Xu X, Yan X. A Convenient and Reliable Multi-Class Classification Model based on Explainable Artificial Intelligence for Alzheimer's Dis-ease. In: 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). IEEE; 2022;671–5.
63. Sha C, Cuperlovic-Culf M, Hu T. SMILE: systems metabolomics using interpretable learning and evolution. BMC bioinformatics. 2021;22(1):1–17.
64. Hammond TC, Xing X, Wang C, Ma D, Nho K, Crane PK, et al. B-amyloid and tau drive early Alzheimer's disease decline while glucose hypometabolism drives late decline. Communications biology. 2020;3(1):1–13.
65. Bloch L, Friedrich CM. Data analysis with Shapley values for automatic subject selection in Alzheimer's disease data sets Cognitive Computation (2024) 16:1–44 43 1 3 using interpretable machine learning. Alzheimer's Research and Therapy. 2021;13(1):1–30.

66. Hernandez M, Ramon-Julvez U, Ferraz F. Withthe ADNIConsortium. Explainable AI toward understanding the performance of the top three TADPOLE Challenge methods in the forecast of Alzheimer's disease diagnosis. PloS one. 2022;17(5):e0264695.
67. Sidulova M, Nehme N, Towards Park CH. Analysis Explainable Image, for Alzheimer's Disease and Mild Cognitive Impairment Diagnosis. In,. IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE. 2021;2021:1–6.
68. Jain V, Nankar O, Jerrish DJ, Gite S, Patil S, Kotecha K. A novel AI-based system for detection and severity prediction of dementia using MRI. IEEE Access. 2021;9:154324–46
69. Loveleen G, Mohan B, Shikhar BS, Nz J, Shorfuzzaman M, Masud M. Explanation-driven HCI Model to Examine the Mini-Mental State for Alzheimer's Disease. ACM
70. Bo¨hle M, Eitel F, Weygandt M, Ritter K. Layer-wise relevance prop-agation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. Frontiers in aging neuroscience. 2019;11:194
71. ombardi A, Diacono D, Amoroso N, Biecek P, Monaco A, Bellantuono L, et al. A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of Mild Cognitive Impairment and Alzheimer's Disease. Brain informatics. 2022;9(1):1–17
72. Ruengchaijatuporn N, Chatnuntawech I, Teerapittayanon S, Sriswasdi S, Itthipuripat S, Hemrungrojn S, et al. An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks. Alzheimer's Research and Therapy. 2022;14(1):1–11.
73. Lai Y, Lin X, Lin C, Lin X, Chen Z, Zhang L. Identification of endo-plasmic reticulum stress-associated genes and subtypes for prediction of Alzheimer's disease based on interpretable machine learning. Frontiers in Pharmacology. 2022;13.
74. Bogdanovic B, Eftimov T, Simjanoska M. In-depth insights into Alzheimer's disease by using explainable machine learning approach.

Scientific Reports. 2022;12(1):1–26

75. Khodabandehloo E, Riboni D, Alimohammadi A. HealthXAI: Collab-orative and explainable AI for supporting early diagnosis of cognitive decline. Future Generation Computer Systems. 2021;116:168–89.
76. Yu L, Xiang W, Fang J, Chen YPP, Zhu R. A novel explainable neural network for Alzheimer's disease diagnosis. Pattern Recognition. 2022;131
77. Kamal MS, Northcote A, Chowdhury L, Dey N, Crespo RG, Herrera-Viedma E. Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes. IEEE Transactions on Instrumentation and Measurement. 2021;70:1–7
78. Ilias L, Askounis D. Explainable identification of dementia from tran-scripts using transformer networks. IEEE Journal of Biomedical and Health Informatics. 2022;26(8):4153–64.
79. Garc´ıa-Gutierrez F, D´ıaz-A´ lvarez J, Matias-Guiu JA, Pytel V, Mat´ıas-Guiu J, Cabrera-Mart´ın MN, et al. GA-MADRID: Design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms. Medical and Biological Engineering and Computing. 2022;60(9):2737–56
80. Bo¨hle M, Eitel F, Weygandt M, Ritter K. Layer-wise relevance prop-agation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. Frontiers in aging neuroscience. 2019;11:194
81. Pohl T, Jakab M, Benesova W. Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease. International Journal of Imaging Systems and Technology. 2022;32(2):673–86.
82. Salih A, Galazzo IB, Cruciani F, Brusini L, Radeva P. Investigating Explainable Artificial Intelligence for MRI-based Classification of De-mentia: a New Stability Criterion for Explainable Methods. In: 2022 IEEE International Conference on Image Processing (ICIP). IEEE; 2022;4003–7.
83. Kou Y, Gui X. Mediating community-AI interaction through situated explanation: the case of AI-Led moderation. Proceedings of the ACM on Human-Computer Interaction. 2020;4(CSCW2):1–27
84. Slijepcevic D, Horst F, Lapuschkin S, Horsak B, Raberger AM, Kranzl A, et al. Explaining machine learning models for clinical gait analysis.

ACM Transactions on Computing for Healthcare. 2021;3(2):1–27

85. Arrotta L, Civitarese G, Bettini C. DeXAR: Deep Explainable Sensor-Based Activity Recognition in Smart-Home Environments. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technolo-gies. 2022;6(1):1–30.
86. Winterburn JL, Voineskos AN, Devenyi GA, Plitman E, de la Fuente-Sandoval C, Bhagwat N, et al. Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multimethod and multi-dataset study. Schizophre-nia Research. 2019;214:3–10