Artificial Intelligence Review (2025) 58:44
https://doi.org/10.1007/s10462-024-11041-5

Applications of deep learning in Alzheimer's disease: a systematic literature review of current trends, methodologies, challenges, innovations, and future directions

Shiva Toumaj[1] · Arash Heidari[2,3] · Reza Shahhosseini[4] · Nima Jafari Navimipour[5,6,7]

Published online: 20 December 2024

Abstract
Alzheimer's Disease (AD) constitutes a significant global health issue. In the next 40 years, it is expected to affect 106 million people. Although more and more people are getting AD, there are still no effective drugs to treat it. Insightful information about how important it is to find and treat AD quickly. Recently, Deep Learning (DL) techniques have been used more and more to diagnose AD. They claim better accuracy in drug reuse, medication recognition, and labeling. This essay meticulously examines the works that have talked about using DL with Alzheimer's disease. Some of the methods are Natural Language Processing (NLP), drug reuse, classification, and identification. Concerning these methods, we examine their pros and cons, paying special attention to how easily they can be explained, how safe they are, and how they can be used in medical situations. One important finding is that Convolutional Neural Networks (CNNs) are most often used for AD research and Python is most often used for DL issues. Some security problems, like data protection and model stability, are not looked at enough in the present research, according to us. This study thoroughly examines present methods and also points out areas that need more work, like better data integration and AI systems that can be explained. The findings should help guide more research and speed up the creation of DL-based AD identification tools in the future.

Keywords Alzheimer's disease · Deep learning · Cognitive impairment · Machine learning · Neuroimaging · Neurodegenerative diseases

1 Introduction

Alzheimer's Disease (AD), the leading cause of dementia, is increasingly posing a significant challenge for global healthcare. The incidence of cases is projected to rise dramatically, increasing from 27 million to an alarming 106 million during the next four decades (Ching et al. 2024). This increase will profoundly impact individuals globally, with around 1 in

Extended author information available on the last page of the article

85 expected to be affected. Regrettably, AD ranks as the third leading cause of mortality globally, behind cardiovascular disease and cancer. It has become one of the most dreaded diseases globally, surpassing even cancer. This neurological disorder progressively impairs cognitive function and memory as neuronal cells degenerate and diminish, which can be severely detrimental. AD is marked by the accumulation of amyloid-beta (Aβ) plaques and tau protein aggregates in the brain, resulting in cognitive deterioration, memory impairment, and speech difficulties. Even though we need to fight AD immediately, the drugs we have now do not offer a fix (Menagadevi et al. 2024). This shows how important it is to find and treat AD early (Bhandarkar et al. 2024). MRI and Positron Emission Tomography (PET) are non-invasive imaging methods that have helped scientists learn a lot about the structural and functional changes in the brain that happen in people with ad and Mild Cognitive Impairment, MCI (Sorour et al. 2024a). PET imaging helps find areas where Aβ proteins are building up, and MRIs give very clear pictures of the structure of the brain that can cause loss due to AD in certain brain areas (Xu et al. 2024). But it is hard to figure out what PET and MRI images mean because different people see different things in the brain and understand them differently. This is why objective and quantitative diagnosis methods were created (Zhou et al. 2024).

Because of this, computer-aided diagnosis systems have become very important. They use Machine Learning (ML) and Deep Learning (DL) to make AD identification more accurate and quick (Srividhya et al. 2024). ML, a branch of Artificial Intelligence (AI), teaches computers to find patterns and guess what will happen next by using data (Bakare et al. 2024). DL is a part of ML that uses complicated methods and needs a lot of training data. It is used in Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN), and Convolutional Neural Networks (CNN) (Xu et al. 2023). Using ML and DL together in medical imaging has a lot of potential to change how diseases are diagnosed and how they are treated (Nanthini et al. 2024). AI-based methods are at the cutting-edge of AD study and technology (Ferrante et al. 2024; Illakiya and Karthik 2023a). This is due to the increased speed of computers, the enhancement of algorithms, and our ability to access vast databases. The newly developed methodologies, encompassing ML and DL, are prepared to address the challenging problems of detecting AD (Bapat et al. 2023; Zhang et al. 2023a). Researchers want to employ AI to facilitate early detection of AD, enhance the efficacy of individualized therapies, and ultimately improve outcomes for individuals with AD, as well as refine diagnostic methodologies using quantitative measures (Pradhan et al. 2024; Illakiya et al. 2024).

Numerous investigations on DL and AD have been published in academic literature. However, there is a noticeable absence of additional and complete studies that were conducted using well-defined criteria. These reviews are critical for informing researchers about the most recent advances in DL-based AD detection systems while also giving guidance for the continued development of domain-specific methods. Research in this field is fast evolving, and previous research, according to our knowledge, did not adequately evaluate explainability, datasets, security, ethical issues, the usefulness of DL–AD applications, medical precision, and clinician evaluations. This article tries to enhance earlier surveys by looking into unique fundamentals and restrictions. Unlike other research, this one focuses on diverse DL–AD application aspects. Because there are several DL application fields and endeavors in AD, this study does a Systematic Literature Review (SLR) of the most current advancements in DL algorithms and evaluation measurements. Our study provides an

overview of the DL approaches that have contributed to numerous AD research inquiries. We categorized these approaches into four categories: AD detection, AD classification, drug repurposing, and Natural Language Processing (NLP). By proposing this categorization, we seek to inform physicians and other healthcare professionals about the use of these methods and raise awareness of the various DL applications in AD. An additional goal of ours is to promote the development of data-driven, computationally based AD training. This study aims to advance readers' understanding of DL applications in AD, the latest developments in the field, the benefits and drawbacks of existing strategies, a comprehensive examination of research methodologies and findings, and areas requiring additional investigation and assessment. The main contributions of this article are as follows:

- We introduce a structured taxonomy to categorize DL methods used for AD detection, classification, drug repurposing, and natural language processing (NLP), highlighting their diverse applications;
- We present an in-depth examination of DL applications in AD, covering their strengths, limitations, and security issues, with a focus on identifying gaps in current literature, particularly in underexplored areas like data privacy and model robustness;
- This paper conducts a systematic review of the current methodologies in DL for AD, including common simulations, evaluation metrics, and the application of XAI frameworks like Grad-CAM;
- We offer an overview of the key challenges facing DL–AD research, such as dataset limitations, model interpretability, and the lack of real-world validations, providing suggestions for future work to address these issues;
- Through the detailed analysis, we identify potential areas of improvement, including the integration of multi-modal data, better evaluation of security measures, and increased focus on clinical applicability.

The structure of the article is as follows: Sect. 2 discusses pertinent publications, while Sect. 3 delves into correlated subjects. Section 4 discusses the methodology and article selection. Section 5 covers article classification alongside an analysis of DL uses and methodologies in AD. Section 6 gives the results and analyses, while Sects. 7 and 8 discuss the remaining challenges and offer conclusions. Table 1 lists the acronyms used throughout the study.

2 Related reviews

Numerous scholars have provided summaries of DL applications for the early detection of AD. We incorporate several survey papers on pertinent subjects within this field to analyze the distinctions between our research and existing literature comprehensively. Previous reviews have mostly concentrated on more basic subjects in AD diagnosis; however, the SLR covers all of the current applications and their features, implementation, and discussion from a practical perspective for the AD domains.

In the realm of ongoing research, Sellappan et al. (Sellappan et al. 2024) emphasized the prospective utilization of AI technology for the classification of AD patients through MRI scans and forecasting the advancement of the ailment. They introduced sophisticated

Table 1 Abbreviation table

Abbreviation	Definition	Abbreviation	Definition
AD	Alzheimer's Disease	LR	Logistic Regression
ADNI	Alzheimer's Disease Neuroimaging Initiative	LSTM	Long Short-Term Memory
AI	Artificial Intelligence	MLP	Multi-Layer Perceptron
AIBL	Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing	MPA	Marine Predators Algorithm
AUC-ROC	Area Under the Receiver Operating Characteristic Curve	MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Networks	NLP	Natural Language Processing
CU	Cognitively Unimpaired	OASIS	Open Access Series of Imaging Studies
DL	Deep Learning	PCA	Principal Component Analysis
DNNs	Deep Neural Networks	PET	Positron Emission Tomography
DT	Decision Tree	RF	Random Forest
EEG	Electroencephalogram	RNN	Recurrent Neural Network
EHR	Electronic Health Record	RQs	Research Questions
FDA	Food and Drug Administration	SCI	Subjective Cognitive Impairment
fMRI	Functional MRI	SHAP	Shapley Additive exPlanations
GAN	Generative Adversarial Networks	SLR	Systematic Literature Review
GBM	Gradient Boosting Machine	SNP	Single Nucleotide Polymorphisms
GNN	Graph Neural Network	SVD	Singular Value Decomposition
Grad-CAM	Gradient-weighted Class Activation Mapping	SVM	Support Vector Machine
GRU	Gated Recurrent Unit	ViT	Vision Transformer
HIPAA	Health Insurance Portability and Accountability Act	XAI	Explainable Artificial Intelligence
LIME	Local Interpretable Model-Agnostic Explanations	XGBoost	Extreme Gradient Boosting

DL methodologies for the processing of medical images in AD diagnosis, aiming to facilitate the standardization and adoption of effective diagnostic approaches within the domain. Also, Shanmugavadivel et al. (Shanmugavadivel et al. 2023) offered a thorough overview of the range of input types, subjects, and classes utilized in AD research by gathering and assessing 50 articles linked to AD diagnosis datasets. Their primary focus was on examining numerous datasets, feature extraction techniques, and parameters employed in neuroimaging. Furthermore, the application of DL techniques and Functional MRI (fMRI) in the diagnosis of AD was examined by Warren and Moustafa (Warren and Moustafa 2023). They then described the preprocessing, classification, and commonly used deep neural network techniques seen in the literature. They also talked about the accuracy, advantages, disadvantages, and prospects of DL and fMRI techniques. In addition, with a focus on early diagnosis, Yao et al. (Yao et al. 2023) gave a thorough review of current advancements in DL techniques utilized for brain MRI images for the categorization of various stages of AD.

To classify and detect AD using brain MRI, Pallawi and Singh (Pallawi and Singh 2023) investigated and compared four distinct DL models: InceptionV3, ResNet50, ResNet101, and DenseNet169. Additionally, Zhao et al. (Zhao et al. 2023) evaluated some common

traditional ML techniques that are employed to classify and diagnose AD utilizing MRI images. Their study reviewed several techniques: CNN, autoencoder, DL, transformer, Support Vector Machine (SVM), Random Forest (RF), and autoencoder. They also examined the various CNN input formats and commonly used feature extractors. Moreover, Zhou et al. (Zhou et al. 2023) reviewed the DL literature from 2010 to early 2023 addressing AD, MCI, and associated disorders. They discussed the main categories of supervised, semi-supervised, and unsupervised techniques designed for a variety of applications in this domain, including the most current advancements like the application of generative models, RNNs, and graph-neural networks.

The existing literature on the application of DL in AD is missing a thorough overview of all the essential aspects of DL applications, and despite some coverage of DL advancements, the literature may lack up-to-date insights into the latest techniques and their application to AD diagnosis. Although previous research offers insightful information, it sometimes lacks comprehensive analyses, comparative assessments, and security strategy concerns for DL applications. By conducting a comprehensive literature review of more than 50 peer-reviewed studies, classifying DL applications into various purposes in AD, and offering a more in-depth analysis of neglected factors like the frequency of CNN techniques and security features in AD–DL applications, the current study addresses this gap in expertise. The main thing that makes our work unique is its thorough approach, which helps us understand how DL can be used in AD more deeply and broadly. In this way, our review is different from other comparison studies because it looks at the uses of DL in AD research in more detail and covers more topics. Unlike other reviews, we have carefully divided DL uses into separate areas, such as NLP, AD recognition, classification, and drug reuse. By putting them into such specific groups, we can better see the unique roles that DL plays in different areas of AD study. Additionally, our SLR goes beyond merely listing techniques; it critically evaluates the challenges, advantages, methodologies, and areas needing improvement within DL–AD research. We have also provided a thorough analysis of the current state of research, identifying key trends such as the predominant use of CNN and the frequent application of Python in simulation environments. Furthermore, the review highlights the importance of accuracy and explainability in DL models, with an emphasis on the usage of Grad-CAM as the leading explainable AI framework. Importantly, we address significant gaps in the literature, such as the lack of security considerations in most studies, and propose future research directions to address these gaps. By integrating these comprehensive insights, our review not only summarizes existing knowledge but also sets the stage for future advancements, making it a valuable resource for researchers and healthcare professionals alike. Therefore, we confront the topic differently than previous works. Table 2 contains a list of relevant studies.

3 Related concepts and background

The relevant survey articles were thoroughly evaluated in an assortment of related fields in the previously indicated section. In this section, we go through several terms and concepts to help our readers understand the subject better. DL has emerged as a transformative force in healthcare, revolutionizing various aspects of patient care, diagnosis, treatment, and medical research. Leveraging advanced neural network architectures, DL algorithms

Table 2 The side-by-side comparison of the related reviews

Authors	Main idea	Scope	Advantage	Disadvantage
Sellappan et al. (2024)	Introducing state-of-the-art DL methods for classifying AD patients	DL approaches for AD diagnosis	– An extensive assessment of the methods	– The selection process for articles that are included is not made apparent
Shanmugavadivel et al. (2023)	A thorough investigation into multiple AI approaches and neuroimaging modalities for AD detection methods	AD detection using neuroimaging methods and AI techniques	– Highlighting potential challenges and providing additional recommendations for enhancement	– A comparative analysis of the papers has not been carried out
Warren and Moustafa (2023)	Exploring DL techniques and fMRI for AD diagnosis	AD detection using fMRI and DL methods	– Illustrating the gaps in subsequent decision-making processes – The limitations and substitute options are examined	– No comparison is drawn between the articles
Yao et al. (2023)	Presenting a review of advances in DL techniques for the categorization of different AD stages	DL methods for AD classification	– The upcoming projects are thoroughly outlined – The limitations and substitute options are examined	– A comparative analysis of the papers has not been carried out
Pallawi and Singh (2023)	Investigating deep CCNs to Classify AD	DNN models for AD classification	– An extensive assessment of the methods	– There is not a thorough and organized evaluation of DL's applications
Zhao et al. (2023)	Evaluating some common conventional ML methods employed for categorizing and predicting AD	ML methods for AD classification and prediction	– A comparative analysis is conducted between the papers – Highlighting potential challenges and providing additional recommendations for enhancement	– The selection process for articles that are included is not made apparent
Zhou et al. (2023)	Acknowledging DL techniques designed for different tasks within the realm of AD	Survey of DL for AD	– The upcoming projects are thoroughly outlined – Illustrating the gaps in subsequent decision-making processes	– No comparison is drawn between the articles
Our work	Exploring DL approaches for Alzheimer's diagnosis	DL applications in Alzheimer's	– Providing a clear categorization of interpretability methods – Offering new insights on XAI in cancer diagnosis – Evaluating existing strategies and identifying areas for improvement	– Primarily emphasizes XAI in cancer detection, potentially overlooking broader applications in other medical areas

excel in tasks such as medical image analysis, disease diagnosis, drug discovery, and personalized medicine. These algorithms can analyze vast amounts of medical data, including Electronic Health Records (EHRs), genomic information, and medical images, to extract valuable insights and assist healthcare professionals in making informed decisions (Yang et al. 2024). DL enables precision medicine approaches that improve patient outcomes and minimize adverse effects by accurately identifying anomalies in medical images, predicting disease progression, and designing personalized treatment plans based on individual patient data. Using smart tech, distant tracking systems, and telemedicine platforms, DL

also makes healthcare easier and more efficient. These technologies constantly gather and analyze bodily data, which makes it easier to find health problems early and get help right away, especially in areas that are hard to reach or do not have enough medical professionals (Sanjay and Swarnalatha 2023). DL improves healthcare operations and management by simplifying processes, resource sharing, and clerical chores. This lowers costs and makes the patient experience better. DL also helps medical research by looking at huge amounts of omics data, finding new biomarkers, and adding to scientific knowledge. This leads to big steps forward in understanding complicated diseases and creating focused treatments. Overall, DL's entry into healthcare has the potential to change the field, improve patient care, and solve global healthcare problems (Yi et al. 2023).

3.1 What is ML?

ML is an important part of AI that focuses on creating algorithms and statistical models that allow computers to use data to make guesses and choices without having to be explicitly programmed to do each job. In ML, computers are trained on named or unstructured information to find trends and draw conclusions (Kadri et al. 2021). There are three main types of ML algorithms: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, algorithms are trained on datasets that have already been labeled. This lets them connect input data to output labels by reducing loss functions such as mean squared error or cross-entropy loss (Ahmadzadeh et al. 2023). On the other hand, unsupervised learning task algorithms use methods like k-means grouping and Principal Component Analysis (PCA) to find unseen structures or trends in datasets that have not been named (Hu et al. 2021). In reinforcement learning, on the other hand, agents interact with their environments and learn how to make choices by repeatedly making mistakes and trying to maximize cumulative rewards by choosing actions that lead to good outcomes and taking feedback from the environment into account (Amini et al. 2021). In Table 3, you can see an outline of ML methods.

Table 3 Definitions of critical terminologies in ML

Terminology	Definition
Supervised learning	Algorithms are trained on labeled data, learning the relationship between inputs and corresponding outputs for prediction tasks
Unsupervised learning	Algorithms find patterns and structures in unlabeled data without explicit guidance and are commonly used for clustering and dimensionality reduction tasks
Semi-supervised learning	Utilizes both labeled and unlabeled data for training, combining aspects of supervised and unsupervised learning
Reinforcement learning	Involves an agent learning to make decisions through interaction with an environment and receiving feedback in the form of rewards or penalties
Deep learning	A subset of ML focusing on neural networks with many layers is capable of learning intricate patterns in data and is particularly successful in tasks like image recognition and NLP

3.2 What is DL?

DL is a subset of ML that uses artificial neural networks with multiple layers (deep architectures) to learn hierarchical representations of data (Kadri et al. 2021). DL algorithms, also known as Deep Neural Networks (DNNs), are capable of automatically learning features and patterns directly from raw data, without the need for manual feature engineering (Fathi et al. 2023). DL models are characterized by their ability to learn complex representations of data through the composition of multiple nonlinear transformations (Uyguroğlu et al. 2024). DL architectures typically consist of an input layer, one or more hidden layers, and an output layer. Each layer comprises multiple interconnected nodes (neurons), and the connections between nodes are associated with learnable parameters (weights and biases) (Jia et al. 2024). DL models are trained using algorithms like Stochastic Gradient Descent (SGD) and backpropagation, which adjust the model parameters to minimize a loss function that measures the difference between predicted outputs and ground truth labels (Vidhya et al. 2023). DL has shown remarkable success in various domains, including computer vision, NLP, speech recognition, and reinforcement learning (Janghel and Rathore 2021). CNNs are widely used in computer vision tasks, RNNs are commonly used in sequential data processing tasks like natural language understanding, and transformer architectures have become popular for tasks requiring attention mechanisms, such as machine translation and text generation (Bae et al. 2023). A selection of DL methods is shown in Table 4. Moreover, Fig. 1 illustrates an overview of ML methods.

3.3 What is the role of DL in AD?

DL algorithms are pivotal in the diagnosis of AD by leveraging diverse data sources such as medical imaging, genetic information, and clinical assessments (Kumar and Nandhini 2021). CNNs analyze neuroimaging data to detect structural changes indicative of AD, while genetic data analysis identifies biomarkers and genetic risk factors (Janghel and Rathore 2021). NLP techniques interpret clinical notes and cognitive tests, aiding clinicians in making accurate diagnoses. Multi-modal fusion techniques integrate data from various sources, enhancing diagnostic accuracy (Alamro et al. 2023). These advancements enable early detection, personalized treatment planning, and disease progression monitoring, underscoring DL's critical role in improving AD diagnosis and patient outcomes while advancing our understanding of this complex neurodegenerative condition. Figure 2 illustrates AD diagnosis using DL algorithms.

3.4 What is XAI?

To clarify how predetermined rules impacted decision-making, researchers proposed Explainable AI (XAI) in the mid-1990s, which gave rise to the idea of interpreting AI systems. In 2004, the term "XAI" was used to describe a system that could explain the fundamental functions of objects that were simulated by computers. This concept contrasts with the present AI approaches' lack of transparency, which makes it difficult for engineers and researchers to explain conclusions made by AI. The development of XAI was delayed, even though several research studies achieved notable advancements. AI algorithms have achieved significant progress in prediction over the years, dominating the significance of

Table 4 Definitions of key terminologies in DL

Terminology	Definition
CNNs	DL models are primarily used for visual tasks such as image recognition and object detection. They consist of convolutional layers that automatically learn features from input images
RNNs	The layout of neural networks is made to handle sequential data by keeping an internal state. They are used a lot in NLP, speech recognition, and analyzing time series
LSTM networks	A type of RNN architecture capable of learning long-term dependencies in sequential data. LSTMs use gated units to remember and forget information over time, making them effective in tasks requiring memory over extended sequences
GRUs	GRUs are a simpler version of RNNs that are meant to handle long-range relationships in linear data, similar to LSTMs. They are better at using computers because they need fewer factors and calculations than LSTMs
Transformer models	A cutting-edge structure for NLP tasks that uses self-attention processes to figure out how things in a series are related to each other. Robots called transformers are very good at things like translating languages and making up texts
GANs	There is a generator and a discriminator network that are taught against each other in generative models. Because GANs can make realistic data samples, they can be used to make images and add to existing data
Attention mechanisms	Techniques used in neural networks to selectively focus on relevant parts of input data. Attention mechanisms allow models to dynamically allocate attention, enhancing performance in tasks involving sequential or spatial data by prioritizing important information

AI's capacity to explain its decisions. However, because AI/ML technologies are becoming widely integrated into daily life, the term "XAI" has acquired acceptance among academics and practitioners (Gao et al. 2023a). To meet social, ethical, and regulatory challenges, AI must now enter a new phase that demands transparency in its internal operations and the ability for users to understand the reasoning behind its conclusions. Various hypotheses have been thrown in recognition of the necessity for a comprehensive description of XAI. However, there is not a single, universally agreed-upon meaning for the phrase because it is frequently associated with endeavors, adjustments, and processes aimed at creating transparent AI and addressing trust concerns. Some explanations supported by reliable organizations can serve as examples of the accepted definition of XAI (Bapat et al. 2023).

3.5 What is the significance of XAI?

XAI has significant advantages and impacts and is essential to the constant development of AI. The ability of XAI to manage legal and ethical challenges associated with AI systems is one of its main advantages. AI's results are made clear in XAI, which promotes openness

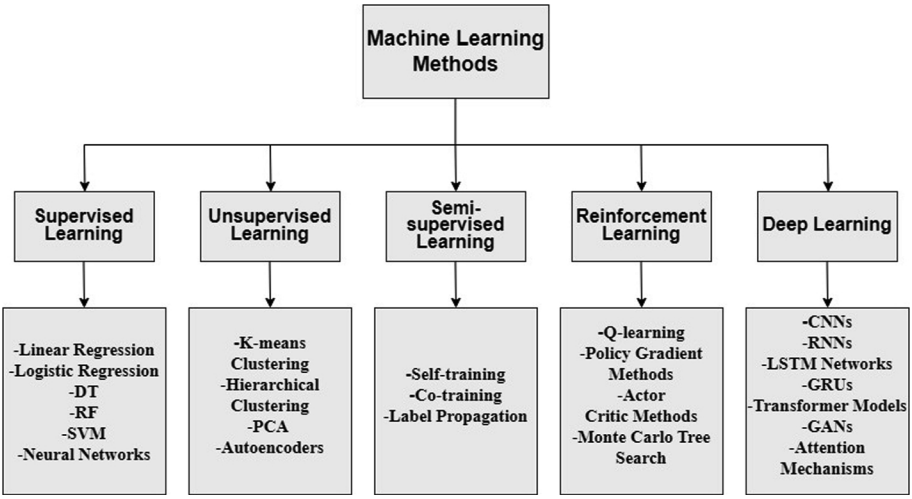

Fig. 1 An overview of ML/DL methods

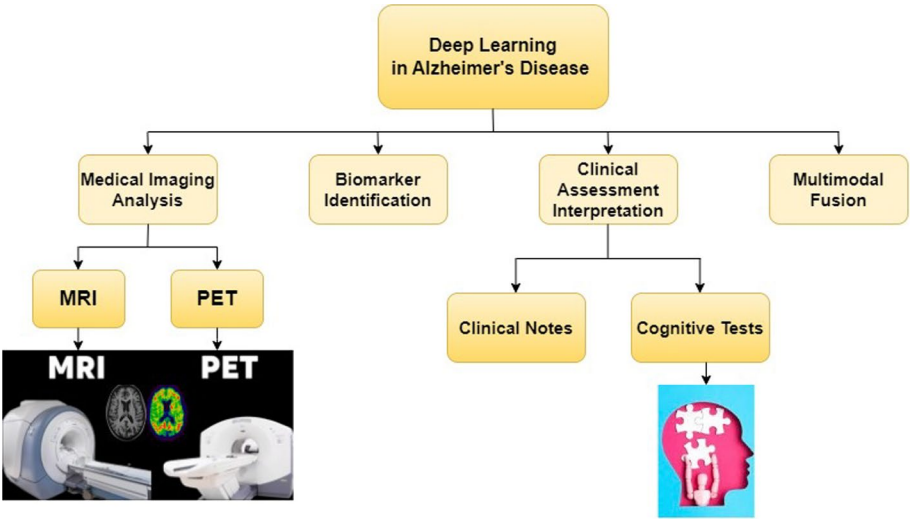

Fig. 2 AD diagnosis using DL algorithms

and reliability. This makes it less likely for prejudice, bias, and unfair results to happen. For AI to be widely accepted and used, people need to believe in it. To do this, you need to be open about it. It is also the law in hospitals and banks that they do it. Scientists can learn more with XAI because it makes it easier to understand how AI models work. This lets scientists find connections between data and how models work, which could lead to new scientific findings and better models (Khare and Acharya 2023). A second way that XAI helps people and AI work together is by letting them talk to each other. This lets people learn from AI, make better decisions, and work with AI to answer tough problems more quickly (Varghese et al. 2023). When people work together and XAI can come up with new ideas,

AI apps are better used. This helps in a lot of areas, like school and health care (Zhang et al. 2021). Finally, XAI is more than just being open; it is the foundation for smart AI study and use. This means AI systems are more like people, which makes them better for society and less likely to cause problems. AI is becoming more and more important in every part of our lives (Santi et al. 2023). XAI is a key part of making AI systems smarter, more responsible, reliable, and useful by closing the gap between what humans can understand and how AI makes decisions (Santi et al. 2023). Popular frameworks in the field of XAI are presented in Table 5.

4 Research methodology

We reviewed terminology and background material in detail in the previous section. To understand the DL–AD domain, this section uses the SLR method, which is an extensive review of all research carried out regarding a specific topic. We then analyze how DL approaches are applied in AD. After that, we evaluate the impact of the methods used to select the studies. The subsections offer more details about the search strategy, such as the criteria for choosing the studies and subjects of research.

4.1 Formalization of questions

The primary goals of the survey include identifying, categorizing, examining, and evaluating key articles in the field of DL applications for AD. Using an SLR, strategies for achieving the previously described goals can be examined for their various components

Table 5 Definitions of key terminologies in XAI

Terminology	Definition
SHAP	A method in Explainable AI that assigns importance values to features based on cooperative game theory, offering insight into how each feature contributes to the model's predictions
LIME	A technique in XAI that explains individual predictions of black-box ML models by creating interpretable surrogate models around specific instances, aiding understanding of model decisions
Grad-CAM	A method in computer vision to visualize and interpret CNN decisions, generating heatmaps to highlight important image regions by analyzing gradients of the predicted class score for feature maps
Saliency map	A visualization tool used in XAI and computer vision to highlight the most important parts of an input, such as an image, that contribute to a model's output prediction, aiding in understanding the model's decision-making process
Integrated gradients	An interpretability technique in ML is used to explain the predictions of a model by computing the integral of the gradients of the model's output concerning its inputs along a path from a baseline or reference state to the actual input, providing insights into feature importance and contributions to predictions

and aspects. We would also need further information about all of the significant issues and challenges that our field is now facing. A list of some of the designated Research Questions (RQs) is as follows:

- ***RQ1:*** What kinds of classifications are accessible for the DL methods used in AD fields? Which of the preceding serves as a particular example?

  The solution to this issue is provided in Sect. 5.

- ***RQ 2:*** Which approaches are used by the researchers to conduct their studies?

  Section 5 *provides an answer to this query.*

- ***RQ 3:*** What conclusions may be made from the data collected from the publications that have been examined? Which articles' features received the greatest attention? What are the most often used DL applications for AD?

  The answer to this question is provided in Sect. 6.

- ***RQ 4***: What are the potential opportunities in this field?

  Section 7 *provides an answer to this question.*

4.2 The article selection procedure

A two-phase process was used to find and select the publications for this research. The selection procedure is shown in Fig. 3. Table 6 illustrates the keywords and phrases utilized in the first phase of the paper search. Items in this collection were discovered using an electronic database search. Journals, conference papers, book chapters, notes, technical studies, and special issues have also been uncovered. The first stage yielded 3042 items.

The total number of publications to be investigated in Stage 2 was determined using two separate methods: (1) studies published before 2017 were eliminated, and (2) duplicated articles were deleted. 1486 papers remaining. The paper titles and abstracts were then examined. The methodology, evaluation, discussion, and conclusion of the publications have all been double-checked to ensure their relevance to the research. In this step, three criteria were utilized to choose articles: (1) review papers were excluded, (2) papers that did not meet our requirements were deleted, and (3) papers that did not mention DL approaches

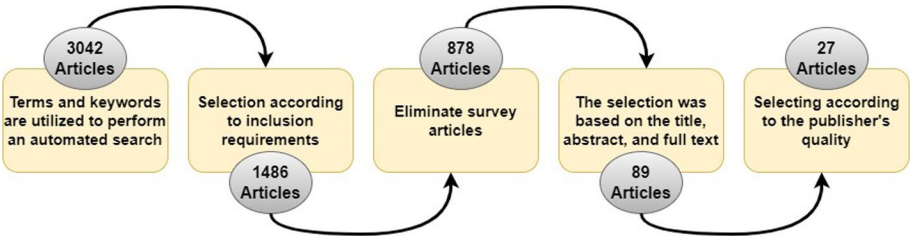

Fig. 3 The steps involved in finding and choosing articles

Table 6 Search phrases and keywords

S#	Search terms and keywords
S1	'DL' and 'Alzheimer's Disease'
S2	'DL' and 'Alzheimer's Disease'
S3	'Deep Learning' and 'Alzheimer's Disease'
S4	'DL' and 'Alzheimer's Disease Detection'
S5	'DL' and 'Alzheimer's Disease Classification'
S6	'DL' and 'Alzheimer's Disease' and 'Drug Discovery'
S7	'DL' and 'Alzheimer's Disease' and 'Drug Repurposing'
S8	'DL' and 'Alzheimer's Disease' and 'NLP'
S9	'DL' and 'Alzheimer's Disease' and 'Natural Language Processing'

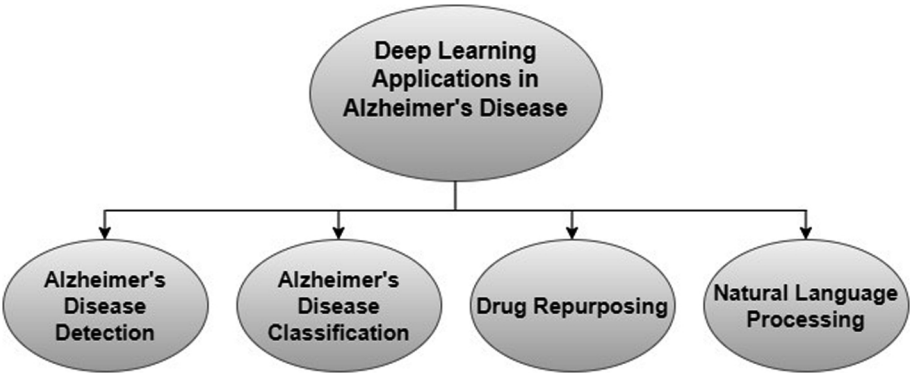

Fig. 4 Making the suggested taxonomy graphical: listing out four different methods for using deep learning in Alzheimer's disease study

were excluded. 89 Articles have currently been selected for further review because they met the stringent requirements. Then, based on the publisher's quality, 28 papers were selected for evaluation and study. IEEE, Elsevier, Wiley, and Springer published 77.8% of the articles that were selected. The ACM and MDPI have the lowest scores. Furthermore, the majority of the chosen publications were published between 2023 and 2024. The next section examines and discusses the DL-AD applications in depth.

5 DL applications for Alzheimer's disease

In the preceding section, we presented an overview of the research methodology and statistics. DL applications in AD are explored in comprehensive detail in this section. The categories that we utilized for arranging DL applications in the field of AD were AD detection, AD classification, medication repurposing, and NLP. Figure 4 illustrates the classification of DL approaches in AD.

5.1 Alzheimer's disease detection applications

Identifying AD has posed significant challenges in the medical industry for an extended period. A timely and accurate diagnosis is crucial for the management and treatment of the condition. DL technology has significantly advanced in recent years, and these developments may aid in the identification of AD. DL algorithms, particularly neural networks, have demonstrated the capability to analyze MRI and PET scan data to identify intricate patterns and characteristics by training these algorithms on extensive datasets of brain images (Kwak et al. 2023). This enables researchers to develop models capable of detecting subtle mutations and problems in the brain indicative of the progression of AD. This novel approach may expedite the identification and treatment of diseases, benefiting patients and enhancing our understanding of their mechanisms. Incorporating DL into clinical practice might significantly transform the identification and management of early AD, Due to the ongoing development of research in this field. Pradhan et al. (Pradhan et al. 2023) used the EHR and Single Nucleotide Polymorphisms (SNP) dataset to explore and learn about several research approaches for the early identification of AD. These datasets are entered into several DNN frameworks. The MRI scans have been analyzed using CNN. The model's accuracy, precision, and recall F1-score have been determined and compared using several types of DL frameworks, including CNN, DenseNet, ResNet50, and VGG, following the development of the confusion matrix. SNP datasets have been proven to have the best accuracy rate, providing clarity on the identification of AD. Also, an early AD detection method was established by Rajasree and Brintha Rajakumari (Rajasree and Brintha Rajakumari 2023). Optimal Bi-GRU (Gated Recurrent Unit), trained by MLP and QDNN results, is used to conclude. Remarkably, the weight of the QDNN model is modified utilizing the recently announced Enhanced Math Optimizer Accelerated Arithmetic Optimization (EMOAOA) approach to increase the detection accuracy of the network. In particular, 95% detection accuracy at the 60th learning rate, 95.5% at the 70th learning rate, 98% at the 80th learning rate, and 98.7% at the 90th learning rate was attained using the suggested EMOAOA+Ensemble Classifier (EC). In addition, Qiao et al. (Qiao et al. 2024) suggested a DNN termed Convolutional-Squeeze-Excitation-Softmax-NET (CSES-NET) in conjunction with multi-channel feature fusion for the diagnosis of AD. They used 1539 participants from the ADNI database for their research. The outcomes of the experiment confirmed that by combining the multi-channel characteristics and extracting nonlinear deep features, the suggested approach enhanced the model's efficacy. Some risk SNP loci that were linked to the pathology of AD and assisted with early prevention and management of AD were also found by the genome-wide association analysis. Moreover, for the identification of discriminative brain regions that could be the biomarkers for early detection in AD progression, Pan et al. (Pan et al. 2022) proposed an adaptive interpretable ensemble model that combines deep 3D CNN and Genetic Algorithm (GA), i.e., 3D CNN+EL+GA. Based on the datasets from the AD Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS), the testing results demonstrated that 3D CNN+EL+GA performed better than other cutting-edge DL algorithms.

Furthermore, Bloch et al. (Bloch and Friedrich 2024) presented a technique to thoroughly examine the traditional ML and DL explanations for early AD diagnosis based on 3D MRIs. A hold-out ADNI test set was used to evaluate seven classical ML models: radial SVM, Decision Tree (DT), Logistic Regression (LR), Extreme Gradient Boosting (XGBoost), RF,

and polynomial SVM. These models were then externally verified for Australian Imaging, Biomarker, and Lifestyle Flagship Study of Ageing (AIBL) and OASIS. Three approaches [permutation importance, Local Interpretable Model-Agnostic Explanations (LIME), and Shapley Additive exPlanations (SHAP)] were used to explain the conventional ML models. Four DL explanations of six DL models [Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM++, LIME, and SHAP] were compared to the outcome explanations. Models for DL and ML achieved the same classification results. Also, using MRI and image augmentation, Alqahtani et al. (Alqahtani et al. 2023a) created DL models, including AlexNet, ResNet-50, GoogleNet (InceptionV3), and SqueezeNet, that are focused on the early diagnosis of AD by Transfer Learning (TL). The average accuracy for AlexNet was 98.05%, the average accuracy for GoogleNet (InceptionV3) was 97.80%, and the average accuracy for ResNet-50 was 91.11%. One of the main problems associated with employing DL is that there is often insufficient data to train a network from the beginning. This is where the TL technique becomes useful.

Singular Value Decomposition (SVD) and PCA are two gene selection methods that Abdelwahab et al. (Abdelwahab et al. 2023) used to identify relevant genes in microarray datasets. They used a CNN as their classifier for AD prediction by utilizing DL concepts. Their methodology comprised processing the dataset using a seven-layer CNN in various configurations. The PCA–CNN model's efficacy was demonstrated by empirical results on the AD dataset, which produced an accuracy of 96.60% and a loss of 0.3503. Furthermore, the SVD–CNN model demonstrated excellent accuracy, achieving 97.08% with a 0.2466 loss. Finally, using DL Techniques, Sorour et al. (Nanthini et al. 2024) proposed an AD–DL technique for early AD detection. The collection is made up of brain MRI images that are used to test and confirm the proposed model. There are five DL models shown that can do binary classification and independent feature extraction. The models are separated into two categories: (1) CNN-without-Aug and (2) CNNs-with-Aug, CNNs-Long Short-Term Memory (LSTM)-with-Aug, CNNs-SVM-with-Aug, and Transfer learning using VGG16-SVM-with-Aug. The results of the tests show that CNN-LSTM is better, with an accuracy rate of 99.92%.

DL methods for AD identification have benefits, such as high accuracy because they use a variety of models and data types. DL makes it possible to use new optimization methods, such as weight tuning and higher recognition rates. DL also makes feature fusion work well and gives clear results that help with diagnosis. DL methods may have trouble with data processing and growth, though, because they are so involved. Not having enough computer power or training data can also be a problem for DL models. Finally, it is still very hard to make sure that private medical data used in DL systems stays safe and private. DL–AD can be used to find people in several ways, as shown in Table 7. One good thing about DL methods for finding AD is that they are very accurate because they use a lot of different models and types of data. To make things better with DL, you can do things like tune the weights and get better recognition rates. Additionally, DL makes feature fusion work well and gives clear outcomes that assist in analysis. Nonetheless, because of their complexity, DL algorithms may encounter difficulties in data analysis and scalability. These models require substantial training data and significant computational power, which can pose challenges in resource-constrained environments. Ultimately, safeguarding the privacy and security of confidential medical data employed in DL systems remains a significant concern in the identification of AD. Table 7 illustrates several applications and characteristics of DL-AD identification.

5.2 Alzheimer's disease classification applications

The field of neuroscience and medical research that investigates AD is highly complex and always evolving. AD is a prevalent form of dementia, making accurate classification essential for research, assessment, and intervention. Increasingly, DL techniques such as CNNs and RNNs are employed to accurately diagnose the various phases of AD by examining intricate medical data, including brain scans and clinical notes. Mahim et al. (Mahim et al. 2024) propose a system that uses both a GRU and a Vision Transformer (ViT) to autonomously and accurately detect indicators of AD in MRI images. The proposed model outperforms prior techniques in both accuracy and speed. It also resolves the problem of class imbalance in the MRI image dataset. The preprocessed Alzheimer's MRI dataset from Kaggle was utilized to train the model, achieving an accuracy of 99.53% for four-class classification and 99.69% for binary classification. Additionally, it proved to have a high accuracy of 99.26% for 3-class on the ADNI Baseline Database. A robust 10-fold cross-validation procedure confirmed their model. In addition, their model became interpretable and explicable by the implementation of XAI approaches.

The pre-trained model that can correctly identify the patient's stage was proposed by Ravi et al. (Srividhya et al. 2023) after using 26 pre-trained Keras DL models to predict the categorization of AD MRI data. ResNet-50v2 has been shown to provide the greatest accuracy (91.84%) and F1-score (0.97) for the AD class. Visualization methods like saliency map and Grad-CAM are applied to the model that best explains the image's region of focus, facilitating the prediction of the image's class. Furthermore, to classify AD into three groups (non-demented, moderate demented, and very mild demented), Saleh et al. (Saleh et al. 2023) suggested utilizing the DenseNet architecture. In comparison to previous methods, their methodology performs better on the MRI dataset and makes use of transfer learning architectures as an underlying model. Their model outperformed earlier approaches by demonstrating, after a thorough examination, that their suggested system model achieves an outstanding Area Under Curve (AUC) of 99% and an accuracy of 96.5%. According to Sekhar et al. (Sekhar et al. 2023) they mixed the Marine Predators Algorithm (MPA) with deep GANs to come up with a new way to use EEG to classify AD. The way they suggest doing things uses GANs' ability to pull out important patterns from raw EEG data to make fake EEG samples. To see what effect the suggested method had, they used a public EEG collection of AD cases and healthy volunteers. The results for classification accuracy, sensitivity, and precision show that deep GANs with MPA work better than the most advanced methods.

Dhaygude et al. (Dhaygude et al. 2024) employed a sophisticated three-dimensional CNN to demonstrate an end-to-end classification technique for AD. It encompasses two additional tasks: the regression of the Mini-Mental State Examination (MMSE) score and the regression of the Clinical Dementia Rating (CDR) Scale score, aimed at enhancing the outcomes of AD categorization and facilitating simultaneous learning. Additionally, an attention mechanism is introduced to improve feature extraction capabilities. According to experimental results using the AD Neuroimaging Initiative dataset, the authors' suggested strategy for classifying AD, performed better than existing techniques in terms of improved classification accuracy and other metrics. Also, a Pathology-Steered Stratification Network (PSSN) was suggested by Xu et al. (Xu et al. 2022a) that used a reaction–diffusion model to include domain knowledge that has been established in AD pathology. Their biological

Table 7 The methods, features, and properties of the DL–AD detection applications

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Pradhan et al. (2023)	Developing DL methods for AD diagnosis	– Accuracy of 99.2%	– Low interpretability	Python	– CNN	–	AD detection	Yes	– EHR – SNP	No
Rajasree and Brintha Rajakumari (2023)	Constructing a four-phase method for early AD identification	– High accuracy	– High complexity – High energy usage	Python	– GRU – MLP – DNN	–	AD detection	Yes	ADNI	No
Qiao et al. (2024)	Providing the CSES-NET-based multi-channel feature fusion approach for early AD detection	– High scalability	– Low robustness	MATLAB	-DNN	–	AD detection	Yes	ADNI	No
Pan et al. (2022)	Proposing an interpretable ensemble model based on an adaptive 3DCNN architecture	– High scalability – High interpretability	– High complexity – High energy usage – High delay	Python	– CNN	Grad-CAM	AD detection	Yes	ADNI, OASIS	Yes
Bloch et al. (2024)	Presenting an organized comparison of the DL and standard ML theories using 3D MRIs	– High robustness – High scalability	– High complexity – High energy usage – High delay	Scikit-learn, PyTorch	– LR – DT – XGBoost –LightGBM – SVM	– Grad-CAM – LIME – SHAP	AD detection	Yes	– ADNI – AIBL – OASIS	No
Alqahtani et al. (2023a)	Suggesting a strategy for DL models focusing on AD early detection	– High accuracy – High scalability	– Low interpretability – Low robustness	Python, MATLAB	AlexNet, ResNet-50, GoogleNet	–	AD detection	Yes	Kaggle	No

Table 7 (continued)

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Abdelwahab et al. (2023)	Creating an accurate predictive DL-based model to assist in AD early diagnosis	– High accuracy – High interpretability	– High complexity – High energy usage – High delay	Python	CNN	–	AD detection	Yes	GSE63060 and GSE63061	No
Sorour et al. (2024b)	Proposing an AD–DL method that employs DL Techniques for early AD detection	Accuracy: 99.92% Precision: 100.00% Recall: 99.50% F1-score: 99.70%	– Low interpretability	Keras, Tensorflow, Scikit-learn	– LSTM – SVM	–	AD classification	Yes	Kaggle	No

Table 8 The methods, features, and properties of the DL–AD classification applications

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Mahim et al. (2024)	Introducing a method for efficient AD classification based on XAI methods and a ViT–GRU model	– High accuracy – High interpretability	– High complexity – High energy usage – High delay	– Python	– ViT – GRU	– LIME – SHAP	AD classification	Yes	– Kaggle – ADNI	No
Srividhya et al. (2023)	Using DL techniques to classify AD into multiple classes	– High interpretability	– Low robustness – Low stability	Python	ResNet-50v2	– Grad-CAM – Saliency Map	AD classification	Yes	ADNI	No
Saleh et al. (2023)	Providing a three-stage AD classification scheme using DL	– Accuracy of 96.5% – AUC of 99%	– Prolonged computational time	TensorFlow	DenseNet	–	AD classification	Yes	Kaggle	No
Sekhar et al. (2023)	Proposing an EEG-based approach to AD classification	– Accuracy of 99.96%	– High complexity	MATLAB	– MPA – GAN	–	AD classification	Yes	– European data format (EDF) – DAT	No
Dhaygude et al. (2024)	Presenting a deep 3D CNN-based end-to-end AD classification technique	– High accuracy	– Low robustness – Low reliability	MATLAB, TensorFlow	CNN	–	AD classification	Yes	ADNI	No
Xu et al. (2022a)	Employing ML and systems biology modeling to identify AD spectrum subgroups	– High robustness	– Low scalability – Low interpretability	TensorFlow	DNN	–	AD classification	No	ADNI	No

Table 8 (continued)

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Bohn et al. (2023)	Exploring multimodal fragility variables that distinguish four cohorts in the AD spectrum	– High robustness	– Low interpretability	– Scikit-learn	RF	– Tree SHAP	AD classification	Yes	– COMPASS-ND	Yes
Shukla et al. (2023)	Suggesting pre-processing techniques to enhance MRI image classification accuracy for AD	– Accuracy: 97.57% – Sensitivity: 97.60%	– Low interpretability	Scikit-learn	– RF – XGBoost – CNN	–	AD classification	Yes	ADNI	No

model filled in the gaps in the sparse imaging data by predicting long-term evolution trajectories that represent each distinctive progression pattern. In terms of intra- and inter-cluster homogeneity and inter-cluster heterogeneity of different clinical ratings, their stratification performs better. Additionally, they discovered six subgroups within the AD spectrum, each of which has a unique biomarker pattern that corresponds with its clinical result.

Furthermore, Bohn et al. (Bohn et al. 2023) investigated a large and representative collection of up to 75 frailty-related parameters that are often utilized in geriatric research. To separate a cohort of older persons who are Cognitively Unimpaired (CU) from those who have Subjective Cognitive Impairment (SCI), MCI, or AD, they employed three computationally competing ML models to find important indicators. They simultaneously tested the relative significance of 83 multi-modal frailty indicators in differentiating the groups utilizing RF analysis. To fully understand the implications of prediction, they used Tree SHAP. With an AUC of 0.98, the AD model revealed exceptional prediction accuracy. Additionally, brain MRI image classification ability has been enhanced by the innovative pre-processing techniques suggested by Shukla et al. (Shukla et al. 2023). They also decreased the time needed for the model to be trained using different pre-existing learning models. After obtaining a dataset from ADNI, it was formatted from 4 to 2D. XGBoost, CNN, and RF are the three learning algorithms they suggested for AD classification. The dataset was utilized to calculate results, which indicate that the accuracy and sensitivity of 97.57% and 97.60%, respectively, surpassed the existing research.

AD classification with DL methods offers high accuracy and automated feature extraction from diverse data sources. DL can reveal novel biomarkers and provide insights into disease progression. Challenges include the need for extensive labeled data, which may be costly and scarce in medical contexts. DL models may also be susceptible to biases in the data and lack transparency, complicating validation and clinical acceptance. Balancing these advantages and disadvantages is crucial for effective DL utilization in Alzheimer's classification. Table 8 lists all of the DL–AD classification applications and their characteristics.

5.3 Drug repurposing applications

Currently, there is no treatment for AD, presenting a significant challenge in contemporary medicine. Experts are exploring innovative solutions to address this significant issue. A promising concept is to utilize existing drugs for the treatment of AD. Researchers can expedite the drug discovery process and maybe identify efficient treatments more rapidly by using medications that have previously been approved for other conditions. DL approaches are utilized to analyze vast datasets and get insights into the intricate mechanisms behind AD (Wu et al. 2022). Researchers want to identify novel therapeutic targets and develop more targeted and effective treatments for AD by integrating DL algorithms with medication repurposing strategies. This interdisciplinary approach has significant potential in combating this severe neurological condition. Zhao et al. (Zhao et al. 2024) developed a Multi-Perspective Neural Network Ensemble Learning (MNNEL-DTA) approach for forecasting drug-target compatibility. This approach employs several DL models, each concentrating on a distinct set of characteristics. Their technique outperformed the latest models by 9.2% and 8.9%, respectively. They employed MNNEL-DTA to identify novel applications for existing medications utilized in the treatment of AD. Research indicated that memantine functions as both an S1R and SYK agonist, whereas glutathione acts as an inhibitor for the

target medin. Notably, memantine has Food and Drug Administration (FDA) approval for AD. Meanwhile, early research on the use of glutathione, which has traditionally been recommended to treat insomnia, has shown promising in treating AD by specifically targeting the main protein.

To find AD risk genes and potential treatment targets, Xu et al. (Xu et al. 2022b) designed an interpretable DL framework entitled NETTAG. This framework takes advantage of data from human protein–protein interactome networks, multi-modal genomics, and Genome-Wide Association Studies (GWAS). The lower incidence of AD associated with gemfibrozil has been demonstrated by combining human interactome network-based prediction with EHR-based patient data validation. Also, a multi-task DL pipeline was built by Hsieh et al. (Hsieh et al. 2023) that first learns AD risk genes and biological connections, and then uses multi-level information on treatment effectiveness to find repurposable therapeutic candidates. Their study conducted mechanistic validation of the top-ranked candidates in neuronal cells. The aim was to discover medication combinations that were both safe and effective in decreasing oxidative stress while maintaining the survival and morphology of neurons. Their pipeline demonstrated how integrating complementary and heterogeneous data/knowledge—such as transcriptome patterns, human interactome, experimental effectiveness, and real-world patient data—may assist in clarifying the process of developing drugs for complicated diseases. Furthermore, the drug repositioning approach utilizing Optimal Transport for AD (DOTA) is a DL technique created by Chyr et al. (Chyr et al. 2022) to repurpose FDA-approved medications that are helpful for AD. By using this method, they forecasted that antipsychotic medications having circadian effects—such as trazadone, olanzapine, suvorexant, aripiprazole, risperidone, and quetiapine—will be effective in treating AD patients. These medications target critical brain receptors, such as serotonin 5-HT2A, dopamine D2, and orexin receptors, which are involved in memory, learning, and cognition.

Moreover, to predict Drug–Target Interactions (DTIs) and AD medication repositioning, Wang et al. (2022a) presented KG-DTI, a knowledge graph-based DL technique. It is shown that, in 5-fold cross-validation, KG-DTI obtains an average ACC of 88.0%, F1-Score of 87.7%, AUROC of 94.3%, and AUPR of 95%. In practice, medications are repositioned to AD by Apolipoprotein E (APOE) target using KG-DTI. Consequently, it is discovered that seven of the ten most highly suggested medications have already been utilized in clinical settings or have been linked to beneficial research for AD. Additionally, to find new applications for licensed treatments for AD, Zang et al. (Zang et al. 2023) simulated trials for hundreds of medications from two major real-world data repositories. They proposed a model selection technique for enhanced baseline covariate balancing and evaluated several propensity score models within the context of inverse probability of treatment weighting. Additionally, they discovered that when it came to covariate balancing, the DL-based propensity score model did not always perform better than logistic regression-based techniques. Finally, they identified five highly rated medications that were initially meant for different uses but may be beneficial for Alzheimer's patients: pantoprazole, gabapentin, atorvastatin, fluticasone, and omeprazole.

AD drug repurposing through DL presents efficiency and cost-effectiveness by analyzing extensive datasets and uncovering potential candidates from existing medications. However, challenges include limited data availability for training models and the risk of oversimplifying the complex biological mechanisms of the disease, potentially leading to inaccurate predictions. Repurposed pharmaceuticals may encounter legal challenges due to

their intended usage differing from their original application, necessitating additional validation and approval procedures. Table 9 includes the various applications of DL–AD drugs and illustrates their respective benefits.

5.4 NLP applications

Recent years have demonstrated significant potential for research at the intersection of NLP and DL addressing AD. Researchers can employ sophisticated ML techniques to analyze huge amounts of written material, such as medical records, clinical notes, and academic publications, to derive critical insights into the diagnosis, progression, and treatment of AD. NLP algorithms can extract valuable information from unstructured text, enabling the identification of significant genetic variables, biomarkers, and clinical indicators associated with the condition (Wu et al. 2023). Researchers want to accelerate Alzheimer's research through the application of NLP and DL. This is anticipated to result in earlier diagnoses, improved patient care, and the development of more efficacious treatment methodologies. Adhikari et al. (Adhikari et al. 2022) extracted recordings from individuals with AD and a control group of healthy individuals to create a dataset in Nepali, a low-resource language. They developed a method to extract information from textual data and established a system utilizing NLP to identify AD patients at an early stage using Nepali transcripts. The researchers determined that patients' speech narratives while describing an image effectively reflect their difficulties in comprehending information.

Additionally, a DL method called the Imaging Genetic deep neural network system (IGnet) was created by Wang et al. (2022b) to automatically classify AD with a combination of genetic sequencing and MRI data. Their suggested method combines computer vision and NLP techniques. A 3D CNN handles the three-dimensional MRI input, while the genetic sequence input is handled by a Transformer encoder. The ADNI dataset was exposed to the suggested methodology. They obtained an area under the receiver operating characteristic curve (AUC-ROC) of 0.924 with the test set and an accuracy of 83.78% in classification using baseline MRI images and specific single-nucleotide polymorphisms on chromosome 19. Also, to assess the early risk of AD based on the image description test, Roshanzamir et al. (Roshanzamir and Aghajan 2021) created transformer-based DL models based on NLP. The Pitt corpus, which includes data from 170 AD patients with 257 interviews and 99 healthy controls with 243 interviews, is used to evaluate the models using picture description test transcripts. Their method is improved by 2.48% using the Large Bidirectional Encoder Representations from Transformers (BERTLarge) embedding with the LR classifier, achieving 88.08% classification accuracy.

Also, Shen et al. (Shen et al. 2022) classified the lifestyle status in AD from clinical notes using six different BERT models: Bert base, Pubmed-BERT (Abs+Ft), Pubmed-BERT (Abs), Bio BERT, UMLS BERT, and Bio-clinical BERT. Weak supervision was utilized along with three traditional ML models: logistic regression, RF, and SVM. Their UMLS BERT model, with precision, recall, and F1 scores of 0.93, 0.93, and 0.92, respectively, performed the best when it came to categorizing the status of physical activity. The bio-clinical BERT model performed the best when it came to categorizing an excessive diet, with precision, recall, and F1 scores of 0.93, 0.93, and 0.93, respectively. Moreover, Clarke et al. (Clarke et al. 2021) have demonstrated that computational analysis of linked speech using NLP and ML indicates disease and might be used as a quick, scalable test for early

Table 9 The methods, features, and properties of the DL–AD drug repurposing applications

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Zhao et al. (2024)	Developing a method for medication repurposing and drug-target affinity prediction in AD	– High accuracy – High Robustness	– High complexity – High energy usage – High delay	Python	– CNN – GNN – LSTM	–	Drug repurposing	Yes	– Davis – KIBA	No
Xu et al. (2022b)	Introducing NETTAG for gene analysis and therapy targets	– High Robustness – High interpretability	– Low scalability	– Scikit-learn	NETTAG	–	Drug repurposing	Yes	NMEDW	No
Hsieh et al. (2023)	Identifying AD medications with the GNN technique	– High accuracy	– High complexity – High energy usage	PyTorch	GNN	PcGrad	Drug repurposing	Yes	CTD STRING AlzPED	Yes
Chyr et al. (2022)	Establishing a DL technique to repurpose FDA-approved medicines that work well for AD	– High accuracy	– Low reliability – Low robustness	Python	DOTA	–	Drug repurposing	Yes	– DrugBank – PharmGKB, the Therapeutic Target Database	No
Wang et al. (2022a)	Providing KG-DTI, for AD medication repositioning	– Accuracy: 88.0% – F1-score: 87.7% – AUROC: 94.3% – AUPR: 95%	– High complexity – High energy usage	Python	KG-DTI	–	Drug repurposing	Yes	DrugBank	No
Zang et al. (2023)	Discovering the applications of medicines that have been approved for AD	– High scalability – High robustness	– High training time	PyTorch, Scikit-learn	– LR – GBM – MLP – LSTM	–	Drug repurposing	Yes	– OneFlorida database – IBM MarketScan database	No

detection. 50 volunteers were enlisted, 25 of whom were Healthy Controls (HC) and 25 of whom were mildly AD or MCI (AD+MCI). They completed five related speech tests. Each transcript had an autonomously extracted high-dimensional set of linguistic characteristics that were used to train SVM for group classification. Performance differed; utilizing picture book narration, accuracy for HC vs. AD+MCI categorization ranged from 62 to 78% while applying overlearned narrative characteristics.

AD detection employing NLP and DL methods presents notable advantages but also presents challenges. NLP enables efficient analysis of textual data from medical records, aiding in early detection by extracting relevant information. DL models integrated with NLP can handle multiple data types, providing a holistic understanding of the disease. Automated screening facilitated by NLP algorithms enhances scalability and cost-effectiveness in early detection initiatives. However, challenges such as data quality and standardization, ethical considerations regarding patient data privacy, and the interpretability of DL models pose significant hurdles. Additionally, resource-intensive model training and generalization challenges across diverse patient demographics limit accessibility and scalability. Validation and regulatory approval processes for NLP and DL-based diagnostic tools demand rigorous studies, delaying market adoption. Despite these challenges, the potential for early intervention and research advancements in AD pathology through NLP and DL methods remains promising. Table 10 shows the methodologies, characteristics, and attributes of AD-related DL and NLP applications.

6 Result and discussion

Previous sections included evaluations, analyses, and classification of 27 works into 4 categories. This section includes a discussion of the findings and a comprehensive examination of the research area, addressing topics like technical TOOLKITS, aspects of design, reliability, perceptive accuracy, computational structure and interpretability, ethical considerations, security, data visualizations, and results analysis.

6.1 Comprehensive overview of datasets for DL applications in AD

Datasets play a crucial role in advancing DL applications in AD research by providing the necessary data for training, validating, and testing ML and DL models. ADNI is a groundbreaking research effort launched in 2004, bringing together researchers from academia and industry to study AD through the use of neuroimaging techniques, biomarkers, and clinical data. ADNI was established as a public–private partnership with funding from the National Institutes of Health (NIH), pharmaceutical companies, and nonprofit organizations (Gao et al. 2023b). Its main goal is to speed up the creation of new treatments for AD by finding signs of disease progression, figuring out how the disease works, and making clinical studies easier by standardizing how data is collected and shared. People give a lot of different kinds of information when they take part in ADNI. As an example, they give genetic information, results from physical checks, MRI and PET scans, and results from cognitive tests. People with AD, MCI, and healthy other people make up ADNI. Experts can see how the disease changes over time with continuous data. Many places in North America keep

Table 10 The methods, features, and properties of the DL and NLP applications for AD

Authors	Main idea	Advantages	Research challenge	Simulation setting	DL methods	XAI methods	Application	Evaluation	Dataset	Security?
Adhikari et al. (2022)	Proposing NLP-based AD early identification	– High accuracy – Low delay	– Low interpretability	Python	– CNN – LSTM	–	NLP	Yes	DementiaBank	No
Wang et al. (2022b)	Integrating 3D CNN for AD classification	– Accuracy of 83.78%	– Low scalability	Python	CNN		NLP	Yes	ADNI	No
Roshan-zamir and Aghajan (2021)	Producing transformer-based models for early AD risk assessment based on NLP	– Accuracy of 88.08% – F1 scores of 87.23%	– Low interpretability – Low robustness	Python	CNN	–	NLP	Yes	DementiaBank	No
Shen et al. (2022)	Assessing NLP models for AD patient lifestyle categorization	– Precision: 93% – Recall: 93%	– Lack of data diversity	Python	– LR – RF – SVM	–	NLP	Yes	CDR of the University of Minnesota	No
Clarke et al. (2021)	Exploring NLP and ML for early AD and MCI detection in speech	– High scalability	– Small study population – Low reliability	Python	SVM	–	NLP	Yes	St. George's NHS Cognitive Clinic	No

records of people with AD so that researchers can get a good sample of the whole community (Alqahtani et al. 2023b).

For the project, different kinds of images are used. Different types of MRI are used to look at different parts of the brain. Functional MRI is used to see how the brain works, and PET is used to find biomarkers like amyloid and tau proteins. Researchers can see and measure changes in the brain that are linked to AD. These changes include the brain shrinking, amyloid plaques building up, and changes in how processes are linked to each other. ADNI also gets full clinical exams from the people who take part in the studies. These include questions about their medical history, how well they can think and remember things, and the things they do every day. Because ADNI wants scientists to work together and share data, its files can be used by anyone in the scientific community. Scientists from all over the world can use ADNI to do new studies, find new biomarkers, and try models that can predict AD because it gives them access to standard, high-quality data (Balasundaram et al. 2023). We now know a lot more about AD thanks to this effort, which helped us find imaging and biochemical signs that show how the disease is getting worse. It has also made it easier to come up with diagnostic standards, plans for clinical trials, and methods for making new drugs for AD. Overall, ADNI is very important for moving forward research on AD because it provides large datasets, encourages researchers to work together, and speeds up the process of turning scientific findings into medical uses. DL methods used on ADNI data could reveal new information about how the disease works, make diagnoses more accurate, and make it easier to customize treatment plans for AD (Zhang et al. 2024).

Kaggle is also a key site in the fields of data science and machine learning, and its wide range of datasets and shared settings make it a useful tool for AD study in many ways. Kaggle has many different types of AD datasets, such as scan data, hospital records, genetic information, and more. There are a lot of tools in this group that will help experts learn more about how complicated the sickness is. The tool has two uses: it stores useful information and brings people together to work on projects together. Researchers, data scientists, and Alzheimer's fans can find, study, and look over data about the disease on Kaggle. People can share their knowledge, work on projects together, and make progress in their field through games, talks, and exchanging ideas on Kaggle (Allada et al. 2024). The setting in Kaggle can be used for a lot of different things. DL is an important part of how we work with data these days. DL helps researchers with their studies of AD by helping them figure out how to predict the disease and how to sort and look at images. Kaggle datasets can be used to train and test DL models. Then, they can look at other scientists' findings and come up with new ways to understand and fight the disease. Researchers can also benefit from Kaggle's busy community, which gives lessons, kernels (notes that can be run), boards, and chances to learn. With Kaggle, experts can push the limits of AD research by sharing their knowledge and working together. This has helped us find the sickness, treat it better, and learn more about it. To make AD research better, Kaggle is very important because it lets academics use different datasets, motivates them to work together, and gives them resources and tools for DL. Getting people to work together and share their ideas on Kaggle helps the ongoing battle against AD and makes things better for patients (Seriramulu et al. 2024).

DrugBank is an extensive database that is very helpful for academics and pharmaceutical businesses that are working on finding new drugs and making new versions of old ones. It has a lot of detailed information on thousands of drugs, both approved drugs and substances that are still being tested. The library has a lot of information about different parts of drugs,

like their effects, combinations, pharmacokinetics, and pharmacodynamics. Researchers can learn more about how drugs work at the molecular level and how they impact living things with this much information. DrugBank is a central library that has all the information about drugs. Its goal is to make it easy to find drugs and do studies on them. Investigators, pharmacologists, doctors, and people who make drugs can learn a lot about the properties, ways that drugs work, and safety factors of drugs here. For example, the library can be used to change drugs, find new targets, do pharmacogenomics, and keep an eye on drugs. Putting together different kinds of drug-related info on one site, DrugBank makes study easier and speeds up the process of making new drugs (Tsuji et al. 2021).

It also gives structural information about drugs and the things they target, such as 2D and 3D images, molecular docking models, and binding site studies (Amini et al. 2024). It may also contain clinical data such as applications, dosages, delivery techniques, adverse effects, and recommendations or contraindications for the medicine. Researchers and data scientists utilize DrugBank data to construct DL models applicable to several aspects of medicine and drug discovery. DL approaches are employed to predict drug-target interactions, identify novel drug candidates, analyze drug response patterns, and optimize drug design. By employing sophisticated machine learning methodologies and integrating DrugBank data with additional scientific information, researchers may identify concealed relationships among drugs, their targets, diseases, and biological mechanisms. This facilitates the development of more personalized medications and treatments for individuals (Shi et al. 2024).

OASIS is a substantial repository of brain datasets mostly comprising MRI images from healthy individuals, those with AD, and individuals with MCI. Cross-sectional and longitudinal data from the cohort illustrate the alterations in the brain's morphology with time. The structural MRI data in OASIS is very elucidative and reveals significant insights into the brain's organization. These scans are especially T1-weighted studies. There are also personal pieces of information like age, gender, and amount of learning, as well as clinical pieces like MMSE scores, CDR scores, and other memory tests. The main goal of OASIS is to make it easy for experts to study AD and how the brain ages by giving them free access to scanning data. By letting anyone use these files for free, OASIS hopes to speed up research and discovery in the areas of brain and AD. Scientists can use OASIS data to study different aspects of AD pathology. They can use it to find biomarkers linked to the start and progression of the disease and to come up with new ways to identify and treat it. A lot of the time, DL methods are used to look at OASIS MRI data. Researchers can use these techniques to get detailed patterns and features from brain pictures, like changes in neuroanatomy, structural problems, and the way brain areas are connected in space (Yu et al. 2023). DL models learned from OASIS data can be used for many things, like sorting the different stages of AD into groups, predicting how the disease will get worse, finding biomarkers linked to the disease, and making computer systems that help doctors make diagnoses. Using OASIS data along with advanced DL methods, researchers can learn more about how AD works and maybe help make changes and medicines that work better. Figure 5 shows the frequency of the dataset that was used for DL–AD apps.

Looking at the numbers of datasets shows that there are many different types of DL used for AD study. ADNI holds a substantial 40% market share, far surpassing its competitors. The prominence of ADNI underscores its significance as a fundamental dataset. It encompasses a diverse array of data types essential for training and evaluating DL models aimed at diagnosing, predicting, and comprehending AD. Kaggle secured second place with a 16%

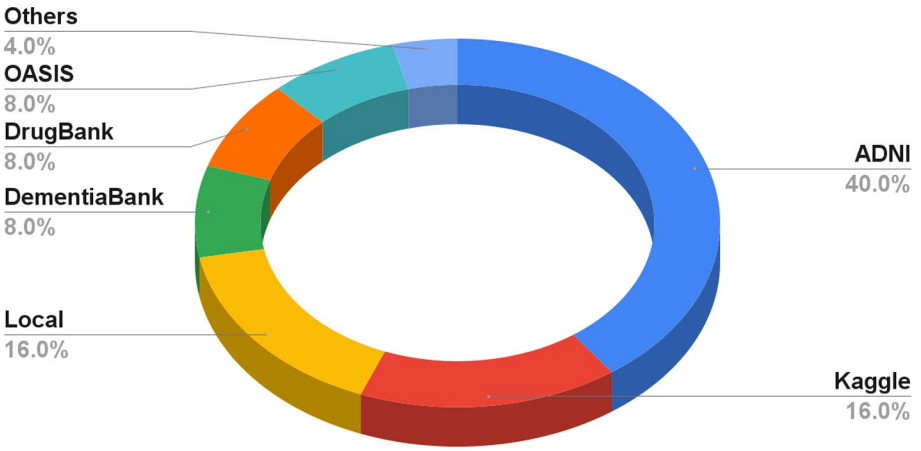

Fig. 5 The frequency of dataset utilized for DL–AD applications

share, underscoring its significance as a supplementary resource for students by providing an extensive array of datasets and challenges that foster innovation in AD. Furthermore, local datasets, including 16%, illustrate the significance of structured data to many study groups. These datasets are presumably designed to address certain research inquiries or objectives, so enhancing the diversity and profundity of data available for exploration by DL. Additionally, platforms like as DementiaBank and OASIS, each possessing an 8% stake, underscore the significance of specialized datasets for various domains of Alzheimer's research. DementiaBank includes audio and video recordings, whereas OASIS contains brain data. Although they constitute a minor portion of the datasets, these datasets are crucial for DL applications, particularly in natural language processing and image analysis. Furthermore, drug-related datasets such as DrugBank, which received an 8% allocation, indicate a growing interest in employing DL to discover novel pharmaceuticals and repurpose existing ones for Alzheimer's treatment. The observation that diverse datasets such as EHR, AIBL, and AlzPED achieved 4% underscores the significance of utilizing a variety of data sources in DL applications. All of these datasets work together to make DL success in AD possible. They make it easier for people from different fields to work together and come up with new ways to diagnose and treat this common neurological disease.

6.1.1 Data augmentation techniques in deep learning for Alzheimer's disease research

For DL models to be trained correctly, they need large datasets to keep them from overfitting and to make sure they work well. Neuroimaging and clinical data collection can be hard to do on a big scale, which can make it hard to get large datasets for AD studies. Data enrichment methods offer an answer by purposely making datasets bigger and more varied, which leads to better model generalization. Some common ways to improve a picture are to rotate, translate, scale, flip, add noise, and stretch the image. Rotation involves turning pictures in small amounts to add variety without changing the content too much. This is useful when brain direction changes a little from one scan to the next. Translation moves pictures along the x or y line to make it look like they were moving slightly while they were being captured. This makes the system more resistant to changes in position. When you zoom in or

out during scaling, you teach models how to spot features at different sizes. This is useful when brain sizes or image resolutions are different. Flipping pictures horizontally or vertically makes mirror images. This can be useful for symmetrical structures, but it needs to be used carefully to avoid making brain tissue look too real. Adding noise to pictures makes them more resistant to errors caused by noisy data. Elastic deformation uses small, random changes to mimic real biological differences. This makes it easier for the model to work for many different people. Also, methods for making fake data, like GANs and VAEs, are used to make new data samples that look a lot like real data. This is done to greatly increase datasets when real data is tested and treatments for this widespread neurological disease.

The best methods for adding to data depend on the type of study being done. Synthetic data generation with GANs and VAEs can make a lot of realistic training data when there is not a lot of real data available. Simple methods like rotation, translation, and scaling can quickly add to small datasets. When working with different imaging sources, adjusting the contrast and brightness makes pictures from those sources more similar, which lowers the unpredictability. Elastic deformation also makes it easier for the model to work with data from different imaging tools. Noise addition improves model performance in real-world settings where noise is common, and simulated data output gives you more data with managed noise levels for situations where you need to be resistant to noise. When studying certain parts of the brain, cropping, and padding make sure that the model learns traits that are unique to those areas, and translation and rotation make sure that the model does not change shape when its position changes slightly. These customized reinforcement methods are very important for getting around the problems that come up with small datasets in DL for AD study. They make sure that strong and useful models are built.

Researchers deal with the problem of limited data access in DL models for AD diagnosis in several different ways. A common method is to add more training samples to existing data by changing it in ways like rotating, translating, and scaling it. This is called "data augmentation". This approach increases the range of the training set and makes the model more useful in real life. Transfer learning is another way. This includes fine-tuning models that have already been trained on large, broad datasets so that they can work with smaller, more specific datasets that are linked to AD. This makes better use of the data from the larger sample to make the job related to AD more effective. Also, methods for making manufactured data, like GANs, are used to provide real-looking fake data that can be used to supplement the limited real data. Federated learning and other collaborative projects let different schools train models using their local data without sharing private data. This makes the dataset bigger while protecting privacy. Researchers are using multi-modal techniques more often to improve training and make up for the fact that some types of data are not always available. These techniques combine different types of data, like genetic, clinical, and brain data. By combining these methods, researchers may be able to get around the problems caused by small datasets and make DL models for AD detection that are more accurate and durable.

6.1.2 Multi-modal biomarker strategies in AD detection

When different types of biomarkers are used together, like genetic information, neuroimaging, Cerebrospinal Fluid (CSF) biomarkers, and cognitive tests, it is much easier to make a correct and reliable diagnosis of AD. Using these various types of data, researchers might be able to take the best parts of each biomarker group to get a more complete and correct

picture of the disease. Brain imaging measures, like MRI and PET tests, can correctly show how the structure and function of the brain change in people with AD. Amyloid-beta and tau diseases show up at the cell level, though, according to signs in the CSF. We can better understand how AD is passed down through families when we know about risk genes like APOE μ4. Cognitive tests can also show how much these changes in the body affect the ability to remember things and think clearly. Advanced data fusion methods can be used to combine information from different sources into accurate prediction models that take into account all the difficult parts of AD. This makes it easier to see how the disease is getting worse, get a faster and more accurate diagnosis, and find the best way to treat it for each person. This all-around method avoids the issues that come up when you rely on just one type of biomarker. Overall, this makes it easier and more accurate to find and treat AD.

6.2 Security

Considering security is crucial when employing DL for AD detection to protect sensitive medical data and ensure the accuracy and impartiality of diagnostic and prediction models. To begin, it is very important to take steps to protect data safety and security to keep patient information safe. Medical data, X-rays, and DNA information are all part of this. Model security is also important so that no one can view or change the model without permission. To do this, methods are used such as model watermarking and encryption of model data. Another important thing is to protect against threats from bad people because inputs that are changed on purpose could cause wrong diagnoses or wrong predictions. When you use antagonistic training and strong model designs on DL models, they are less likely to be attacked in these ways. This makes them more useful for studying AD and using them in clinical settings. Also, mistakes in collecting data and making model predictions are big issues that could make health gaps even worse. Consequently, it is important to prevent biases by employing diverse and representative training datasets and routinely assessing the model's impartiality among ethnic groups. Cybersecurity is increasingly vital as it protects the stability and availability of systems connected to Active Directory applications (Roncero-Parra et al. 2024). This involves protecting medical apparatus, digital platforms, and communication networks from unauthorized access and cybercriminals. In the end, following the rules and standards for medical device security and data protection is what makes sure that DL technologies are used legally and ethically in AD research and clinical settings. This fosters trust and enhances healthcare outcomes for all individuals. To effectively address these security concerns and promote the proper utilization of DL in AD management, collaboration among all stakeholders is essential. This includes researchers, physicians, legislators, and cybersecurity specialists. Figure 6 illustrates the frequency of protection utilization for DL–AD applications.

In DL applications for identifying AD, security considerations were neglected 88.9% of the time, whereas they were addressed in 11.1% of cases. This shows a big gap in how these applications are secure. This difference brings up the possible flaws and risks that come with using DL models in AD studies and clinical practice. To begin, the vast majority of times when security was not taken into account makes me worry about how to keep private medical data safe. Personal data like medical records, X-rays, and genetic information are involved in AD and can be stolen if they are not properly protected. Not paying attention to safety measures like data protection, access controls, and anonymization methods raises the

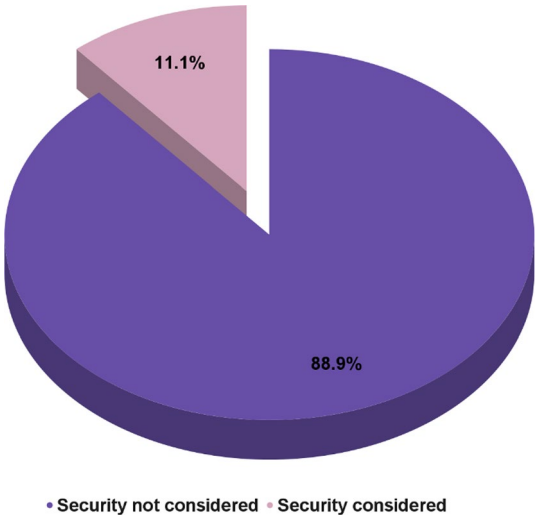

Fig. 6 The frequency of security utilized for DL–AD applications

chance of data theft, privacy breaches, and people getting into your data without permission. This has important moral and legal effects, mostly when it comes to protecting patient privacy and following rules like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA).

Also, the fact that security is not considered much in DL apps for AD makes me wonder about the dependability and strength of diagnostic and predictive models. A lot of things can go wrong with model predictions, including attacks from enemies, mistakes in gathering data, and threats that can be found online. In particular, warring strikes could lead to wrong evaluations or forecasts, which would be bad for patients and make health gaps worse. Strong cybersecurity measures are needed to protect healthcare services and make sure they are available and working properly. This is because cyber threats are aimed at systems that are linked together, such as medical devices and communication networks. Now that we know these things, security must become very important in DL's AD systems right away. Cybersecurity experts, researchers, doctors, and politicians must all work together to fix the flaws that are already there, put in place strong security measures, and promote responsible rollout practices. Users can protect patient data, make sure models are accurate, and work for fair healthcare outcomes in AD management by building security in every stage of development and operation.

For safety reasons, the AD models used in the study must be kept safe. This is because they hold private data and correct recognition is very important. An all-around plan can keep these models safe. This includes keeping the data safe, making sure the models work as they should, and making sure they can keep going even if something goes wrong. Homomorphic encryption and other types of data encryption keep data private and safe while it is being handled and stop people from getting to it without permission. You can also use federated learning to train models with samples that are spread out. A lot of schools can train models this way without having to share raw data. This keeps the patients' identities safe. To keep the model's identity safe, antagonistic training can be used to make it better at resisting attacks from bad people, like giving it bad inputs to trick it. Regular checks and the use of blockchain technology could make the data and model updates safer by making sure

that only approved changes are made. Any possible security holes can be found and fixed quickly by seriously restricting access, checking the security often, and keeping a close eye on what the model does at all times. It might be better to keep privacy, integrity, and availability high by adding these safety features to the different DL models that are used in AD studies. In the long run, this could keep patient data safe and make analysis tools more accurate.

6.3 Analysis of results

To learn more about the tools and methods that experts in the DL–AD field like to use, we can look at how different modeling settings are used in this field. The numbers give us a good idea of what the study group thinks and how they feel about things. Figure 7 clearly shows that Python is the clear winner, with a comfortable majority share of 53.1%. Python's popularity is due to several things, such as its flexibility as a computer language, its large library of tools designed for data analysis and machine learning, and its easy-to-understand structure that speeds up development. Additionally, the broad use of DL tools like TensorFlow and PyTorch, which work perfectly with Python, makes it even more popular among experts. Nonetheless, MATLAB appears to be lagging behind Python, although maintains a 12.5% market share, which remains significant. Historically, MATLAB was highly regarded in scientific and technical disciplines because of its robust analytical tools and extensive library. Nonetheless, it appears that Python and other more adaptable and community-backed alternatives are gaining traction.

A 9.4% allocation to Scikit-Learn indicates that conventional ML techniques remain relevant in the DL for the AD detection domain; nonetheless, there is a discernible transition toward DL methodologies. This alteration is occurring because DL models are more adept at managing complex data patterns and extracting valuable insights from large datasets. PyTorch and TensorFlow have comparable market shares of 9.4% and 12.5%, respectively. This same score indicates that each of these premier DL techniques has comparable

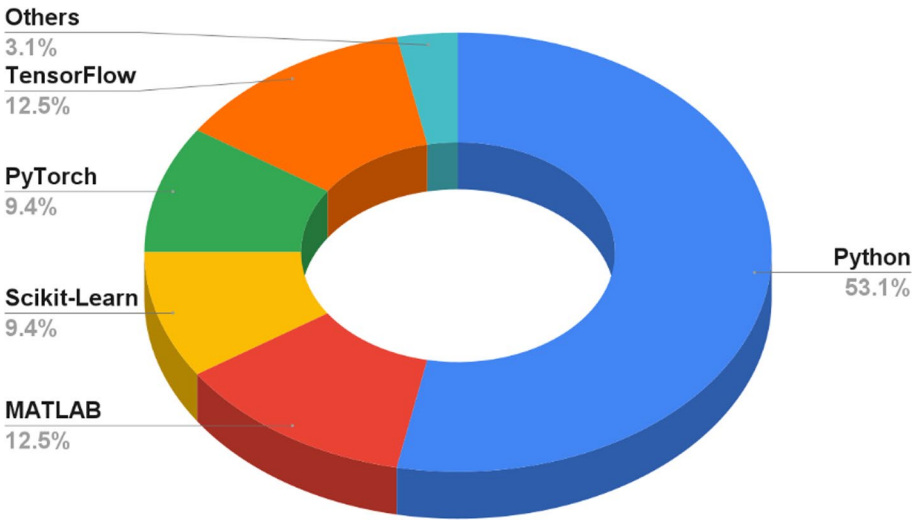

Fig. 7 The utilization patterns of different simulation configurations in the DL–AD domain

popularity. Researchers likely select between PyTorch and TensorFlow according to the requirements of their research. For instance, PyTorch may be employed when dynamic computation graphs are required, but TensorFlow may be utilized when a comprehensive ecosystem with robust industry backing is desired. The little allocation (3.1%) to alternative packages and tools, such as Keras and pandas, indicates their utility, but less frequent usage compared to the previously mentioned resources. Keras, frequently used with TensorFlow for constructing neural networks, demonstrates TensorFlow's potency in the domain of DL and AD detection. Ultimately, Python emerged as the premier programming language for DL applications in AD research, with TensorFlow and PyTorch recognized as the foremost DL frameworks. MATLAB remains available; however, its usage has diminished due to the emergence of more versatile and community-supported alternatives such as Python. Various researchers employ distinct tools and packages according to their requirements; nonetheless, the prevailing tendency indicates that Python-based environments are the most favored in this domain.

A lot more ML and DL methods are being used in AD research, and different models are being used to study different parts of the disease. CNN is the most popular of these methods, with 25.7% of all uses. This shows how well they work for processing complex spatial data like brain imaging scans. CNNs have shown promise in getting useful information from brain data, which can help find clues and trends in how diseases develop. 11.4% of the methods used are SVM and LSTM. SVMs are known for being able to handle large amounts of data and have been used successfully in classification tasks, such as telling the difference between the different stages of AD based on different neuroimaging and clinical traits. LSTM networks, a type of RNN, on the other hand, are great at handling sequential data. This makes them useful for looking at trends in time in continuous studies or following how a disease gets worse over time. RF, LR, and DNN together make up 8.6% of the methods used. RF algorithms are good at working with different kinds of data and have been used in Alzheimer's studies to help choose features and put them into groups. Classical statistical methods like LR are still useful because they are easy to understand and use. They are often used in studies that want to find out how risk factors affect disease results. With their many secret layers, DNNs are a strong way to find complicated trends in different kinds of data, such as DNA, imaging, and clinical data. This helps move precision medicine methods in Alzheimer's research forward.

Each MLP, GNN, and boosting method like XGBoost adds 5.7% to the range of DL uses in AD. MLPs are a basic type of feedforward neural networks. They are flexible and have been used for many tasks, from figuring out what an image is about to detecting cognitive decline. GNNs have gotten a lot of attention because they can model the complicated links that exist in brain connection networks. This assists us in understanding how dysfunctions in neural networks contribute to AD. Certain boosting algorithms, such as XGBoost, provide ensemble learning techniques that combine many weak learners to enhance classification precision. This enhances the reliability of detection models for Alzheimer's research. Ultimately, alternative methodologies such as GAN, decision trees, GBM, Light-GBM, ResNet, and DenseNet provide an additional 2.9% to the domain of DL in Alzheimer's research. There are several approaches to exploring innovative strategies for disease identification, biomarker discovery, and comprehension of disease mechanisms. In conclusion, the diverse array of ML and DL methodologies employed in AD research underscores the complexity and various aspects of the condition. Researchers can investigate various

aspects of Alzheimer's pathogenesis by leveraging the unique capabilities of each approach. This enables early identification and diagnosis, understanding of disease progression, and identifying potential therapeutic targets. Integrating these methodologies and optimizing their synergistic effects will advance Alzheimer's research, ultimately resulting in improved screening instruments, more tailored treatment strategies, and enhanced outcomes for those afflicted with this devastating condition. Figure 8 illustrates several ML and DL strategies employed in the domain of DL for the identification of AD. The percentages displayed provide significant insights into the common use of different explainability and interpretability frameworks in the field of DL applications for AD research. Analysis of Fig. 9 indicates that Grad-CAM is the most prevalent, accounting for 45.5% of the utilization. This approach is highly beneficial for visualizing and comprehending DL models since it highlights the most significant regions of brain images for tasks such as prediction or classification. Researchers are keen to identify disease-related features in three-dimensional space to further our understanding of AD pathogenesis. This is evidenced by the extensive utilization of Grad-CAM. Trailing closely, SHAP constitutes 27.3% of traffic, indicating a significant growth rate. SHAP enables the determination of the rationale behind certain assertions by providing the significance values of attributes. When it comes to AD, SHAP helps figure out which signs or traits are most important for diagnosing or tracking the disease. This high usage rate shows that more and more people are realizing how important it is to understand how different features work together in DL models. This makes AI-driven predictions easier to understand and more open. With 18.2% usage, LIME is another important part of the interpretability environment in the AD study. By changing the data that it is given, LIME comes

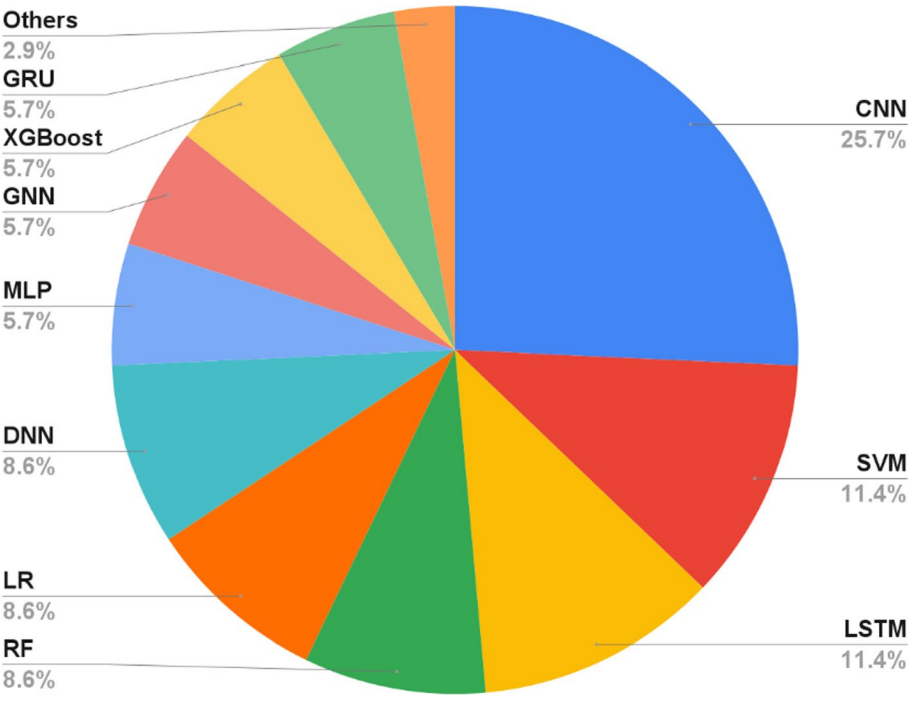

Fig. 8 Implementation of various ML–DL methods and how frequently they appear in DL–AD medical publications

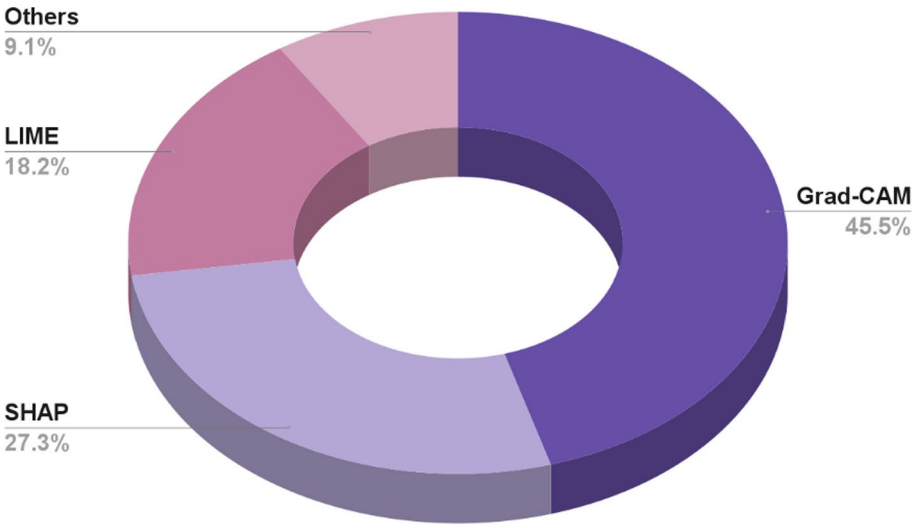

Fig. 9 Utilization of diverse frameworks for explainability and interpretability within the DL–AD domain

up with reasons for each estimate. In this case, LIME is a useful tool for figuring out why certain brain pictures are thought to show signs of the disease. Even though it is not used as much as Grad-CAM and SHAP, the fact that LIME exists shows how important model interpretability is in this particular field.

Other methods, like saliency maps, make up the last 9.1%. These methods help find interesting parts of brain images so that more research can be done on them by showing where the model focuses its attention in the input data. The fact that these methods are included, even though they are only used a small portion of the time, shows the wide range of ways that have been studied in AD research to understand how DL models make decisions. In analysis, the fact that Grad-CAM is the most popular method suggests that the main goal is to locate disease-related features in space. On the other hand, the high acceptance rates of SHAP and LIME show that the focus is shifting to understanding the importance of specific features or biomarkers. Additionally, the addition of other methods shows the multidimensional approach experts in AD use to improve the practical usefulness of AI systems for detection and treatment. Researchers are trying to close the gap between AI-driven forecasts and clinical decision-making by using these interpretability methods together. This will help us understand and treat AD better in the long run. All of these results show how important it is for data to be able to be understood and explained for AD study to move forward.

It is common for DL to be used for AD, as shown by the large number of studies that look at it (92.6%). This shows that the field cares a lot about being careful and correct. This focus on review shows that people understand how complicated and variable AD is. However, strong evaluation methods are needed to make sure that results are accurate and can be used by others. A lot of tests need to be done on DL models to show that they can be used in many cases, including with various datasets, groups of people, and stages of disease. A lot of different tests are done on models to see how well they work and how well they can be used in real clinical settings. This testing process not only makes people more sure that the models are correct, but it also makes sure that they can help doctors identify, predict, and

treat people with AD. A lot of people review things, which also makes it easy to see how the field has grown over time. Through systematic comparison of several models using standard assessment metrics, researchers may identify successful aspects, areas for improvement, and promising avenues for further investigation. As this process of enhancement is reiterated, more sophisticated and beneficial DL approaches are developed specifically to address the challenges posed by AD.

Nonetheless, it is crucial to discuss the limited research that disregards their methodologies. If outcomes are not examined, individuals may begin to question their accuracy and reliability, perhaps hindering advancement in the discipline. Researchers must recognize the significance of stringent review processes and include them in their methodologies to ensure the validity and effect of their investigations. Standardized review procedures are essential for elucidating growth and comparing outcomes across various investigations. Establishing standardized review metrics, standards, and datasets in the area facilitates collaboration, enhances uniformity in effort, and simplifies replication. Adhering to standard review standards enables researchers to conduct more effective comparisons, expedite information sharing, and ultimately enhance the use of DL in AD. Collaborative efforts are underway to enhance patient outcomes since review mechanisms are prevalent in several DL applications for AD. Researchers can enhance DL models by concentrating on comprehensive evaluation techniques. This will ultimately result in the development of superior instruments for detecting and treating AD. Figure 10 provides a comprehensive analysis of the frequency of rate components in DL applications for AD.

The several forms of DL applications utilized in the AD study demonstrate diverse methodologies for comprehending, diagnosing, and treating this severe neurological condition. AD detection and classification, each with eight applications, are evident domains requiring further focus. To try to find things, DL models look at medical imaging data such as MRI scans, PET scans, and CT scans. These models help find early warning signs and patterns of AD, which is important for getting help and treatment right away. In classification, on the other hand, DNA markers and clinical signs are used to put people into different stages or groups of the disease. DL methods make it easier to make complicated classification models that can put people in the right groups based on huge amounts of complex data. This helps

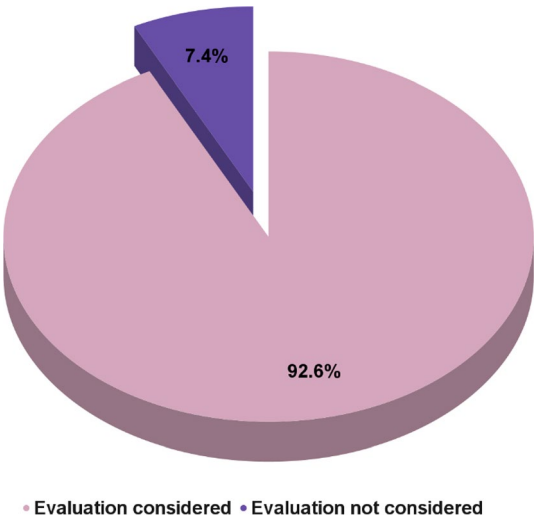

Fig. 10 The evaluation frequency employed by applications related to DL–AD

with making meds more specific. The process of reusing drugs, which can be done in six different ways, is also becoming known as a way to speed up the search for new treatments for AD. A lot of biology and chemical data is used by DL models to find drugs that could be used to treat the disease. These methods use trends and links in big datasets that have not been seen before to speed up and lower the cost of making new drugs. This could lead to better treatment options for patients.

The fact that NLP has five uses also shows how important text mining and data extraction is to the study of AD. A lot of written data from medical records, study papers, and patient files is used by NLP methods to find useful information. We can use these new ideas to find useful signs, understand how diseases work, and come up with study questions. New information can be found and choices can be made more quickly with NLP-based methods that pull data from a lot of different text sources. Figure 11 shows data that show a setting in the DL–AD area that is growing and changing. These results show that DL apps used to study AD are linked. DL is a key tool for learning more about diseases and coming up with new ways to identify and treat them. Some of these methods are early discovery, classification, drug reuse, and text mining. People from different fields are working together and using cutting-edge computer technology to deal with the complex issues that come with AD. The ultimate goal is to improve patient outcomes and quality of life.

6.4 Ethical concerns

There are ethical worries about using DL in AD that touch on some important issues that need careful thought and direct action. Concerns about privacy and data protection are at the top of this list. DL systems depend on large datasets that often contain private medical data. So, there are good reasons to be concerned about how this information is gathered, kept, and used, as well as the risks of privacy breaches or the wrong use of personal health information. The moral effects go beyond the basic idea of informed consent, especially for people with AD who might not be able to make decisions as well as they used to. It is hard to get people to give real consent for the use of their data in DL research, which makes me

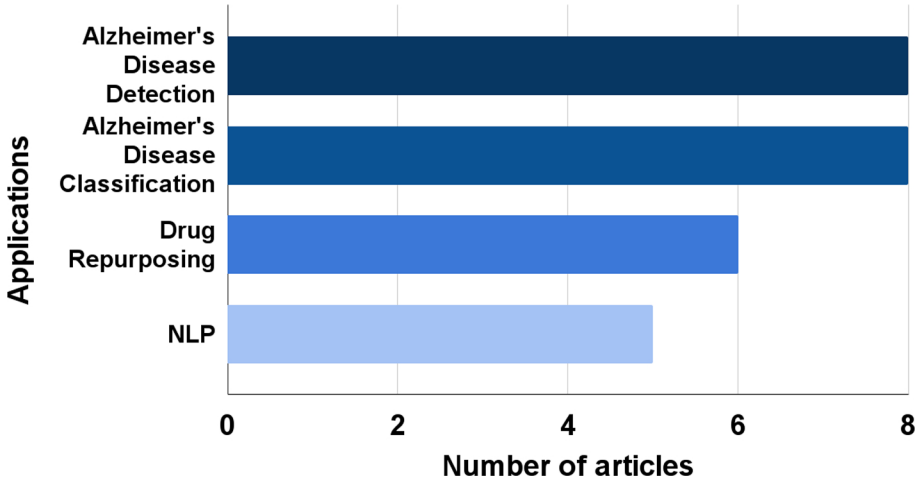

Fig. 11 The utilization of DL–AD applications across different distributions

wonder if substitute consent is enough to protect vulnerable groups. Ethical concerns also come up because the training data used for DL models has built-in biases. Biases can cause mistakes or make current differences in healthcare results worse. This shows how important it is to make sure that records are representative and fair to lower these risks.

Another important thing to think about is how DL algorithms can be understood and who is responsible for them. Because these models lack explanation, it is hard to figure out how they make decisions, which makes it hard to hold them accountable when mistakes happen or bad things happen. To keep healthcare gaps from getting worse, everyone needs equal access to DL technology and AD detection and treatment. Adding DL to the evaluation and treatment of AD also makes people wonder how it will affect the bond between the doctor and the patient. Patients might not feel comfortable depending on technology alone for important healthcare choices. This could change how patients and healthcare workers trust each other and talk to each other. Although DL algorithms are better at finding AD, there is a chance that they will diagnose or treat too many people with it. Overdiagnosis could cause patients to go through worry and care that they do not need, and remedies suggested based on incomplete data could lead to too much medicine and its negative effects. To solve these ethical problems, scholars, doctors, lawmakers, and ethicists need to work together. It is very important to set clear rules and standards for the development and use of DL applications in AD that put patient privacy, fairness, openness, and equality first. These apps need to be constantly checked and evaluated to make sure they follow ethical guidelines and give users real benefits while minimizing any possible harm.

6.5 Incorporation of recent studies on transformers and large language models

Originally intended for NLP tasks, transformers' capacity to manage intricate data linkages via self-attention processes has led to critical applications in the area of AD diagnosis. Long-range connections and complex patterns in data are particularly well captured by these models, which are essential for detecting minute changes linked to AD. Neuroimaging data with a lot of dimensions, like MRI and PET scans, has been handled with Vision Transformers (ViTs), a type of transformer designed for image processing. A new study shows that ViTs are better than regular CNNs at finding early signs of AD because they can describe and capture features better. For example, researchers discovered that ViTs did a better job than cutting-edge CNN models at sorting MRI data into groups of healthy controls and early-stage AD cases. This is because the transformer can focus on specific parts of the brain and record small changes in anatomy that are indicative of AD development. Large Language Models (LLMs), such as GPT-3 and BERT, have changed how text data is processed and analyzed. These models point to Alzheimer's study in new paths. These models can understand and write human language, which makes them useful for looking at clinical notes, electronic health records, and other random text data that contains important patient histories and symptoms. LLMs may be able to find trends and risk factors for AD by getting useful information from huge amounts of text data. For example, one study used LLMs to look at long-term medical data to predict the start and course of AD and found language signs that are linked to cognitive loss. LLMs may also help find new drugs by looking at data from clinical trials and academic papers to find possible therapy targets and find new uses for existing drugs to treat AD. In a study, researchers showed that data from clinical trials can be used to guess how well different drug molecules will work. This opens up new

treatment options for AD that have not been tried before. When transformers and LLMs are added to AD recognition systems, researchers can diagnose and understand AD in a whole new way. These advanced models are the best at handling and analyzing mixed data, which includes neuroimaging, clinical records, genetic information, and patient-reported results. By mixing the written data processing power of LLMs with the image analysis power of transformers, researchers may be able to make detailed models that show the whole picture of a patient's health. It was possible to make a combination model that greatly improved the accuracy of diagnoses and the number of early detections. For example, a new study that looked at both MRI images and clinical notes merged ViTs with BERT. AD may be easier to understand and record because it uses more than one mode of communication. The scale and constant learning of transformers and LLMs keep diagnostic tools up-to-date and effective. This is because these models can adapt to new data and lead to new research findings. These technologies do more than help with early detection. They also make personalized treatment plans possible by giving full patient profiles and predicting how each person will respond to treatments. We included these cutting-edge developments in our survey to give you the most up-to-date information and to show how transformers and LLMs have the potential to completely change how AD is diagnosed and treated, which will ultimately lead to better patient outcomes and a deeper understanding of the condition.

6.6 Innovative preprocessing techniques in MRI data analysis for AD classification

To make MRI data analysis more accurate and faster for classifying AD, we need to use advanced filtering methods. An intensity normalization method is used to make sure that all MRI images have the same intensity numbers on a standard scale. This step is necessary because differences in how the scanner is set up and how the patient is aligned could cause problems with the data. These changes are lessened by intensity normalization, which lets DL models learn from the input without being affected by differences that do not matter. MRI pictures are cleaned up even more with skull stripping, a more advanced preparation method that removes non-brain tissues. By isolating the specific part of the brain, this process makes the data less complicated and lets the models focus on the relevant physical structures linked to AD.

Using bias field correction is another new way to do preprocessing. MRI images often have uneven intensities because the magnetic field changes, which could lead to mistakes in the analysis. These problems are fixed with bias field correction methods like the N4ITK algorithm, which makes the levels more evenly spread across the pictures. Also, spatial normalization aligns MRI data from different people by putting them in the same physical space. Individual brain scans are turned into a common brain template as part of this process. This lets voxel-level comparisons be made between subjects. Researchers using this method are very close to finding common patterns of brain atrophy linked to AD. Also, data expansion techniques are used to make the MRI collection bigger and more varied than it is. Some methods for adding to data, like randomly spinning, moving, and scaling it, can make different copies of the original data. This makes it easier for DL models to generalize. Advanced enhancement methods, like elastic deformations and creating malicious training data, make the models more resilient by simulating real changes in the body and likely dangerous situations. Utilizing these advanced preparation techniques, researchers could significantly enhance the quality and reliability of MRI data, hence optimizing the performance

of DL models in identifying AD. These strategies reduce the challenges arising from diverse data sources and ensure that the models are trained on clean, consistent, and representative datasets. This ultimately results in more precise and dependable medical outcomes.

7 Key challenges and future works

This section assesses the mentioned articles to provide final views on accomplishments, prevalent concepts, and possible obstacles in the domain of DL applications within the AD sector. Although several limitations have been discussed in the analysis of the DL–AD trial in the previous sections, they were not the only issues. This part accordingly addresses them, along with potential research opportunities for DL applications and various approaches that may be explored for their effective management in the future.

7.1 Open issues and key challenges

DL applications in AD research have shown promise in various areas, including diagnosis, prognosis, classification, drug discovery, and understanding of disease mechanisms. However, there are several open issues and key challenges that researchers continue to face in this field. Here are some of the key challenges in applying DL to AD:

7.1.1 Data quality and quantity

There are many problems with the quality and amount of data in DL used for AD study. These problems are very important for making models that are dependable and accurate. To begin, the quality of the data used to train DL models has a big effect on how well they work and how well they can be used in other situations. Accessing high-quality data in AD studies can be especially hard because the disease is so variable and complicated. MRI and PET readings, which are types of neuroimaging data, can have noise, errors, and differences in how they are saved and interpreted. It is also possible for clinical data, like brain tests and DNA markers, to be wrong or missing values, which makes the training process even more difficult. To reduce errors and make DL models more reliable, it is important to make sure that the data is correct, consistent, and full by using strict quality control measures, data preparation methods, and expert annotation (Ye et al. 2023).

Second, the amount of data that can be used to train DL models is very important for making sure that the models can do enough and can generalize well. However, AD samples are often not very big, especially for rare disease groups and certain community categories. Overfitting happens when models remember the noise in the training data instead of the patterns and connections that lie beneath it. Small datasets can cause this. Also, because AD is so different for each person, it can be hard to get sample statistics that include people with different disease states, demographics, and genetic backgrounds. Dealing with the lack of data through methods like data augmentation, joint data-sharing, and multi-site studies can help make more and different types of training data available for developing DL models. Using transfer learning and pretraining on big datasets from related fields, like general medical imaging or neurodegenerative diseases, can also make DL models more useful and efficient in AD research. By focusing on data quality and amount, researchers can make DL

applications for AD detection, prediction, and treatment more reliable, useful, and able to be used in real life.

7.1.2 Interpretability and explainability

The ability to understand and explain DL models is very important for their growth and use in AD studies. These ideas are about being able to understand and defend the choices a model makes by showing the traits and trends it has picked up from the data. In AD, where correct identification and prediction are very important for patient care, models that can be understood can give us useful information about how the disease works biologically and what factors raise the risk of getting it. One way to make things easier to understand and explain is to use attention mechanisms. These draw attention to the most important parts or traits of brain data that help the model make predictions. By seeing these focus maps, scientists and doctors can learn more about the specific parts of the brain that are damaged by AD and figure out how the model makes its decisions. Feature importance methods like SHAP or LIME can also be used to find the most important features in the input data and measure how they change the model's output. Adding these methods for interpretability to DL models not only makes them clearer but also builds trust and acceptance among doctors and other partners, which makes it easier to use these models in clinical practice.

Also, being able to understand and explain things is very important for making sure that DL models for AD detection and prediction are correct and getting better. By knowing why, a model makes the statements it does, researchers can find possible flaws, confusing factors, or data limits. This lets them improve the model design and make it more generalizable and stable. Furthermore, interpretable models make it easier for doctors and data scientists to work together, because subject experts can provide useful ideas and information that can help with the development and understanding of the model. Finally, researchers can make these models more clear, reliable, and useful for patients by emphasizing how easy they are to understand and explain in DL research for AD. This will allow them to be used in everyday clinical practice and personalized patient care (Bloch and Friedrich 2024).

7.1.3 Generalization and robustness

Generalization means that a model can do well on new data from the same distribution as the training data that it has not seen before. In the case of AD, this means that a DL model that was trained on a certain dataset should be able to correctly identify or guess how people from different groups will do. But it can be hard to generalize because AD is heterogeneous, meaning that it shows up differently in different people and may depend on things like age, genetics, and other health problems. DL models need to be made in a way that captures the basic patterns and traits of AD that are common across a wide range of populations. This is to make sure that the model's results are accurate and consistent across a range of clinical settings and patient groups. On the other hand, robustness is how stable and resilient a DL model is to changes or perturbations in the input data. In AD study, strong models are needed to deal with noise, flaws, and other types of variation that are common in clinical data, like changes in image quality or patient profiles. Overfitting is less likely to happen with robust models, and they can keep working well even when they get data that is not distributed in the same way as the training data. It takes careful model building, regulariza-

tion methods, and data addition strategies to make sure that the model is exposed to a lot of different situations and can handle new problems that have not been seen before. By putting both generalization and stability at the top of the list when creating DL models for AD, researchers can make these models more reliable and useful for clinical use, which will eventually lead to better patient care and results (Eslami et al. 2023).

7.1.4 Data imbalance and bias

Data mismatch and bias are big problems that make it hard to build and use DL models for AD studies. According to the idea of "data imbalance", AD records often have an uneven spread between classes, with a lot more samples from healthy people than from people with AD. This difference in class sizes can cause models to make skewed predictions; the model may be more accurate for the majority class but less accurate for the minority class. So, the model might miss important trends and traits related to AD, which would make it less accurate at diagnosing and predicting how the disease will progress. Data bias in AD studies can also come from different places, such as differences in demographics, differences in clinical outcomes, and sample errors. For example, datasets may mostly show certain racial or medical groups, which can cause models to make skewed guesses that do not work well across different groups of people. Differences in imaging techniques or clinical exams are another example of a bias in the way data is collected that can cause regular mistakes and make model training harder. To fix data mismatch and bias, you need to carefully think about sampling strategies, data preparation methods, and mathematical changes to make sure that DL models are trained on datasets that are fair and representative. Additionally, researchers must actively look for and reduce biases during the model development process to make sure that the model works equally well for all patient groups and clinical situations (Agarwal et al. 2023).

7.1.5 Multi-modal integration

It is called multi-modal integration when information from different sources, such as neuroimaging, genetic markers, clinical exams, and cognitive tests, is used together in an AD study. This helps make it easier to diagnose, predict, and keep track of diseases. We can learn different things about different parts of the disease from each type of data. We can learn more about how AD starts and gets worse by putting all of these different kinds of data together. For instance, MRI and PET scans show how the brain changes in people with AD, both in terms of how it looks and how it works. Some genetic factors, such as the APOE genotype, make it more likely that someone will get the illness. AD symptoms and problems with working can be picked up by clinical exams and memory tests. This helps to find out what state the disease is in and keep an eye on it. Using DL to mix these different kinds of data makes it possible to study AD in a fuller way. This makes it easier to find the disease, plan specific treatments, and keep track of how it changes over time (Chen et al. 2022).

To get the most out of different data sources in the AD study, multi-modal interaction also comes with several technology issues that need to be fixed. Some of these problems are preprocessing and normalizing the data to make sure that it can be used with different types of data, extracting and choosing features to find biomarkers and other useful information from each data source, and using model fusion techniques to effectively combine data from

different types of data while keeping their unique qualities. DL techniques like multi-modal neural networks, attention mechanisms, and graph-based approaches show promise for combining different types of data and figuring out how they are connected in complicated ways. It is possible to learn new things about how AD works by combining different types of data that complement each other. This could lead to the discovery of strong biomarkers for early diagnosis and prognosis, as well as the creation of targeted interventions and personalized treatment plans that are specific to each patient's needs (George et al. 2023).

7.1.6 Longitudinal analysis

In AD studies, longitudinal analysis looks at how the shape, function, and cognitive skills of patients' brains change over time. This method is very important for figuring out how diseases develop, finding early signs of cognitive decline, and checking how well treatments work. DL models are very important in longitudinal analysis because they can find time correlations in longitudinal data and guess how diseases will progress in the future. However continuous analysis comes with some problems, such as data sparsity (because not all patients get regular follow-up exams) and handling missing values (because patients may drop out of studies or miss visits). There are also big methodological problems with matching and standardizing data across multiple time points and taking into account differences between and within subjects. DL methods like RNNs and longitudinal CNNs have been suggested as ways to deal with these problems by modeling time sequences and recording complex longitudinal patterns in the development of AD (Xia et al. 2023).

Furthermore, longitudinal analysis enables the identification of biomarkers associated with disease progression and treatment response, facilitating personalized medicine approaches in AD management. DL models trained on longitudinal data can identify subtle changes in brain imaging and clinical assessments that precede symptomatic onset, enabling early intervention and potentially delaying disease progression. Moreover, longitudinal analysis provides insights into the heterogeneity of AD trajectories, helping to stratify patients into distinct subtypes and tailor treatment strategies accordingly. However, interpreting longitudinal DL models and translating their findings into clinical practice remain ongoing challenges. Enhancing model interpretability and validation through collaboration with clinicians and incorporating domain knowledge into model development are essential steps toward realizing the full potential of longitudinal analysis in DL applications for AD research and patient care.

7.1.7 Clinical validation and translation

Clinical evaluation and translation are very important steps in the process of making DL apps for AD study. During the clinical validation phase, the performance of DL models is carefully checked using separate datasets that reflect the wide range of patients and clinical situations that happen in real healthcare settings. To find out how useful the model is for diagnosing or predicting, different measures like sensitivity, specificity, accuracy, and AUC-ROC must be looked at. In addition, clinical confirmation includes checking how DL models affect clinical decision-making, such as how well they help doctors with evaluation, risk assessment, and planning treatments. During this phase, it is important to understand

the limits and possible flaws of DL models. This tells doctors and researchers how reliable and applicable model results are to a wide range of patient groups and healthcare situations.

Once DL models have undergone rigorous clinical validation, the next step is translation into real-world clinical practice. This involves integrating the validated models into existing healthcare workflows and infrastructure, ensuring seamless interoperability with EHRs, medical imaging systems, and diagnostic tools (Kim et al. 2022). Additionally, translation requires addressing regulatory and legal considerations, such as obtaining approval from regulatory agencies (e.g., FDA) and ensuring compliance with data privacy and security regulations (e.g., HIPAA). Moreover, successful translation of DL models into clinical practice necessitates comprehensive training and education for healthcare professionals to effectively utilize these technologies and interpret their outputs in the context of patient care. Collaborative efforts between researchers, clinicians, healthcare administrators, and regulatory bodies are essential for overcoming the logistical, technical, and regulatory challenges associated with the translation of DL applications from research settings to routine clinical care in AD and other healthcare domains.

7.1.8 Privacy and security

Privacy and security concerns are paramount in the development and deployment of DL applications for AD research. AD datasets often contain sensitive medical information, including neuroimaging data, genetic profiles, and clinical histories, which pose significant risks to patient privacy if not adequately protected. DL models trained on such data may inadvertently reveal personal health information, leading to breaches of confidentiality and potential harm to individuals. Moreover, the increasing prevalence of data breaches and cyberattacks in healthcare settings underscores the need for robust security measures to safeguard AD datasets from unauthorized access and exploitation. Techniques such as encryption, access controls, and secure multi-party computation can help mitigate these risks by ensuring that sensitive data remains confidential and protected from malicious actors (Golovanevsky et al. 2022).

Furthermore, the integration of DL models into clinical workflows requires careful consideration of privacy and security implications to maintain patient trust and compliance with regulatory requirements. Healthcare organizations must implement comprehensive data governance policies and procedures to govern the collection, storage, and sharing of AD data, ensuring compliance with laws such as HIPAA in the United States. Additionally, transparent communication with patients regarding the use of their data for research purposes is essential to uphold ethical standards and promote patient autonomy (Wang et al. 2022c). Collaborative efforts between researchers, healthcare providers, and cybersecurity experts are essential to develop and implement robust privacy-preserving techniques that balance the need for data access and innovation with the protection of patient privacy and security in AD research.

7.1.9 Integration with clinical workflows

When DL models are seamlessly added to current healthcare systems and practices, this is called "integration with clinical workflows". Integrating these models is important to make sure that healthcare workers can use the information they give them to make good decisions

every day. The ability of DL and HER systems to work together technically is an important part of this problem. DL models need to be able to access and evaluate patient data kept in EHRs. This data can be organized or unstructured and includes a lot of different types of data, such as clinical notes, imaging reports, lab tests, and personal information. To make sure that these systems can talk to each other and share data, we need to create standard interfaces and protocols, as well as strong data preparation processes that can handle the variety and complexity of clinical data.

Moreover, integration with clinical workflows also entails considerations regarding scalability and real-time performance. DL models must be able to handle large volumes of data and provide timely insights to support clinical decision-making. This requires optimizing model architectures and algorithms for efficiency and deploying them on scalable computing infrastructure that can handle the computational demands of real-time inference (Zhang et al. 2023b). Additionally, DL models need to be continuously updated and refined based on feedback from healthcare professionals and new data sources, highlighting the importance of establishing feedback loops and mechanisms for model maintenance and improvement. Ultimately, successful integration with clinical workflows relies not only on technological advancements but also on addressing organizational and cultural barriers within healthcare institutions, fostering collaboration between data scientists and healthcare professionals, and ensuring that DL solutions align with the needs and priorities of clinical practice.

7.1.10 Ethical and societal implications

The ethical and societal implications of DL applications in AD research encompass a range of complex considerations that extend beyond technical challenges. Firstly, the use of DL models in Alzheimer's research raises important questions surrounding patient privacy, autonomy, and consent. As these models rely on large datasets containing sensitive medical information, ensuring the protection of patient data privacy becomes paramount. Researchers must navigate the ethical complexities of obtaining informed consent for data sharing and analysis, particularly considering the cognitive impairments and vulnerability of many AD. Additionally, there is a need to address concerns regarding data ownership, control, and potential exploitation, ensuring that individuals retain autonomy over their health information while still enabling collaborative research and innovation (Kim et al. 2023).

Furthermore, the deployment of DL technologies in AD diagnosis and treatment introduces potential biases and disparities that could exacerbate existing healthcare inequities. These biases may arise from various sources, including imbalanced datasets, algorithmic decision-making, and disparities in access to healthcare resources. Researchers must actively mitigate biases in model development and validation, employing techniques such as algorithmic fairness and bias detection to ensure equitable outcomes across diverse patient populations (Mohamed Yusof et al. 2023). Moreover, there is a broader societal responsibility to address the equitable distribution and accessibility of DL technologies, ensuring that advancements in Alzheimer's research benefit all individuals, regardless of socioeconomic status, geographic location, or other demographic factors. Engaging with stakeholders, including patients, caregivers, advocacy groups, and policymakers, is essential for promoting transparency, accountability, and inclusivity in the development and deployment of DL applications for AD. Addressing these open issues and key challenges requires interdisciplinary collaboration between computer scientists, neuroscientists, clini-

cians, and other stakeholders. By leveraging advances in DL, combined with insights from AD research and clinical practice, we can potentially improve early detection, diagnosis, and treatment monitoring of AD, ultimately enhancing patient outcomes and quality of life.

7.2 Future works

The future of DL uses in AD study is very bright. They could help us learn more about, find, diagnose, and treat this crippling neurological disease. Here is a full list of all the books that are expected to come out in the future in this field:

7.2.1 Early diagnosis and prediction

To fight this very bad brain disease, it is very important to find and identify AD early on. It is possible that DL tools could help us find AD earlier, even before any signs show up. DL models can use data from brain scans, memory tests, and genetic markers to find small trends and signs that point to changes in the disease that causes AD. It might be possible to find people who are very likely to get AD with these models. This would make it possible to find problems early and make personalized care plans. You can use DL to make models that show how likely it is that someone will go from MCI to AD. This helps doctors help and care for people in a more specific way. These prediction models could use more study to get better and more reliable. More types of data could be added to get a fuller picture of the risk, and DL could be used to find new biomarkers for AD that show up early. DL methods could also help us figure out the tricky road of AD growth. This would help us make better tools for predicting the disease and better care plans for each person. For a better understanding of how diseases develop over time, longitudinal data such as medical records, brain scans, and biomarker measurements are very useful. DL algorithms can look at these various types of data streams to discover time trends and biomarker profiles that are connected to various stages of AD. They can then figure out the best way to treat the disease based on how they think it will grow. Data mining techniques like functional connectivity and structural brain network modeling can also be used by DL models to help us figure out the chemical processes that make AD worse. Scientists can make personalized medicines that can slow or stop the disease from getting worse by learning more about how illnesses work. This will finally lead to better results for people with AD. So, more research could be done in this area to make combined models that use multi-modal continuous data, make it easier to predict how a disease will progress, and turn study results into ideas that can be used in the real world (Kang et al. 2023).

7.2.2 Disease progression modeling

In AD, disease progression modeling uses DL methods to look at long-term data that shows different parts of the disease's development. This method lets researchers keep an eye on small changes in biomarkers, brain patterns, and clinical measures over time. This helps them figure out how AD progresses from its early stages to later dementia. DL models can find complicated patterns in continuous data, which lets us describe how diseases progress in different people. This could help group patients based on their risk levels and guide specific treatment plans. These models can also help find important biomarkers and imaging

markers linked to different stages of AD progression. This will shed light on the underlying neurobiological mechanisms and make it easier to create targeted interventions that can slow or stop the disease's progression. Also, using DL methods to model how the disease progresses could help improve the plans of clinical trials and ways to measure outcomes in AD studies. By showing the different ways that diseases get worse and finding biomarkers that can predict when a patient will get worse, these models can help choose the best groups of patients for clinical trials and make outcome measures more accurate. This makes drug development faster and cheaper. DL-based disease progression models can also make it easier to test disease-modifying therapies by giving numbers that show how well treatments work and how quickly diseases get worse. This speeds up the process of bringing promising interventions from preclinical studies to clinical practice. Overall, using DL methods to model disease progression is a strong way to figure out how AD develops, help with early intervention strategies, and eventually lead to the creation of effective treatments for this terrible neurological disorder.

7.2.3 Drug discovery and development

The way drugs are found and made for AD could change a lot if DL is used. DL algorithms use very large datasets from genes, proteomics, metabolomics, and structural biology to find new drug targets, guess how drugs will interact with targets, and speed up the screening of possible medical chemicals. Big sets of biology data can be used with these models to find small trends. Researchers can use this to find small DNA marks that are linked to the growth of AD. It is also easier to look into large areas of chemistry with DL, which helps find small chemicals that could be used as medicines. In silico drug screening can help DL algorithms rank potential chemicals for further tests to prove their safety. It might be cheaper take less time and happen faster with this method than with the usual ones. Researchers may work on improving and simplifying DL models in the future. They may also combine various types of data to learn more about how complicated AD is and use what they learn from DL studies to come up with smart ways to make drugs.

DL methods could also change how specific medicines are used to treat AD. DL models can make sure that every patient gets the best medicine with the fewest side effects by looking at multiple types of patient data, such as genetic traits, biomarker trends, and clinical results. Based on each patient's unique traits and how the disease develops, these models can help doctors figure out the best ways to treat them. Also, DL algorithms can help find clues about how well a drug will work. This lets early action and control of AD happen. We might make treatment programs that can change how they do therapy based on real-time data about patients as time goes on. We might also take the patients' thoughts and ideas into account when we make decisions. When digital learning and personalized treatment work together, they might be able to help people with AD get better results and have a better quality of life.

7.2.4 Neuroimaging analysis

To study AD and find it in people, neuroimaging is very important. DL methods, especially CNNs, are now good ways to automatically look at different kinds of brain scans, like fMRI, PET, and structural MRIs. From large sets of data, these models can learn complicated

trends and traits. This helps correctly identify and group issues connected to AD, such as the shrinking of the hippocampus, the building of amyloid-beta, and changes in functional connectivity networks. Brain study that uses DL should get better in the future, which should help us learn more about how AD works by finding small signs that are linked to different types and stages of the disease. Another thing that could help make evaluations and predictions more accurate is adding continuous information and mixed image data to DL models. People who are at risk or in the early stages of AD might be able to get better care and start treatment earlier. DL is also used for more than just detection in brain studies for AD. Researchers also use it to learn how the disease affects the brain. Using large neuroimaging datasets, DL models can find intricate patterns and links between different parts of the brain. These details help us comprehend the adjustments in form and purpose that happen when AD gets worse. The study's results not only teach us more about AD, but also point us in the direction of possible treatment and drug research areas. Making DL models that both therapists and researchers can understand can also help them figure out the brain traits that make model predictions work. This makes it easier to use study results in real life. Multidisciplinary teams made up of neuroscientists, radiologists, and computer scientists will need to work together to move the field of brain research using DL forward and speed up the creation of new ways to diagnose and treat AD.

7.2.5 Personalized treatment planning

Personalized treatment planning in AD is a big change in the way healthcare is done because it aims to make treatments fit the specific traits and needs of each patient. DL is very important in this area because it looks at different kinds of patient data, like genetic profiles, biomarker patterns, brain results, and clinical records, to help make choices about treatment. These models can find groups of patients who might react differently to different treatments. This lets doctors decide which treatments to use first and how to make them more effective and safer. DL methods also make it possible to add real-time information about patients to treatment algorithms. This lets therapy plans be changed automatically as the disease gets worse. Healthcare professionals can improve the quality of life for people with AD by using the power of personalized medicine to get the best results from treatments and reduce side effects.

Also, individual treatment plans could change the way clinical studies and drug research are done in AD. A one-size-fits-all method is common in traditional clinical studies, which can cause different people to respond differently to treatment and have high dropout rates. DL models can divide patient groups into groups based on genetic subtypes, how the disease progresses, and how well a treatment works. This lets researchers make more focused and effective clinical trials. By finding signs that can determine whether a treatment will work or whether the condition will get worse, these models can fill study groups with patients who are most likely to benefit from new medicines. This increases statistical power and decreases sample sizes. DL can also help find new drug mixtures or new uses for current chemicals by finding patient subgroups with unique pharmacological profiles or effects that work well together. In the end, individual treatment planning could speed up the creation of effective AD treatments and start a new era of precision medicine in neurological diseases (Hu et al. 2023).

7.2.6 Digital biomarkers and remote monitoring

Digital biomarkers and online tracking are new areas of study that are growing quickly in the field of AD. They offer creative new ways to find small changes in patients' cognitive and functional skills. These biomarkers include different digital signs that come from smartphones, wearable tech, and other digital devices. They provide constant amounts of data that can help us understand how diseases are progressing and how well treatments are working. For example, movement analysis using wearable sensors can find early signs of motor weakness linked to AD. Speech and language patterns caught by smartphone apps may show minor language changes that are a sign of cognitive decline. Using DL methods to look at these digital biomarkers has a huge potential to improve finding and keeping an eye on AD early, allowing early treatments to stop the disease from getting worse and better results for patients. Digital biomarkers also make online tracking possible, which is a cost-effective and scalable way to provide healthcare. This is especially true for people with AD who may have trouble making regular doctor visits with their caretakers. DL algorithms allow researchers to create predictive models that constantly look at live data from wearable tech and other digital sources. These models can then warn healthcare workers about changes from normal trends that may need more testing or action. These digital signals can also help with personalized treatment plans by giving doctors real-time information on how patients are responding to interventions. This lets them make therapies fit each patient's needs and tastes. As digital biomarkers and remote monitoring continue to develop, it will be important for clinicians, data scientists, and technology developers to work together across disciplines to test these methods in large-scale clinical studies and make them a normal part of clinical practice. This will ultimately lead to better outcomes for people with AD (Illakiya and Karthik 2023b).

7.2.7 XAI and clinical interpretability

Within the field of AD study, XAI involves creating DL models that not only make correct predictions but also give good reasons for those forecasts. This is very important in healthcare situations where decisions made by AI models need to be easy for both doctors and patients to understand and analyze. In AD, where complicated neural processes cause the disease to get worse, XAI can help us understand the traits and trends that make up the model's predictions. For instance, DL models that look at scan data might point out biomarkers or parts of the brain that show damage from AD. Being able to see how the sickness works helps doctors decide how to best treat their patients. It is also possible to make DL models more useful in real healthcare settings by making them easier for doctors to understand. Clinicians often decide what to identify and how to treat a patient by looking at a lot of information about the patient, their medical history, and what they know about the subject. DL algorithms can make better clinically useful guesses and ideas when they use subject knowledge to build and review models. One example is that a DL model for AD might look at more than just brain data. Memory tests, family history of risk, and the patient's medical history may also be looked at. This would show the whole picture of how the patient is doing. In the long run, this way makes care for people with AD better because it makes healthcare workers more open to and accepting of AI systems. It also makes sure that AI-driven findings are in line with clinical practice and patient needs.

7.2.8 Data sharing and collaboration

It is important to share data and work together to improve the study of DL uses in AD. Sharing DNA, clinical, brain, and omics data is possible when researchers, businesses, and groups work together. These datasets are very important for making deep-learning models that work well in a wide range of situations. Researchers can reach larger and more different groups of people when they share data. Because of this, they can make models that are more accurate and effective for early evaluation, prediction, and planning of treatment. Standards for data types and methods are also pushed by collaboration. This makes sure that studies are all the same and can be compared. People from different fields are more likely to work together, which not only makes it easier to repeat study results but also lets thoughts and skills from many fields be used to solve the hard problems of AD. In the DL for AD project, people do more than just share data when they work together. They get to know companies, healthcare workers, and support groups for patients. DL technologies can be easier to add to healthcare systems when businesses work together. This can help study results be used in the real world. Doctors and patient support groups work together to make sure that DL models are made with the end user in mind, taking into account real-life patient tastes and clinical needs. DL in AD research can have a bigger effect and speed up progress toward better diagnosis, medicines, and eventually better results for people with this deadly neurological disease by encouraging teamwork across disciplines and community involvement.

Future works in DL applications for AD will likely focus on improving the accuracy and interpretability of predictive models, advancing our understanding of disease mechanisms, accelerating drug discovery efforts, and translating research findings into clinical practice for personalized diagnosis and treatment of AD. Collaboration and interdisciplinary efforts will be key to realizing the full potential of DL in addressing the complex challenges posed by AD (Fathi et al. 2024).

8 Conclusion and limitations

This study looked at several DL uses in AD. These included key areas like detecting AD, classifying AD, reusing drugs, and NLP. We started by going over some basic DL ideas and then talked about the newest developments in the field. We asked important "what," "why," and "how" questions to get more information, which led to a thorough examination of DL–AD applications from different points of view. We found important trends and problems by doing a thorough data analysis. In particular, CNNs were the most common method, being used in 25.7% of the cases. Python was the most common computer language, used in 53.1% of DL–AD studies. Grad-CAM was the most common XAI system, used in 45.5% of cases, mostly for jobs like finding AD and categorizing it. Additionally, our review showed that there are big holes in the present study, mainly when it comes to security. Alarmingly, 88.9% of the studies that were looked at did not talk about security issues. This shows that we need to build DL models for AD in a way that is more aware of security issues right away. Even though progress has been made, there are still important issues that need to be dealt with right away, such as privacy, accessibility, fairness, and transferability, to make sure that AI is used in healthcare responsibly and effectively. We also found some problems with the literature we looked at, like how hard it was to find papers that were not written in

English, how incomplete the accounts of the algorithms were, and how there were not any standard ratings for new methods. Taking care of these problems is very important for the long-term growth and success of DL uses in AD studies. In the end, this study adds important new information to the field of DL in AD, which is changing very quickly. It will help guide future research and lead to even more new ideas. We think that this growing body of study will lead to more useful, safe, and understandable DL models that will help find Alzheimer's disease earlier and give each person a better chance of getting better care.

However, we highlight some drawbacks, such as difficulties obtaining non-English publications, inadequate algorithmic descriptions in some of the reviewed studies, and the lack of comparable evaluations for new methods. The development and sustainability of DL applications in AD will depend on how well these issues are addressed in the future.

Acknowledgements The authors would like to thank Dr. Amin Roshanaei for their support of this research.

Author contributions ST, RH, AH, and NJN contributed equally to the conception and design of the study. ST, RH, AH, and NJN contributed equally to the acquisition and analysis of data. ST, RH, AH, and NJN contributed equally to drafting the text or preparing the figures.
All authors contributed eqaully.

Funding Not Applicable.

Data availability The paper contains all of the data.
No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
The authors declare no competing interests.

Ethical approval Not Applicable.

References

Abdelwahab MM, Al-Karawi KA, Semary HE (2023) Deep learning-based prediction of Alzheimer's disease using microarray gene expression data. Biomedicines 11(12):3304

Adhikari S et al (2022) Exploiting linguistic information from Nepali transcripts for early detection of Alzheimer's disease using natural language processing and machine learning techniques. Int J Hum-Comput Stud 160:102761

Agarwal R, Sathwik AS, Godavarthi D, Ramesh JVN (2023) Comparative analysis of deep learning models for multiclass Alzheimer's disease classification. EAI Endorsed Trans Pervasive Health Technol. https://doi.org/10.4108/eetpht.9.4334

Ahmadzadeh M et al (2023) Neuroimaging and machine learning for studying the pathways from mild cognitive impairment to Alzheimer's disease: a systematic review. BMC Neurol 23(1):309

Alamro H, Thafar MA, Albaradei S, Gojobori T, Essack M, Gao X (2023) Exploiting machine learning models to identify novel Alzheimer's disease biomarkers and potential targets. Sci Rep 13(1):4979

Allada A, Bhavani R, Chaduvula K, Priya R (2024) CSCOOT: competitive swarm coot optimization-based CNN transfer learning for Alzheimer's disease classification. Int J Intell Syst Appl Eng 12(7s):337–349

Alqahtani S, Alqahtani A, Zohdy MA, Alsulami AA, Ganesan S (2023a) Severity grading and early detection of Alzheimer's disease through transfer learning. Information 14(12):646

Alqahtani N et al (2023b) Deep belief networks (DBN) with IoT-based Alzheimer's disease detection and classification. Appl Sci 13(13):7833

Amini M, Pedram M, Moradi A, Ouchani M (2021) Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN). Comput Math Methods Med 2021:1–15

Amini R, Moradi S, Najafi R, Mazdeh M, Taherkhani A (2024) BACE1 inhibition utilizing organic compounds holds promise as a potential treatment for Alzheimer's and Parkinson's diseases. Oxid Med Cell Longev 2024:6654606

Bae JB et al (2023) A case–control clinical trial on a deep learning-based classification system for diagnosis of amyloid-positive Alzheimer's disease. Psychiatry Investig 20(12):1195

Bakare R, Shete VV, Kompatsiaris I, Tsolaki M (2024) Revolutionizing the Alzheimer's disease stage diagnosis through AI-powered approach. Int J Intell Syst Appl Eng 12(13s):407–416

Balasundaram A, Srinivasan S, Prasad A, Malik J, Kumar A (2023) Hippocampus segmentation-based Alzheimer's disease diagnosis and classification of MRI images. Arab J Sci Eng 48(8):10249–10265

Bapat R, Ma D, Duong TQ (2023) Predicting four-year's Alzheimer's disease onset using longitudinal neurocognitive tests and MRI data using explainable deep convolutional neural networks. J Alzheimer's Dis 97(1):1–11

Bhandarkar A, Naik P, Vakkund K, Junjappanavar S, Bakare S, Pattar S (2024) Deep learning based computer aided diagnosis of Alzheimer's disease: a snapshot of last 5 years, gaps, and future directions. Artif Intell Rev 57(2):1–62

Bloch L, Friedrich CM (2024) and Alzheimer's Disease Neuroimaging Initiative, Systematic comparison of 3D deep learning and classical machine learning explanations for Alzheimer's Disease detection. Comput Biol Med 170:108029

Bohn L, Drouin SM, McFall GP, Rolfson DB, Andrew MK, Dixon RA (2023) Machine learning analyses identify multi-modal frailty factors that selectively discriminate four cohorts in the Alzheimer's disease spectrum: a COMPASS-ND study. BMC Geriatr 23(1):837

Chen Y et al (2022) Contrastive learning for prediction of Alzheimer's disease using brain ^{18}F-FDG PET. IEEE J Biomed Health Inform 27(4):1735–1746

Ching WP, Abdullah SS, Shapiai MI, Islam AM (2024) Transfer learning for Alzheimer's disease diagnosis using EfficientNet-B0 convolutional neural network. J Adv Res Appl Sci Eng Technol 35(1):181–191

Chyr J, Gong H, Zhou X (2022) DOTA: deep learning optimal transport approach to advance drug repositioning for Alzheimer's disease. Biomolecules 12(2):196

Clarke N, Barrick TR, Garrard P (2021) A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. Front Comput Sci 3:634360

De Santi LA, Pasini E, Santarelli MF, Genovesi D, Positano V (2023) An explainable convolutional neural network for the early diagnosis of Alzheimer's disease from ^{18}F-FDG PET. J Digit Imaging 36(1):189–203

Dhaygude AD et al (2024) Knowledge-based deep learning system for classifying Alzheimer's disease for multi-task learning. CAAI Trans Intell Technol 9(4):805–820

Eslami M, Tabarestani S, Adjouadi M (2023) A unique color-coded visualization system with multimodal information fusion and deep learning in a longitudinal study of Alzheimer's disease. Artif Intell Med 140:102543

Fathi S, Ahmadi A, Dehnad A, Almasi-Dooghaee M, Sadegh M, Alzheimer's Disease Neuroimaging Initiative (2023) A deep learning-based ensemble method for early diagnosis of Alzheimer's disease using MRI images. Neuroinformatics 22(1):1–17

Fathi S, Ahmadi A, Dehnad A, Almasi-Dooghaee M, Sadegh M, Alzheimer's Disease Neuroimaging Initiative (2024) A deep learning-based ensemble method for early diagnosis of Alzheimer's disease using MRI images. Neuroinformatics 22(1):89–105

Ferrante M, Boccato T, Toschi N (2024) Enabling uncertainty estimation in neural networks through weight perturbation for improved Alzheimer's disease classification. Front Neuroinform 18:1346723

Gao X, Liu H, Shi F, Shen D, Liu M (2023a) Brain status transferring generative adversarial network for decoding individualized atrophy in Alzheimer's disease. IEEE J Biomed Health Inform 27(10):4961–4970
Gao X, Cai H, Liu M (2023b) A hybrid multi-scale attention convolution and aging transformer network for Alzheimer's disease diagnosis. IEEE J Biomed Health Inform 27(7):3292–3301
George A, Abraham B, George N, Shine L, Ramachandran S (2023) An efficient 3D CNN framework with attention mechanisms for Alzheimer's disease classification. Comput Syst Sci Eng 47(2):2097–2118
Golovanevsky M, Eickhoff C, Singh R (2022) Multimodal attention-based deep learning for Alzheimer's disease diagnosis. J Am Med Inform Assoc 29(12):2014–2022
Hsieh K-L, Plascencia-Villa G, Lin K-H, Perry G, Jiang X, Kim Y (2023) Synthesize heterogeneous biological knowledge via representation learning for Alzheimer's disease drug repurposing. Iscience 26(1):105678
Hu J, Qing Z, Liu R, He K (2021) Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease. Front Neurosci 14:626154
Hu Z, Li Y, Wang Z, Zhang S, Hou W (2023) and Alzheimer's Disease Neuroimaging Initiative, Conv-Swinformer: integration of CNN and shift window attention for Alzheimer's disease classification. Comput Biol Med 164:107304
Illakiya T, Karthik R (2023a) Automatic detection of Alzheimer's disease using deep learning models and neuro-imaging: current trends and future perspectives. Neuroinformatics 21(2):339–364
Illakiya T, Karthik R (2023b) A dimension centric proximate attention network and swin transformer for age-based classification of mild cognitive impairment from brain MRI. IEEE Access 11:128018–128031
Illakiya T, Karthik R, Alzheimer's Disease Neuroimaging Initiative (2024) A deep feature fusion network with global context and cross-dimensional dependencies for classification of mild cognitive impairment from brain MRI. Image vis Comput 144:104967
Janghel R, Rathore Y (2021) Deep convolution neural network based system for early diagnosis of Alzheimer's disease. IRBM 42(4):258–267
Jia H et al (2024) Assessing the potential of data augmentation in EEG functional connectivity for early detection of Alzheimer's disease. Cogn Comput 16(1):229–242
Kadri R, Bouaziz B, Tmar M, Gargouri F (2021) CrossViT wide residual squeeze-and-excitation network for Alzheimer's disease classification with self attention ProGAN data augmentation. Int J Hybrid Intell Syst 17(3–4):163–177
Kang W, Lin L, Sun S, Wu S (2023) Three-round learning strategy based on 3D deep convolutional GANs for Alzheimer's disease staging. Sci Rep 13(1):5750
Khare SK, Acharya UR (2023) Adazd-Net: automated adaptive and explainable Alzheimer's disease detection system using EEG signals. Knowl Based Syst 278:110858
Kim JS et al (2022) Deep learning-based diagnosis of Alzheimer's disease using brain magnetic resonance images: an empirical study. Sci Rep 12(1):18007
Kim H-B et al (2023) Modulation of associations between education years and cortical volume in Alzheimer's disease vulnerable brain regions by Aβ deposition and APOE ε4 carrier status in cognitively normal older adults. Front Aging Neurosci 15:1248531
Kumar SS, Nandhini M (2021) Entropy slicing extraction and transfer learning classification for early diagnosis of Alzheimer diseases with sMRI. ACM Trans Multimed Comput Commun Appl 17(2):1–22
Kwak MG et al (2023) Self-supervised contrastive learning to predict the progression of Alzheimer's disease with 3D amyloid-PET. Bioengineering 10(10):1141
Mahim S et al (2024) Unlocking the potential of XAI for improved Alzheimer's disease detection and classification using a ViT-GRU model. IEEE Access 12:8390–8412
Menagadevi M, Devaraj S, Madian N, Thiyagarajan D (2024) Machine and deep learning approaches for Alzheimer disease detection using magnetic resonance images: an updated review. Measurement 226:114100
Mohamed Yusof NIS, Awaluddin NA, Fauzi FM (2023) Insight into the structure and physicochemical properties of potent chemokine receptor 5 inhibitors for the discovery of novel Alzheimer's disease drugs. Cent Nerv Syst Agents Med Chem Formerly Curr Med Chem-Cent Nerv Syst Agents 23(2):95–108
Nanthini K, Tamilarasi A, Sivabalaselvamani D, Suresh P (2024) Automated classification of Alzheimer's disease based on deep belief neural networks. Neural Comput Appl 36:1–15
Pallawi S, Singh DK (2023) Review and analysis of deep neural network models for Alzheimer's disease classification using brain medical resonance imaging. Cogn Comput Syst 5(1):1–13
Pan D et al (2022) Adaptive 3DCNN-based interpretable ensemble model for early diagnosis of Alzheimer's disease. IEEE Trans Comput Soc Syst 11(1):247–266
Pradhan N, Sagar S, Singh AS (2023) Analysis of MRI image data for Alzheimer disease detection using deep learning techniques. Multimed Tools Appl 83:1–24
Pradhan N, Sagar S, Jagadesh T (2024) Advance convolutional network architecture for MRI data investigation for Alzheimer's disease early diagnosis. SN Comput Sci 5(1):167

Qiao J et al (2024) Deep neural network CSES-NET and multi-channel feature fusion for Alzheimer's disease diagnosis. Biomed Signal Process Control 87:105482
Rajasree R, Brintha Rajakumari S (2023) Ensemble-of-classifiers-based approach for early Alzheimer's Disease detection. Multimed Tools Appl 83:1–29
Roncero-Parra C, Parreño-Torres A, Sánchez-Reolid R, Mateo-Sotos J, Borja AL (2024) Inter-hospital moderate and advanced Alzheimer's disease detection through convolutional neural networks. Heliyon 10(4):e26298
Roshanzamir A, Aghajan H, Soleymani Baghshah M (2021) Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. BMC Med Inform Decis Mak 21:1–14
Saleh AW, Gupta G, Khan SB, Alkhaldi NA, Verma A (2023) An Alzheimer's disease classification model using transfer learning DenseNet with embedded healthcare decision support system. Decis Anal J 9:100348
Sanjay V, Swarnalatha P (2023) An overview of deep learning approaches for Alzheimer's disease classification: a review. J Adv Res Appl Sci Eng Technol 33(1):122–140
Sekhar J, Rajyalakshmi C, Nagaraj S, Sankar S, Saturi R, Harshavardhan A (2023) Deep generative adversarial networks with marine predators algorithm for classification of Alzheimer's disease using electroencephalogram. J King Saud Univ Comput Inf Sci 35(10):101848
Sellappan S, Anand SP, Shadrach FD, Krishnasamy B, Karra R, Annamalai U (2024) A survey of Alzheimer's disease diagnosis using deep learning approaches. J Auton Intell 7(3)
Seriramulu V et al (2024) Review of MR spectroscopy analysis and artificial intelligence applications for the detection of cerebral inflammation and neurotoxicity in Alzheimer's disease. Med J Malays 79(1):102–110
Shanmugavadivel K, Sathishkumar V, Cho J, Subramanian M (2023) Advancements in computer-assisted diagnosis of Alzheimer's disease: a comprehensive survey of neuroimaging methods and AI techniques for early detection. Ageing Res Rev 91:102072
Shen Z et al (2022) Classifying the lifestyle status for Alzheimer's disease from clinical notes using deep learning with weak supervision. BMC Med Inform Decis Mak 22(1):1–11
Shi Y, Chen Z, Huang L, Gong Y, Shi L (2024) A network pharmacology approach to reveal the key ingredients in Scrophulariae Radix (SR) and their effects against Alzheimer's disease. Heliyon 10(3):e24785
Shukla GP, Kumar S, Pandey SK, Agarwal R, Varshney N, Kumar A (2023) Diagnosis and detection of Alzheimer's disease using learning algorithm. Big Data Min Anal 6(4):504–512
Sorour SE, Abd El-Mageed AA, Albarrak KM, Alnaim AK, Wafa AA, El-Shafeiy E (2024a) Classification of Alzheimer's disease using MRI data based on deep learning techniques. J King Saud Univ Comput Inf Sci 36(2):101940
Srividhya L, Sowmya V, Ravi V, Gopalakrishnan EA, Soman KP (2023) Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease. Multimed Tools Appl 83:1–24
Srividhya L, Sowmya V, Ravi V, Gopalakrishnan EA, Soman KP (2024) Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease. Multimed Tools Appl 83(6):16799–16822
Tsuji S et al (2021) Artificial intelligence-based computational framework for drug-target prioritization and inference of novel repositionable drugs for Alzheimer's disease. Alzheimer's Res Ther 13(1):92
Uyguroğlu F, Toygar Ö, Demirel H (2024) CNN-based Alzheimer's disease classification using fusion of multiple 3D angular orientations. Signal Image Video Process 3:1–9
Varghese A, George B, Sherimon V, Shuaily A, Salim H (2023) Enhancing trust in Alzheimer's disease classification using explainable artificial intelligence: incorporating local post hoc explanations for a glassbox model. Bahrain Med Bull 45(2):1471–1478
Vidhya R et al (2023) Alzheimer's disease detection using residual neural network with LSTM hybrid deep learning models. J Intell Fuzzy Syst 45(6):1–15
Wang S, Du Z, Ding M, Rodriguez-Paton A, Song T (2022a) KG-DTI: a knowledge graph based deep learning method for drug–target interaction predictions and Alzheimer's disease drug repositions. Appl Intell 52(1):846–857
Wang JX, Li Y, Li X, Lu Z-H (2022b) Alzheimer's disease classification through imaging genetic data With IGNet. Front Neurosci 16:846638
Wang C et al (2022c) Early diagnosis of Alzheimer's disease and mild cognitive impairment based on electroencephalography: from the perspective of event related potentials and deep learning. Int J Psychophysiol 182:182–189
Warren SL, Moustafa AA (2023) Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. J Neuroimaging 33(1):5–18
Wu Y, Liu Q, Qiu Y, Xie L (2022) Deep learning prediction of chemical-induced dose-dependent and context-specific multiplex phenotype responses and its application to personalized Alzheimer's disease drug repurposing. PLoS Comput Biol 18(8):e1010367

Wu W, Holkeboer KJ, Kolawole TO, Carbone L, Mahmoudi E (2023) Natural language processing to identify social determinants of health in Alzheimer's disease and related dementia from electronic health records. Health Serv Res 58(6):1292–1302
Xia W, Zhang R, Zhang X, Usman M (2023) A novel method for diagnosing Alzheimer's disease using deep pyramid CNN based on EEG signals. Heliyon 9(4):e14858
Xu E et al (2022a) Pathology steered stratification network for subtype identification in Alzheimer's disease. Med Phys 51(2):1190–1202
Xu J et al (2022b) Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. Cell Rep 41(9):111717
Xu F, Zheng Q, Shi J, Yan K (2023) Pre-training and ensembling based Alzheimer's disease detection. Technol Health Care 32(1):1–17
Xu E et al (2024) Pathology steered stratification network for subtype identification in Alzheimer's disease. Med Phys 51(2):1190–1202
Yang Z, Li K, Gan H, Huang Z, Shi M, Zhou R (2024) An Alzheimer's disease classification network based on MRI utilizing diffusion maps for multi-scale feature fusion in graph convolution. Math Biosci Eng 21(1):1554–1572
Yao Z et al (2023) Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. Eur J Radiol 165:110934
Ye H, Zhu Q, Yao Y, Jin Y, Zhang D (2023) Pairwise feature-based generative adversarial network for incomplete multi-modal Alzheimer's disease diagnosis. Vis Comput 39(6):2235–2244
Yi F et al (2023) Identifying underlying patterns in Alzheimer's disease trajectory: a deep learning approach and Mendelian randomization analysis. EClinicalMedicine 64:102247
Yu L, Liu J, Wu Q, Wang J, Qu A (2023) A Siamese-transport domain adaptation framework for 3D MRI classification of gliomas and Alzheimer's diseases. IEEE J Biomed Health Inform 28(1):381–402
Zang C et al (2023) High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data. Nat Commun 14(1):8180
Zhang X, Han L, Zhu W, Sun L, Zhang D (2021) An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE J Biomed Health Inform 26(11):5289–5297
Zhang Y, He X, Liu Y, Ong CZL, Liu Y, Teng Q (2023a) An end-to-end multimodal 3D CNN framework with multi-level features for the prediction of mild cognitive impairment. Knowl Based Syst 281:111064
Zhang X, Han L, Han L, Chen H, Dancey D, Zhang D (2023b) sMRI-PatchNet: a novel efficient explainable patch-based deep learning network for Alzheimer's disease diagnosis with Structural MRI. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3321220
Zhang L, Wang L, Liu T, Zhu D (2024) Disease2Vec: encoding Alzheimer's progression via disease embedding tree. Pharmacol Res 199:107038
Zhao Z et al (2023) Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: a review. Front Comput Neurosci 17:1038636
Zhao L, Li Z, Chen G, Yin Y, Chen CY-C (2024) Multi-perspective neural network for dual drug repurposing in Alzheimer's disease. Knowl Based Syst 283:111195
Zhou Q, Wang J, Yu X, Wang S, Zhang Y (2023) A survey of deep learning for Alzheimer's disease. Mach Learn Knowl Extr 5(2):611–668
Zhou J, Zhao M, Yang Z, Chen L, Liu X (2024) Exploring the value of MRI measurement of hippocampal volume for predicting the occurrence and progression of Alzheimer's disease based on artificial intelligence deep learning technology and evidence-based medicine meta-analysis. J Alzheimer's Dis 97(3):1–14

Authors and Affiliations

Shiva Toumaj[1] · Arash Heidari[2,3] · Reza Shahhosseini[4] · Nima Jafari Navimipour[5,6,7]

✉ Arash Heidari
Arash_heidari@ieee.org

Shiva Toumaj

shiva_tumaj@gmail.com

Reza Shahhosseini
Rshahh93@gmail.com

Nima Jafari Navimipour
nima.navimipour@khas.edu.tr

[1] Urmia University of Medical Sciences, Urmia, Iran

[2] Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

[3] Department of Computer Engineering, Faculty of Engineering and Natural Science, İstanbul Atlas University, Istanbul, Türkiye

[4] Istanbul Medipol University, Istanbul, Turkey

[5] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Kadir Has University, Istanbul, Turkey

[6] Future Technology Research Center, National Yunlin University of Science and Technology, Douliou 64002, Yunlin, Taiwan

[7] Research Center of High Technologies and Innovative Engineering, Western Caspian University, Baku, Azerbaijan