Check for updates

Understanding machine learning applications in dementia research and clinical practice: a review for biomedical scientists and clinicians

Yihan Wang[1,2], Shu Liu[1,2,3], Alanna G. Spiteri[1], Andrew Liem Hieu Huynh[4,5], Chenyin Chu[1,2], Colin L. Masters[1], Benjamin Goudey[2,3], Yijun Pan[1,2]* and Liang Jin[1,2]

Abstract

Several (inter)national longitudinal dementia observational datasets encompassing demographic information, neuroimaging, biomarkers, neuropsychological evaluations, and muti-omics data, have ushered in a new era of potential for integrating machine learning (ML) into dementia research and clinical practice. ML, with its proficiency in handling multi-modal and high-dimensional data, has emerged as an innovative technique to facilitate early diagnosis, differential diagnosis, and to predict onset and progression of mild cognitive impairment and dementia. In this review, we evaluate current and potential applications of ML, including its history in dementia research, how it compares to traditional statistics, the types of datasets it uses and the general workflow. Moreover, we identify the technical barriers and challenges of ML implementations in clinical practice. Overall, this review provides a comprehensive understanding of ML with non-technical explanations for broader accessibility to biomedical scientists and clinicians.

Keywords Alzheimer's dementia, Alzheimer's disease, Dementia subtyping, Diagnosis, Disease progression, Machine learning

Introduction

Alzheimer's disease (AD), the major cause of dementia, is a progressive neurodegenerative disorder that predominantly affects older people [1]. The accumulation of amyloid-beta (Aβ) and formation of neurofibrillary tangles marked by tau phosphorylation in the brain are the key hallmarks of AD [1]. Clinically, the disease can be divided into three stages: 1) preclinical AD i.e., cognitive unimpaired (CU) people with amyloid accumulation in the brain, 2) prodromal or mild cognitive impairment (MCI) and 3) Alzheimer's dementia (ADem) [1]. This disease trajectory can vary between individuals, and preclinical AD can occur 15–20 years prior to ADem [1].

Observational longitudinal dementia datasets have been collected in diverse age groups across several (inter)national dementia cohorts (Table 1), providing rich information that enhances the granularity and scope of data science research. These datasets encompass a broad spectrum of information including biomarkers, genetics, neuropsychological evaluations, neuroimaging, omics, etc. (Table 2). Traditional statistical methods, constrained by

*Correspondence:
Yijun Pan
yijun.pan@unimelb.edu.au
[1] The Florey Institute of Neuroscience and Mental Health, 30 Royal Parade, Parkville, VIC 3052, Australia
[2] Florey Department of Neuroscience and Mental Health, The University of Melbourne, 30 Royal Parade, Parkville, VIC 3052, Australia
[3] The ARC Training Centre in Cognitive Computing for Medical Technologies, The University of Melbourne, Carlton, VIC 3010, Australia
[4] Department of Aged Care, Austin Health, Heidelberg, VIC 3084, Australia
[5] Department of Medicine, Austin Health, University of Melbourne, Heidelberg, VIC 3084, Australia

Table 1 Major longitudinal datasets used in ML-dementia

	Dataset	Participants	Year	Region	Follow-ups	Age	Data type	Focus/Aim
1	Framingham Heart Study [6]	15,000+	1948-present	USA	Every 2–8 years (vary by cohort)	13–62 at baseline (vary by cohort)	Demographic; Genetics; Lifestyle; Medical history	Cardiovascular and chronic diseases
2	BLSA [7]	3,000+	1958-present	USA	Every 1–4 years (vary by age groups)	20–80+	Cognitive assessments; Demographic; Genetics; Lifestyle; Medical history	Aging
3	Rotterdam Study [8]	15,000+	1990-present	Netherlands	Variable	45+	Biomarkers; Clinical; Cognitive assessments; Demographic; Genetic; Lifestyle	Cardiovascular, neurological, ophthalmological and endocrine diseases
4	ROSMAP [9]	3,000+	1997-present	USA	Every 12 months	65+	Cognitive assessments; Demographic; Genetics; Metabolomics; Imaging; Physical activity; Proteomics	Aging, Alzheimer's disease
5	CLHLS [10]	80,000+	1998-present	China	Variable	65+	Biomarkers; Clinical; Cognitive assessments; Demographic; Lifestyle; Physical activity	Healthy longevity
6	NACC [11]	40,000+	1999-present	USA	Every 12 months	19+	Biomarker; Cognitive assessments; Demographic; Genetics; Imaging	Dementia
7	WRAP [12]	1,500+	2001-present	USA	Every 24 months	40–65 at baseline	Biomarker; Cognitive assessments; Demographic; Genetics; Lifestyle; Metabolomics; Imaging	Alzheimer's disease
8	ADNI [13]	2,000+	2004-present	USA and Canada	Vary by phase	55–95	Biomarker; Cognitive assessments; Demographic; Imaging; Omics	Alzheimer's disease
9	AIBL [14]	3,000+	2006-present	Australia	Every 12–18 months	50+	Biomarker; Cognitive assessments; Demographic; Imaging; Omics	Alzheimer's disease
10	UK Biobank [15]	500,000+	2006-present	UK	Variable	40–69 at baseline	Cognitive assessments; Demographic; Genomics; Metabolomics; Imaging; Physical activity	Broad medical and health research
11	OASIS [16]	2000+	2007-present	USA	Vary by sub-project	18–96	Biomarker; Clinical; Cognitive assessments; Imaging	Accessible imaging data
12	CLSA [17]	50,000+	2008-present	Canada	Variable	45–85	Cognitive assessments; Demographic; Lifestyle; Medical History	Aging
13	BNA [18]	30,000+	2009-present	France	Variable	Mean age of 76	Clinical; Cognitive Assessments; Demographic	Alzheimer's disease

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 3 of 21

Table 1 (continued)

	Dataset	Participants	Year	Region	Follow-ups	Age	Data type	Focus/Aim
14	TILDA [19]	8,000+	2009-present	Ireland	Every 24 months	50+	Biomarker; Cognitive assessments; Demographic; Lifestyle; Social Engagement;	Aging
15	ALFA [20]	2,500+	2013-present	Spain	Variable	45–75 at baseline	Biomarker; Cognitive assessments; Demographic; Genomics; Lifestyle; Imaging;	Early pathophysiology of Alzheimer's disease
16	EMIF-AD MBD [21]	1000+	2013-2018	Several European countries	Variable	50+ at baseline	Biomarker; Clinical; Demographic; Genomics; Metabolomics; Imaging; Proteomics;	Alzheimer's disease
17	A4 [22]	5,000+	2014-2020	USA	Variable	65–85	Clinical; Cognitive assessments; Demographic; Imaging;	Amyloid-related memory problems

This is not an exhaustive list of all datasets available for AD research globally. The datasets were selected based on the following criteria: 1) open-access observational datasets, 2) more than 1,000 participants, and 3) prominent use in ML-dementia models developed over the last 10 years. The list is ordered by the year in which the cohort study began

Abbreviations: A4 Anti-Amyloid Treatment in Asymptomatic Alzheimer's study, *ADNI* Alzheimer's Disease Neuroimaging Initiative, *AIBL* The Australia Imaging, Biomarker and Lifestyle Study, *ALFA* Alzheimer and Families project, *BLSA* The Baltimore Longitudinal Study of Aging, *BNA* The French National Alzheimer Database, *CLHLS* Chinese Longitudinal Healthy Longevity Survey, *CLSA* The Canadian Longitudinal Study on Aging, *EMIF-AD MBD* European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery, *NACC* The National Alzheimer's Coordinating Center, *OASIS* Open Access Series of Imaging Studies, *ROSMAP* Religious Orders Study and Rush Memory and Aging Project, *TILDA* The Irish Longitudinal Study on Ageing, *WRAP* Wisconsin Registry for Alzheimer's Prevention

Table 2 Types of data commonly used in ML-dementia

Input data Categories	Assessments/ Techniques	Cost	Invasiveness	Targets	Datasets[b]
Biopsy	Brain tissue biopsy	AU$1,601.8[a]	High	Amyloid plaques, neurofibrillary tangles, etc	-
Blood/CSF biomarker	Lumbar puncture	AU $109[a]	High	CSF-Aβ & CSF-tau [23], neurofilament light chain, etc	2, 3, 4, 5, 6, 7, 8, 9, 11, 17
	Blood testing (e.g., HRMS [24])	US$200—$500 [25]	Minimum	Neurofilament light chain; plasma $A\beta_{42}$/ $A\beta_{40}$ [26], p-tau 181, p-tau231, and p-tau217 [27], etc	2, 3, 5, 8, 9,15, 16
Omics (blood/serum, urine, saliva)	Genomics	US$990 for Alzheimer's disease, familial, plus APOE panel [28]	None/minimum	APP, APOE, PSEN1, PSEN2, etc	1, 2, 3, 4, 5,7, 10, 15
	Metabolomics	US$100—US$500 [29]	None/minimum	Amino acids, carbohydrates, fatty acids, etc	2, 4, 7,8, 10, 15
	Proteomics	$132 [30]	None/minimum	GFAP [31] and LTBP2 [32] etc	2, 4, 8, 15
	Transcriptomics	AU$200 [33]	None/minimum	MicroRNAs, mRNA levels, etc	2, 5, 10, 15
Cognitive assessments	Neuropsychological evaluation	Vary by healthcare system	None	ADAS-Cog, CDR, MMSE, etc	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17
	Questionnaire	Labor cost	None	ECOG and FAQ, etc	
Demographics or clinical data from EHR	Census, observational cohort study, clinical history	Labor cost	None	Age, economic status, education, gender, lifestyle, medical history, race, etc	1, 2, 3, 4, 5, 6,7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
Imaging	DTI	US$97 [34]	None	White matter integrity, brain network connectivity, etc.	3, 6, 7, 11
	Retinal imaging	AU$55[a]	None	Network complexity, tortuosity, vessel calibers, etc	14
	CT	AU$342.95[a]	None	Structural images	1
	MRI	AU$426.5[a]	None	Functional MRI: tracking alterations in blood flow linked to neural activity [35] structural MRI: brain anatomy images [36]	3, 4, 6, 8, 11, 15, 17
	PET	AU$605.05[a]	None/moderate	Aβ-PET, FDG-PET, tau-PET [37], etc	2, 6, 7, 8, 9, 11, 17
Physiological monitoring	EEG	AU$358.45[a]	None	Brain waves, electrical activity of the brain, etc	1, 17
	ECG	AU$167.55[a]	None	Electrical activity of the heart, etc	1, 2, 10, 17
Speech and language [38]	Microphones & recording	-	None	Acoustic, prosodic, etc	-

Abbreviations: *Aβ* amyloid-beta, *ADAS-Cog* Alzheimer's Disease Assessment Scale—Cognitive Subscale, *APOE* apolipoprotein E, *APP* amyloid precursor protein, *CDR* Clinical Dementia Rating, *CSF* cerebrospinal fluid, *CT* computed tomography, *DTI* diffusion tensor image, *ECG* electrocardiogram, *ECOG* Everyday Cognition Scale, *EEG* electroencephalogram, *EHR* electronic health records, *FAQ* Functional Activities Questionnaire, *FDG-PET* fluorodeoxyglucose positron emission tomography, *GDF15* growth differentiation factor 15, *GFAP* glial fibrillary acidic protein, *HMRS* high-resolution mass spectrometry, *LTBP2* latent transforming growth factor beta binding protein 2, *MMSE* Mini-Mental State Examination, *MRI* magnetic resonance imaging, *PSEN1* presenilin 1, *PSEN2* presenilin 2, *PET* positron emission tomography

[a] prices were obtained from Australia Medicare Benefits Schedule website in June 2024 [39]

[b] refer to Table 1 for dataset number

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 5 of 21

rigid assumptions and a limited ability to handle complex interactions have shown limitations in processing these multi-modal datasets, prompting an exploration of more adaptive and comprehensive techniques such as machine learning (ML) [2]. ML is a class of algorithms that enable computers to analyze data and make decisions by identifying patterns specific to tasks [3]. These techniques can detect subtle patterns and trends in large datasets, significantly enhancing the effectiveness and productivity of data-driven research. In addition, ML has already proven successful in tracking disease, including market-ready products (e.g., Vivid E80 [4]) and FDA-approved devices (e.g., Apple's Atrial Fibrillation History Feature [5]).

The development of anti-Aβ monoclonal antibodies, such as donanemab [40] and lecanemab [41], has shown promising results in reducing cognitive decline in early treatment scenarios. This underscores the importance of timely intervention. ML can enhance early detection accuracy and personalized stimulation by determining the most effective timepoint to adminster antibodies in the right patients, thereby maximizing their therapeutic benefits. However, it must be noted that while ML can aid in identifying individuals likely to benefit, our global health systems are not fully equipped to provide these early interventions. Monoclonal antibodies require costly monitoring for brain bleeds, which presents challenges not only in funding the necessary scans but also in accessing scanners within a reasonable distance for patients. A recent study showed that novel biomarkers including microRNAs, metabolites and proteins have been identified using ML approaches [42]. Furthermore, it has been demonstrated that patient-level simulations by ML can predict disease trajectories [43], estimate the likelihood of transitioning from MCI to ADem [44] or even successfully forecast the time-to-event outcomes survival probability for MCI participants [45].

Here we provide a comprehensive overview of ML application in dementia (ML-dementia) using non-technical terms to enhance accessibility to a broad readership. Specifically, we evaluate ML from a historical perspective and discuss typical workflows, successful applications within 5 years and challenges—highlighting the evolving utility of ML in biomedical research to enhance diagnosis and management of dementia.

Machine learning
Types of ML
ML includes a variety of algorithms designed to learn from data to meet a predefined goal, such as identifying patterns or making predictions about future states. The model updates its settings or '(hyper-)parameters' based on feedback from performance metrics known as 'loss functions', which assesses the accuracy of the model's predictions compared to actual outcomes. Once the model is optimally trained, it can use real-world data to achieve the predefined task [46]. ML techniques are primarily divided into three categories: unsupervised learning, supervised learning, and reinforcement learning, with the first two being more commonly used in dementia research. These categories are discussed in detail below and their advantages and limitations are summarized in Table 3**.**

Supervised learning
Supervised learning explores the relationship between input features and the corresponding target outputs, also known as labels. In dementia research, supervised learning can be further categorized based on the predictive target, for instance, classification tasks dealing with categorical labels (e.g., ADem vs CU), regression tasks handling numerical labels (e.g., Clinical Dementia Rating—Sum of Boxes [CDR-SB] and Mini-Mental State Examination [MMSE]). Once the model is trained, it can then make predictions on unlabelled data of the same input.

Unsupervised learning
Unsupervised learning operates on unlabelled data, which focuses on uncovering patterns or relationships without considering any predefined labels. This approach includes 1) clustering tasks such as identifying subtypes of dementia based on biological, neuropsychological, and demographic features and 2) data compression such as using principal component analysis to simplify and summarize complex data.

Reinforcement learning
Reinforcement learning (RL) is used to learn and improve decision making by continuously receiving feedback through interaction with external conditions and observing the response. This approach is less commonly used than the supervised and unsupervised methods. RL can be classified as model-free and model-based types; model-free RL operates without a predefined model, while model-based RL is preferred for incorporating domain knowledge (i.e., existing clinical knowledge). RL could mainly be employed to simulate and predict cognitive states, as well as to estimate the probability of transitioning between cognitive states.

Statistical analysis versus ML approaches
Traditional statistical methods include a hypothesis-driven approach and statical inference (i.e., generalizing findings from a subset of data to a large population). Such approach relies on strong assumptions about the data, e.g., the data follows a normal distribution to fit existing

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 6 of 21

Table 3 Examples of machine learning models

Subtypes	Model name	Advantages	Limitations
Supervised learning [47]			
Linear models	Linear regression; Logistic regression; Lasso; Ridge; Elastic net, etc	Simple; Computationally efficient; Suitable for linear relationship	Sensitive to outliers
Non-linear models	Support vector machine; k-nearest neighbors	Can capture features with non-linear relationships	Computationally intensive; Possible to overfit
Tree-based models	Random forest; Decision trees; eXtreme gradient boosting	Robust to outliers; Can capture non-linear relationships and interactions between features	Possible to overfit; Less interpretable
Probabilistic models	Naive bayes; Bayesian linear regression	Can be used on small datasets; Robust to noise	Assumption of independence between features; Prior knowledge in data distribution must be known
Deep learning models	Multilayer perceptron; Convolutional neural networks; Recurrent neural networks; Transformers; Attention	Capable of learning complex patterns and relationships; Good for image and speech recognition	Requires large amounts of data; Computationally intensive; Lack of interpretability
Unsupervised learning [48]			
Clustering	k-Means; Hierarchical clustering	Can discover patterns and groups for unlabelled data	Sensitive to scale of data; Assumes certain data distributions
Dimensionality reduction	Principal component analysis; t-SNE	Reduces data complexity	Lack of interpretability
Reinforcement learning [49]			
Model-based	Dyna-Q; Monte carlo tree search	Efficiency	Suitable only for simpler environments; Effectiveness highly depends on the accuracy of the model
Model-free (Value-based)	Q-Learning; State-Action-Reward-State-Action	Flexible; Suitable for more dynamic environment; Less sensitive to the initial conditions and hyperparameters	Data and computationally intensive; May not handle stochastic environments effectively
Model-free (Policy-based)	REINFORCE; Deterministic Policy Gradient	Flexible; Supports both discrete and continuous action spaces	Data and computationally intensive; Slow to converge; High variance

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 7 of 21

theoretical models [50]. However, these traditional statistical methods often encounter practical challenges in complex real-world scenarios, as the assumptions made may not be satisfied in clinical practice [2]. In contrast, ML adopts a more data-driven approach with minimal assumptions, and it concentrates on prediction rather than inference [2]. However, statistical models and ML techniques sometimes overlap; e.g., both methods often employ linear and logistic regression models to meet statistical goals or to achieve simple linear predictions in ML contexts. It must be noted that ML possesses the capability to process and analyze extensive and complex datasets, such as omics data, effectively uncovering patterns or capturing interactions that might be omitted or overlooked by the traditional statistical analysis [2]. Therefore, ML is often beneficial to clinical research, where data is inherently multidimensional with a diverse array of variables.

The history and typical workflow of ML techniques in dementia research and clinical applications

Prior to the year 2000, research primarily focused on clarifying the genetic and biochemical foundations of AD, with significant emphasis on the roles of Aβ and familial genetic mutations [51]. In the subsequent decade (2000–2010), scholarly attention shifted towards differentiating AD from CU mostly using ML model such as support vector machines alongside brain imaging techniques [52]. In the following five years or so, researchers focused on predicting clinical progress in MCI patients using multi-kernel support vector machine (SVM, a ML model) with longitudinal data from magnetic resonance imaging (MRI) and positron emission tomography (PET) [53].

Since then, ML or deep learning, a subset of ML that uses neural network to simulate the learning process of human [54], has been used to classify disease subtypes and stages. Similar to how the human brain employs interconnected neurons for information processing, neural networks in ML use nodes (artificial neurons) and their interconnections to mimic the brain's structure and functionality. This design facilitates pattern recognition and decision-making. For instance, Ramzan et al. [55] utilizes resting-state function MRI with Residual Network architecture to classify AD into: CU, significant memory concern, early-MCI, MCI, late-MCI, and ADem. In more recent years, the adoption of advanced deep learning architectures, such as time-series models has expanded. For example, hybrid deep learning frameworks based on Bidirectional Long Short-Term Memory models leverage multimodal data (i.e., MRI, PET, and neuropsychological evaluation) to enhance the classification of CU and early MCI [56]. A timeline summarizing the use of ML in dementia research is presented in Fig. 1.

The general workflow to build and apply the ML-dementia model is summarized in Fig. 2, which can be separated into six key steps, including 1) Intended application, 2) Data selection, 3) Data pre-processing, 4) Model Construction, 5) Model evaluation, and 6)
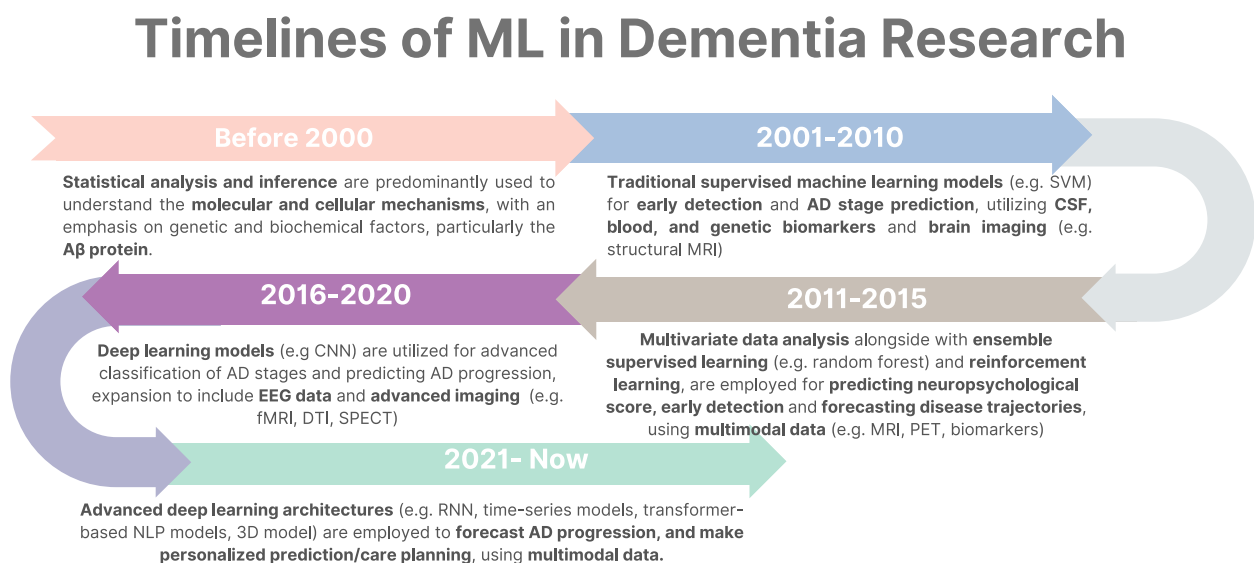
Fig. 1 Timelines of ML in dementia research. Aβ = amyloid-beta; AD = Alzheimer's dementia; CNN = convolutional neural network; CSF = cerebrospinal fluid; DTI = diffusion tensor image; EEG = electroencephalogram; fMRI = functional magnetic resonance imaging; MRI = magnetic resonance imaging; NLP = natural language processing; PET = positron emission tomography; RNN = recurrent neural network; SPECT = single-photon emission computed tomography; SVM = support vector machine. This figure is created using Canva (www.canva.com)

Machine Learning Workflows in Clinical Settings
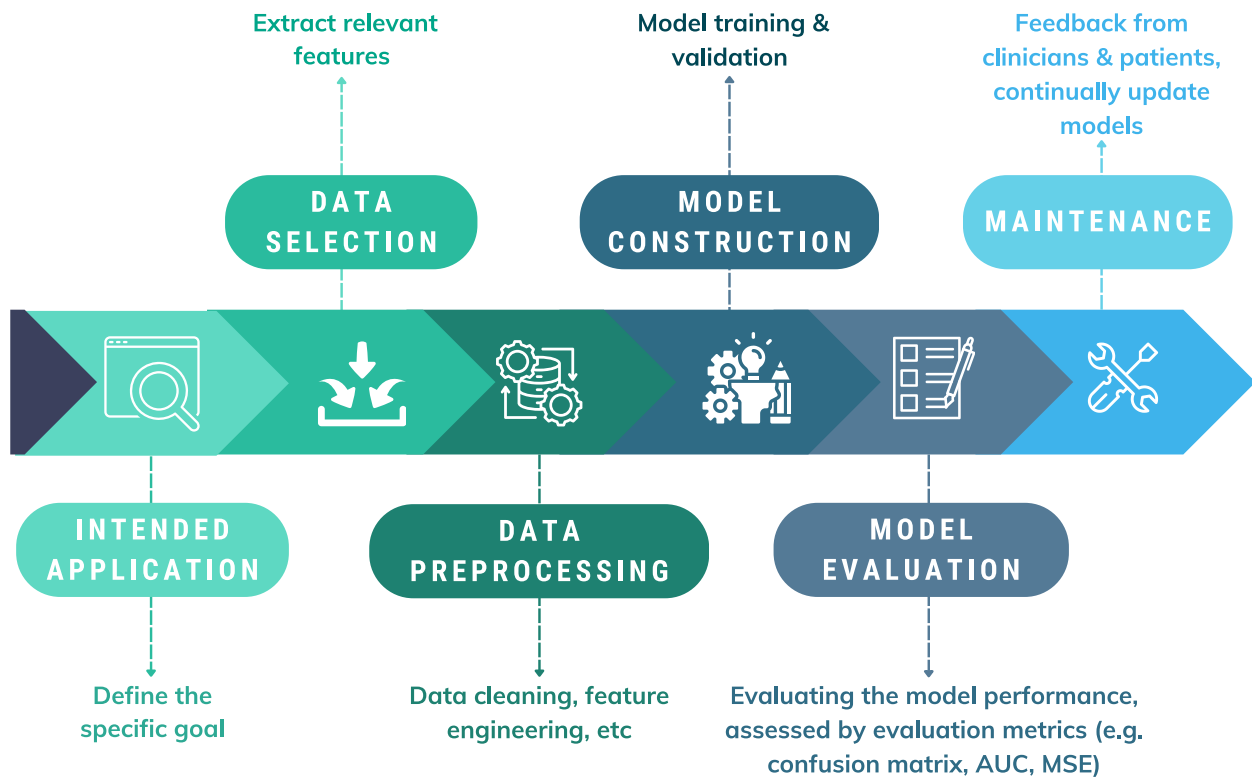
Fig. 2 General machine learning model workflows in clinical settings. AUC = area under the curve; MSE = mean squared error. This figure is created using Canva (www.canva.com)

Maintenance. We have provided a detailed description for each step in Supplementary Material – ML workflow.

Data used in ML-dementia studies

Several observational dementia datasets have been used for ML model construction and validation (Table 1), such as the Australian Imaging, Biomarker and Lifestyle (AIBL) study [57] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) study [13]. These datasets are often longitudinal, involving thousands of participants, spanning several decades with regular follow-ups, and some are still actively recruiting. These datasets feature a diverse range of participant demographics, typically focusing on middle-aged adults from various racial, ethnic and educational backgrounds. Each dataset has a distinct focus. For instance, Open Access Series of Imaging Studies [OASIS] [16] concentrate on brain imaging, while the Religious Orders Study and Rush Memory and Aging Project [ROSMAP] [9] aim to understand aging processes. Data collection and testing within the same dataset can vary depending on the project's phases or aims. For example, ADNI adapts its data collection strategies across five phases, and OASIS divides its datasets

to address specific research goals. While most datasets listed in Table 1 primarily address AD, others such as the UK Biobank [15] and the Framingham Heart Study [6], provide a broader insight across various health outcomes within larger cohorts.

A variety of data/sample collection methods have been employed in these studies, which can be categorized as per their level of invasiveness (Table 2). Invasive methods, such as cerebrospinal fluid collection through lumbar puncture, are commonly used to obtain biomarkers (Aβ and tau) and markers of neurodegeneration [1]. The AT(N) 2018 framework [58], categorizes the progression of AD into different stages based on specific combinations of these biomarkers (Table 4). Compared to lumbar puncture, venous blood collection is considered less-invasive, and often used for biomarker research and omics (genomics, transcriptomics, proteomics, and metabolomics) analysis [59]. Non-invasive methods such as MRI and PET are employed to study brain structure and Aβ levels [1]. Neuropsychological evaluation (Table 5) are also non-invasive, which are quantitative measures of cognitive

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 9 of 21

Table 4 2018 NIA-AA research framework [58] for biological definition of Alzheimer's disease

A	T	N	CU (Cognitively unimpaired)	MCI (Mild Cognitive Impairment)	Dementia
-	-	-	Cognitively unimpaired	MCI not caused by Alzheimer's disease	Dementia not caused by Alzheimer's disease
		+			
	+	-			
		+			
+	-	-	Preclinical Alzheimer's disease	Prodromal Alzheimer's disease	Alzheimer's dementia
		+			
	+	-		Alzheimer's pathologic change with MCI	Alzheimer's pathologic change with dementia
		+		MCI suspects not caused by Alzheimer's disease	Dementia suspects not caused by Alzheimer's disease

A: amyloid-beta levels detected by PET or cerebrospinal fluid analysis

T: tau pathology evidenced by tangles and PET or cerebrospinal fluid biomarkers

N: neurodegeneration indicated by MRI atrophy, ^{18}F-Fluorodeoxyglucose—PET hypometabolism, or high cerebrospinal fluid tau

Table 5 Examples of neuropsychological tests for dementia research or clinical diagnosis

Method	Domain	Name of assessments
Questionnaire	Global functioning and behavior	Everyday Cognition (ECOG), Functional Assessment Questionnaire (FAQ), etc
	Psychiatrics	Geriatric Depression Scale (GDS), Neuropsychiatric Inventory (NPI), etc
Neuropsychological evaluation	Global functioning and behavior	Clinical Dementia Rating-Sum of Boxes (CDR), Mini Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), etc
	Test battery	Alzheimer's Disease Assessment Scale (ADAS), Cogstate Brief Battery (CBB), etc
	Language	Boston Naming Test (BNT), Rey Auditory Verbal Learning Test (RAVLT), etc
	Memory	Rey Auditory Verbal Learning test (RAVLT); Logical memory IIA Delayed (LOGIMEM), etc
	Visuospatial ability	Clock Drawing Test, Hooper Visual Organization Test, etc
	Recognition and processing speed	Face Recognition Tests, Benton Visual Retention Test, etc
	Executive functioning	Wisconsin Card Sorting Test, Stroop Test, Trail Making Test (Part B), etc
	Emotional and personality assessment	Minnesota Multiphasic Personality Inventory, Beck Depression Inventory, etc

Table 6 CDR-SB and MMSE scores for cognitive health classification

Cognitive health	Substage	CDR-SB Score (0–18) [62]	MMSE Score (0–30) [63]
CU (cognitive unimpairment)	-	0	30
MCI (mild cognitive impairment)	Questionable impairment	0.5–2.5	26–29
	Very mild dementia	3.0–4.0	
Alzheimer's dementia	Mild dementia	4.5–9.0	21–25
	Moderate dementia	9.5–15.5	11–20
	Severe dementia	16.0–18.0	0–10

CDR-SB Clinical Dementia Rating-Sum of Boxes, *MMSE* Mini-Mental State Examination

functions across various disease stages (Table 6) [60]. Demographic information, lifestyle data and medical history are often self-reported or collected using questionnaires and are used as baseline predictors in the majority of studies [61].

Existing ML-dementia models using non/less-invasive data

The following section reviews ML models using input data collected via non-/moderately invasive approaches. These data include demographics (age, gender, ethnicity, family history), medical history, neuropsychological

Wang *et al. Alzheimer's Research & Therapy* (2024) 16:175

Page 10 of 21

evaluation, blood (omics, biomarkers), and brain imaging. Studies published between 2019 and 2024 were selected based on uniqueness in methodology, which is summarized in Table 7 and Fig. 3.

Dementia subtyping

AD is the major cause of dementia, followed by vascular dementia, frontotemporal dementia, and dementia with Lewy bodies [90]. Accurate differential diagnosis is important for clinicians to offer the most suitable care options to the patients [91]. Recent studies utilizing ML and deep learning models have shown relative high accuracy in differential diagnoses by incorporating metabolomics [67] and neuroimaging [64–66] (Table 7A). For instance, Qiang et al. [67] established the associations between 249 metabolites and type of dementia (all-cause dementia, ADem, and vascular dementia) using UK Biobank data. The study employed Cox proportional hazard models and light gradient boosting machine algorithms to generate a metabolic risk score. This score when combined with demographic and neuropsychological test scores achieved an AUC of 0.85 (AUC approaching 1 indicates excellence in discrimination) for the classification of different types of dementia. By employing neuroimaging data, Castellazzi et al. [92] used the adaptive neuro-fuzzy inference systems to distinguish between ADem and vascular dementia. This achieved over 84% accuracy using a combination of features from resting-state functional MRI and diffusion tensor imaging. Moreover, another independent research group [65] achieved ~80% accuracy in differentiating dementia with Lewy bodies from ADem using structural MRI data and a residual neural network. Finally, Nguyen et al. [66] introduced an innovative approach, by integrating 3D U-Nets with a multi-layer perceptron classifier to discern ADem from frontotemporal dementia through structural MRI images, attaining an AUC of 0.94.

Although these studies achieved high diagnostic accuracies (~80%), only Nguyen et al. [66] validated their model using an external dataset. This raises concerns about the generalizability of these findings and suggests that potential cohort bias cannot be ruled out. It is crucial to further validate these models prior to clinical trial and implementation. Moreover, these studies appear to focus on the differential diagnosis between vascular dementia and ADem (Qiang et al. [67] and Catellazzi et al. [92]) and between frontotemporal dementia and ADem (Nguyen et al. [66]). Future research could explore the possibility of differentiating multiple subtypes of dementia using a single model. Furthermore, all these studies, except Qiang et al. [67], leveraged advanced imaging techniques to capture intricate details of the brain. The reliance on high-resolution imaging data necessitates substantial resources, making it challenging to implement the new technology in clinics.

Disease staging

Predicting disease stages using either a binary classification (CU vs ADem, CU vs MCI+ADem, CU vs MCI, MCI vs ADem) or CU/MCI/ADem classification is commonly used in ML-dementia. These typically employ omics data [69, 74], neuropsychological evaluation [70], and neuroimaging [68, 70, 71] (Table 7B). Mahendran et al. [74] demonstrated that deep belief network-based approach (accuracy 82%) outperformed SVM (accuracy 78%) and Naïve Bayes (accuracy 76%) in binary classification of CU and ADem using their multi-omics data. In another study, Wang et al. [69] utilized six differentially expressed metabolites, three metabolic pathways and a random forest model to differentiate the MCI+ADem group from CU, and they achieved an AUC of 0.77. MRI data have also been employed to facilitate disease classification. For instance, Naz et al. utilized only structural MRI data [71], and achieved a classification accuracy of 99.27, 98.89 and 97.06% for MCI/ADem, ADem/CU, and MCI/CU, respectively. To generate more complex models, multimodal data (e.g., demographic, medical history, brain volume, neuropsychological evaluation and genetics) have been integrated, such as convolutional neural network model for disease stage classification. For example, using multimodality, Venugopalan et al. [70] achieved a classification accuracy of 83% for CU, 74% for MCI and 85% for ADem.

We noted that model development in most of these studies were challenged by an imbalanced dataset, with AD and MCI often being underrepresented compared to CU individuals due to disease prevalence. Interestingly, Naz et al. [71] manually balanced the dataset by eliminating some of the CU participant data (CU=95, MCI=146, ADem=95). However, this approach reduces the overall dataset size, possibly leading to the model not capturing all critical features for accurate classification [93]. Model overfitting is also expected from using such a small dataset [94]. Future studies could focus on enriching AD and MCI participant data; however, this is currently less practical due to a lack of harmonized datasets that allows data pooling. An alternative approach is to intentionally recruit MCI and ADem participants, as done by Kwak et al. [77]; however, these data may be less suitable for studying the onset and progression of AD. Another major issue is that the classification accuracy is usually less satisfactory for differentiating MCI from AD, as has been reported by Wang et al. [69] and Naz et al. [71]. Using multimodal data could be a potential solution [70], nonetheless, future studies are required to confirm whether their observations are dataset dependent.

Table 7 Dementia ML models using non/less-invasive data as input predictors

A. Dementia subtyping

Study	Participants	Algorithm(s)	Predictors	Findings/Results
Castellazzi et al. (2020) [64]	NINCDS2-ARDA criteria ($n_{ADem}=33$, $n_{VaD}=27$, $n=15$ for testing)	ANN, SVM, Adaptive Neuro-Fuzzy Inference System	Diffusion tensor imaging and MRI	ADem vs. VaD with AUC=0.853 using ANFIS algorithm; multimodal input feature set (e.g., DTI+rs-fMRI metrics) have better performance than a unimodal feature set
Nemoto et al. (2021) [65]	The Brain Functions Laboratory, Inc. ($n_{DLB}=101$, $n_{ADem}=69$, $n_{CU}=38$)	RNN	MRI	ADem vs. DLB with ACC=79.2%
Nguyen et al. (2023) [66]	ADNI2($n_{CU}=190$, $n_{ADem}=149$) NIFD($n_{CU}=136$, $n_{FTD}=150$) NACC($n_{CU}=2182$, $n_{ADem}=485$, $n_{FTD}=37$)	Ensemble of 3D U-Nets	MRI	ADem vs. FTD with AUC=0.938 evaluate in-domain, and AUC=0.916 evaluating out-of-domain
Qiang et al. (2024) [67]	UK biobank ($n_{total}=274,160$; $n_{ACD}=5274$; $n_{ADem}=2346$; $n_{VAD}=1221$)	CPH; LightGBM	Demographics, genetics, plasma metabolome	ACD/ADem/VaD with AUC=0.857 for ACD; AUC=0.861 for ADem; AUC=0.873 for VaD

B. Disease staging

Study	Participants	Algorithm(s)	Predictors	Findings/Results
Marzban et. al. (2020) [68]	ADNI ($n_{CU}=185$; $n_{MCI}=106$; $n_{ADem}=115$)	CNN	Diffusion tensor imaging and MRI	AUC=0.84 for MCI/CU; AUC=0.94 for ADem/CU
Wang et al. (2020) [69]	ROSMAP brain sample ($n_{CU}=51$, $n_{MCI}=31$; $n_{ADem}=37$) serum sample ($n_{CU}=446$; $n_{MCI/ADem}=120$) serum progression sample ($n_{CU}=356$; $n_{CUp}=90$)	RF	Plasma metabolome;	CU vs. MCI/ADem with AUC=0.772 for metabolite level model; health vs. MCI/ADem with AUC=0.731 for metabolomic pathway level model
Venugopalan et. al. (2021) [70]	ADNI ($n_{CU}=598$; $n_{MCI}=699$; $n_{ADem}=707$)	DT, RF, SVM, KNN for classification; Auto-encoder and 3D CNN for preprocessing	3D MRI images, cognitive assessments, demographics and genetic (single nucleotide polymorphisms)	Single modality for the best performance: $ACC_{CU}=83\pm7\%$; $ACC_{MCI}=74\pm6\%$; $ACC_{ADem}=85\pm3\%$ Combine three modalities for the best performance: $ACC_{CU}=88\pm2\%$; $ACC_{MCI}=80\pm2\%$; $ACC_{ADem}=87\pm2\%$
Naz et al. (2022) [71]	ADNI ($n_{CU}=95$, $n_{ADem}=95$, $n_{MCI}=146$)	AlexNet, GoogLeNet, VGG-16/19, ResNet-18/50/101, MobileNetV2, InceptionV3, Inception-ResNet-V2 and DenseNet201	MRI	MCI vs. ADem with AUC=0.993 CU vs. ADem with AUC=0.989 CU vs. MCI with AUC=0.970
Rye et. al. (2022) [72]	ADNI ($n_{MCIs}=357$, $n_{MCIc}=321$)	Ensemble based model; RF	Cognitive assessments, genetic APOE status, hippocampal volume	MCIs vs. MCIc using RF ACC=0.746
Hashmi & Barukab (2023) [73]	OASIS ($n_{CU}=639$, $n_{very-mild-ADem}=645$, $n_{mild-ADem}=662$, $n_{moderate-ADem}=624$)	Deep learning reinforcement learning for active learning; XGBoosting	MRI	CU: F1-score=0.82 Very-mild-ADem: F1-score=0.86 Mild-ADem: F1-score=0.90 Moderate-ADem: F1-score=0.94

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 12 of 21

Table 7 (continued)

| Study	Participants	Algorithm(s)	Predictors	Findings/Results
Mahendran et. al. (2023) [74]	Gene Expression Omnibus data under National Centre for Biotechnology Information ($n_{ADem}=439$, $n_{non-ADem}=257$) DNA methylation ($n_{ADem}=68$, $n_{non-ADem}=74$)	Deep Belief Network	Genomic, e.g., DNA methylation and microarray gene expression	ADem vs. non-ADem: ACC=82%

C. Disease progression/trajectory prediction

Study	Participants	Algorithm(s)	Predictors	Findings/Results
Beltrán et al. (2020) [75]	ADNI ($n=815$)	CART, RF, GB & SVM	Cognitive assessments, demographic, and plasma proteomics	MCIs vs. MCIc with AUC=0.71
Jiang et al. (2020) [76]	ADNI ($n_{MCIs}=165$; $n_{MCIc}=137$)	Functional ensemble survival tree	Cognitive assessments, demographic	Simulate the trajectory of cognitive assessments; AUC=0.847 for predicting the AD converter for MCI participants
Kwak et al. (2021) [77]	SMG-SNU Boramae Medical Center ($n_{CU}=285$; $n_{MCI}=1057$; $n_{ADem}=1300$)	LR; SVM	Cognitive assessments	MAE=0.46 for predicting CDR-SB value for MCI; MAE=0.39 for predicting CDR-SB for ADem;
Lian et. al. (2021) [78]	ADNI-1 for training ($n_{CU}=226$; $n_{MCIs}=225$; $n_{MCIc}=165$; $n_{AD}=181$) ADNI-2 for testing ($n_{CU}=185$; $n_{MCIs}=234$; $n_{MCIc}=37$; $n_{ADem}=143$)	Multitask weakly-supervised attention network	MRI	RMSE=1.501 for CDR-SB, RMSE=5.701 for ADAS-Cog; RMSE=2.244 for MMSE
Mofrad et. al. (2021) [79]	ADNI ($n_{MCIs}=333$, $n_{MCIc}=333$, $n_{MCIs}=333$, $n_{MCIc}=333$)	Ensemble based model with soft voting strategy	Cognitive assessments, MRI image	MCIs vs. MCIc: ACC=76±4% for cognitive assessments only; ACC=77±3.7% adding MRI features; CUs vs. CUc: ACC=56±6% for cognitive assessments only; ACC=61±5.7% adding MRI features;
Saboo et al. (2021) [80]	ADNI ($n_{CU}=52$; $n_{AD}=23$; $n_{eMCI}=58$; $n_{LMCI}=27$)	Reinforcement learning	Cognitive assessments, demographic, MRI and PET	MAE=0.537 for predicting MMSE 10-year cognition progression trajectory
Mukherji et. al. (2022) [81]	ADNI ($n_{CU}=200$, $n_{MCI}=400$, $n_{ADem}=200$)	RNN; LSTM; MLP	Cognitive assessments	MCIs vs. MCIc with ACC=77.88% for predicted cognitive states 3-year after
Bucholc et al. (2023) [82]	NACC-UDS (2-year model: $n_{MCIs}=656$; $n_{MCIp}=112$; 3-year model: $n_{MCIs}=656$; $n_{MCIp}=65$;)	LR, SVM, RF	Cognitive assessments	Predict several cognitive assessment score with best performance of ACC=87.5%
Zou et al. (2023) [83]	ADNI ($n_{MCIs}=601$; $n_{MCIp}=330$) NACC ($n_{MCIs}=1742$; $n_{MCIp}=759$)	Multivariate functional mixed model framework	Cognitive assessments, demographic, and genetic APOE status	Simulate the trajectory of cognitive assessments and the time to dementia onset; integrated AUC=0.839 for instantaneous model landmark times at 3

D. Prediction of the brain levels of amyloid-beta and tau

Study	Participants	Algorithm(s)	Predictors	Findings/Results

Table 7 (continued)

Study	Dataset	ML method	Features	Results
Palmqvis, et al. (2019) [84]	Train: BioFINDER (n=346) Validation: ADNI (n=170) Test: ADNI (n=661)	LASSO, Logistic Regression	Cognitive assessments, demographic, genetics and plasma proteomics;	Predict $A\beta+$ in PET or CSF: Training: AUC=0.85 with delayed recall mode+plasma $A\beta_{42}/A\beta_{40}$ Testing: AUC=0.83 with delayed recall mode+plasma $A\beta_{42}/A\beta_{40}$
Langford et al. (2020) [85]	A4 (n=1323 for $A\beta+$; n=3163 for $A\beta-$)	XGBoost	Cognitive assessments, demographic and genetic APOE status	Predict $A\beta+$ in PET: AUC=0.6 for a web-based battery; AUC=0.74 with APOE and cognitive assessments
Shan et al. (2021) [86]	ADNI (n_{SMC}=170; n_{eMCI}=317; n_{LMCI}=236)	KNN, DT, SVM, RF	Clinical, cognitive assessments, demographic and genetic APOE status	Predict $A\beta+$ in PET for 6 separate groups with the best ACC=90.4% for the SMC male group
Janelidze et al. (2021) [87]	BioFINDER (n_{CU}=182; n_{MCI}=104) ADNI (n=59 for $A\beta+$; n=63 for $A\beta-$)	Logistic regression	Plasma biomarker	Predict $A\beta$ status: AUC=0.86 in CSF; AUC=0.83 in PET With immunoprecipitation-coupled mass spectrometry developed at Washington University achieve the best in all 8 assays
Lew et. al. (2023) [88]	ADNI (n=1027 for amyloid; n=375 for Tau; n=1239 for fluorodeoxyglucose)	logistic regression for binary classification	Cognitive assessment, demographic, genetic APOE status and MRI	AUC=0.79 for amyloid-beta; AUC=0.73 for tau; AUC=0.86 for neurodegeneration
Zhang et. al. (2022) [89]	European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery (n_{CU}=311; n_{MCI}=184; n_{AD}=386)	7-layer neural network	Demographic, genetic APOE status and proteomics	AUC=0.782 for Amyloid, AUC=0.674 for p-tau, and AUC=0.734 for t-tau. AUC=0.831 for A+T+N+ vs. A-T-N-

Abbreviations: A4 Anti-Amyloid Treatment in Asymptomatic Alzheimer's disease, *ADNI* Alzheimer's disease Neuroimaging Initiative, *NACC-UDS* National Alzheimer's coordinating center uniform data set, *NIFD* National institute of Fashion Designing, *NINCDS-ADRDA* National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association, *ROSMAP* Religious Orders Study and Rush memory and Ageing; *SMG-SNU* Seoul National University – Seoul National University, *ACC* accuracy, *Aβ+* amyloid-beta positive, *Aβ-* amyloid-beta negative, *ACD* all-cause dementia, *ADem* Alzheimer's dementia, *AFT* accelerated failure time model, *ANN* artificial neural network, *APOE* apolipoprotein E, *AUC* area under the receiver operating characteristic curve, *CART* category and regression tree, *CPH* cox proportional hazard regression model, *CSF* cerebrospinal fluid, *CU* cognitive unimpairment, *CUs* cognitive unimpairment stable (i.e., no progression to MCI or AD), *Cup* cognitive unimpairment progression (i.e, progression from CU to MCI/AD), *DLB* dementia with Lewy bodies, *DT* decision tree, *eMCI* early mild cognitive impairment, *GB* gradient boosting, *KNN* k-nearest neighbour, *LASSO* least absolute shrinkage and selection operator, *LightGBM* light gradient boosting machine, *LMCI* late mild cognitive impairment, *LR* linear regression, *LSTM* long short-term memory, *MAE* mean absolute error, *MCI* mild cognitive impairments, *MCIs* mild cognitive impairment stable, *MCIp* mild cognitive impairments progression to Alzheimer's dementia, *MLP* multilayer perceptron, *MMSE* Mini Mental State Examination, *PET* positron emission tomography, *RF* random forest, *RNN* recurrent neural networks, *SMC* significant memory concern, *SVM* support vector machine, *VaD* vascular dementia, *XGBoosting* eXtreme gradient boosting

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 14 of 21
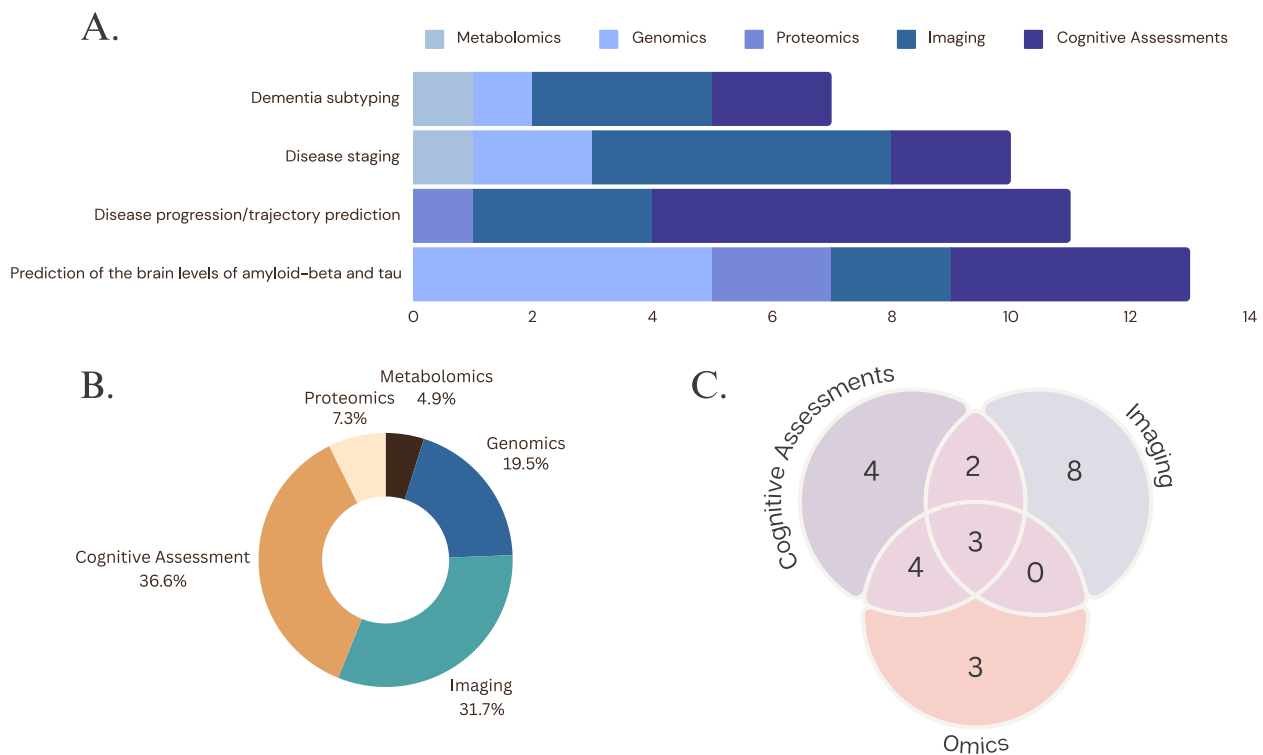
Fig. 3 Types of data used in ML models. **A** Counts of various data types used in four major ML-dementia applications; **B** Donut chart showing the distribution of data types used in the selected studies (Table 7); **C** Venn diagram illustrating the overlap of data types used in the selected studies shown in Table 7. This figure is created using Canva (www.canva.com)

Disease progression/trajectory prediction

The prediction of future disease states or neuropsychological outcomes can be achieved using classification and regression models, as well as simulating disease trajectories using more complex deep learning models (Table 7C). Most classification models categorize MCI-to-dementia progressors and non-progressors. For example, Rye et.al. [72] achieved a 75% of accuracy in predicting whether MCI participants progress to dementia using a random forest model, where neuropsychological evaluation, hippocampal volume and Apolipoprotein E (APOE) genotype were used as input features. An ensemble model was employed by Mofrad et al. [79] for such prediction, where MRI and neuropsychological evaluation were used to achieve a 77% accuracy. Regression models often employ neuropsychological evaluation, such as CDR-SB, ADAS-Cog, and MMSE [77, 78, 82], to estimate disease severity over time. For example, Lian et al. [78] employed a multitask weakly-supervised Attention Network, which is a regression model that built on structural MRI data collected from CU, MCI progressor, MCI non-progressor, and ADem participants to predict 3-year future CDR-SB, ADAS-Cog, and MMSE scores. This model has

achieved promising results, with a root-mean-squared error of 1.5, 5.7, and 2.2 for each score, respectively.

For disease trajectory simulation, Bucholc et al. [82] has combined unsupervised and supervised learning techniques, where participants were categorized by their cognitive score trajectories (stable vs deterioration over 2–3 years). The trajectories of each category were then analyzed using random forest, support vector machine, and linear regression (supervised). This approach achieved a ~90% accuracy in predicting seven different neuropsychological test scores over 1-year and 2-year intervals, from the correspondent baseline scores. A more complex model, Long Short-Term Memory Recurrent Neural Networks, was used by Mukherji et al. [81] to simulate the trajectory for five neuropsychological tests. This model achieved a prediction accuracy of 85 and 83% for 2-year and 4-year, respectively. Recent work has also focused on dynamically predicting the risk of dementia onset. This is typically achieved using a Cox model, combined with functional data analysis to model longitudinal neuropsychological outcomes. For example, Jiang et al. [76] utilized the functional ensemble random survival forest to characterize the joint effects of neuropsychological evaluation in predicting disease progression,

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 15 of 21

specifically to predict the time to AD conversion in individuals with MCI and to provide personalized dynamic predictions. This approach achieved an AUC of approximately 0.90 over an average follow-up period of 31 months. Similarly, Zou et al. [83] proposed a multivariate functional mixed model framework to simultaneously model multiple longitudinal neuropsychological outcomes and the time to dementia onset, achieving an integrated AUC of over 0.80, with the mean time to visit being 1.12 years.

Mukherji et al. [81], Bucholc et al. [82] and Lian et al. [78] predict disease progression over a fixed interval, while Jiang et al. [76] and Zou et al. [83] simulate disease progression. It should be noted that simulation methods introduce higher variance and complexity compared to fixed interval models [95]; however, they can predict disease status at any time point, whereas fixed interval models can only predict disease status at the end of the interval. Different models may suit varying clinical needs or patient expectations, each balancing its own advantages and limitations. In addition, these complex models are prone to overfitting [94], capturing noise that does not generalize to unseen data. This issue could be exacerbated in studies where the training datasets are relatively small, such as that for Jiang et al. [76] (165 MCI stable, 137 MCI progressor). We have also noted that most of these models, except Lian et al. [78], involve various neuropsychological tests, which often differ between studies. This makes it challenging for external validation and comparison between different models. Future studies should consider developing models based on neuropsychological tests that are routinely used in clinics for easier evaluation, validation and potential implementation.

Predicting Aβ and tau levels in the brain
ML models have shown promise in predicting AD biomarkers with reasonable accuracy (Table 7D). For predicting Aβ and p-tau levels in the brain, the problem is often simplified into a binary classification, e.g., normal vs high or negative vs positive. Langford et al. [85] employed the extreme gradient boosting algorithm, a scalable tree boosting model to predict Aβ PET positivity (standardized uptake values $\geq$ 1.15) from demographics (age, education, gender and family history), four neuropsychological tests and APOE genotype., An AUC of 0.74 was achieved. Palmqvist et al. [84] used plasma $A\beta_{42}/ A\beta_{40}$ ratios, APOE genotype, and neuropsychological tests for a logistic regression with a lasso penalty model, and achieved an AUC of 0.83. In contrast, Lew et. al. [88] employed a logistic regression model for binary prediction of PET results (high versus low Aβ or p-tau) using MRI and other data (e.g., demographic, APOE genotype, neuropsychological tests and hippocampal

volumes etc.). This resulted in an AUC of 0.79 for Aβ and 0.73 for p-tau. Using a seven-layer neural network, 3,635 plasma proteins, age and APOE genotype for the same prediction, Zhang et al. [89] achieved a lower AUC score for Aβ (AUC=0.78) and p-tau (AUC=0.67). Their performance is relatively lower than the other studies, which could possibly be due to high feature-to-sample ratio (3000 proteins in 800 participants), which can complicate model training and validation.

Notably, a universally accepted threshold to determine binary classification is lacking. For example, Langford et al. [85] used a threshold of 1.15, while Palmqvist et al. [84] adopted a threshold of 0.738. Whether this would have impacted the prediction performance of the model is unclear. Future studies should consider standardizing this threshold to enable comparisons between models. Another issue with these studies is that the datasets used for model training are relatively small (e.g., 300 participants for Palmqvist et al. [84] and 800 participants for Zhang et al. [89]), possibly due to cost constraints associated with PET and MRI. Research funding bodies could play a role in encouraging (inter)national collaboration and data sharing, as well as endorsing standard data formats (especially for those high-cost experiments) to increase the size of datasets for more robust results.

Challenges and future directions
ML has been applied to clinical data analysis for more than two decades, and its widespread adoption in clinical research and healthcare has noticeably accelerated. This section will discuss the technical barriers, and the anticipated challenges and potential solutions to applying ML in clinical practice for dementia (summarized in Table 8).

Clinical data quality
Given the complex set up of longitudinal studies and heterogenous disease pathology, missing values, outliers, data imbalance are inevitable. Missing data is often due to incomplete responses, data collection errors, technical issues and participant withdrawal [96]. Data scientists either disregard participants with missing data or use imputation techniques (e.g., mean imputation, multiple imputation by chained equations, etc. [97]). Outliers normally result from errors from record, measurement or misclassification. Statistic techniques, such as *z*-scores and interquartile range or box plot are used to detect outliers. Once identified, common approaches involve removing outliers, adjusting into specific percentile, or applying transformations to reduce the skewness of the data distribution [98]. Data imbalance is a commonly encountered issue for dementia dataset, as MCI and ADem occur in a smaller population compared to CU. When MCI/ADem cases are significantly

Table 8 Challenges, solutions and future directions

Challenges	Solutions/future directions
Missing data	Utilize data imputation technique like mean imputation, multiple imputation by chained equations, etc
Data imbalance	Utilize resampling techniques like Synthetic Minority Over-sampling Technique, etc
Diagnostics error	Expand the use of subjective diagnostics criteria
Non-uniform longitudinal data	Data harmonization
Lack of generalizability	Develop global criteria that balance scientific rigor and practical feasibility
Exclusion of diverse populations	Encourage global collaborative efforts among researchers, clinicians, and regulatory bodies, strategic recruitment of people from culturally and linguistically diverse background
Computational burdens	Utilize efficient algorithm design, high-performance computing resources, and distributed computing platforms
Patient acceptance	Increase public awareness, ensure data transparency, security, and provide psychological support
Clinician acceptance	Offer ML training to medical students and clinicians, develop explainable AI techniques, and involve clinicians in co-design of ML tools to enhance usability and trust
Lack of interpretation for ML-dementia applications	Implement and promote explainable AI techniques like LIME and SHAP to make ML decision-making transparent
Ethical and regulatory considerations	Advocate for local and international ethical guidelines and regulatory compliance, ensure continuous monitoring post-deployment

underrepresented compared to CU, it can lead to a biased model performance, where ML models trained on imbalanced data may prioritize the majority and struggle to accurately predict the minority [99]. To address this issue, resampling techniques such as Synthetic Minority Over-sampling Technique [100] can be employed.

The quality of clinical data used to train ML models directly impacts the soundness of the model. The diagnoses are performed by clinicians and neuropsychologists [101, 102], which can sometimes introduce human errors into the dataset. This is because diagnosis is complicated by that 1) preclinical AD is difficult to detect [103], 2) MCI can be misclassified [104], and 3) vascular dementia, Lewy body dementia, and frontotemporal dementia are sometimes misdiagnosed as ADem [105]. Moreover, some neuropsychological tests are influenced by practice effects [106] (repeated testing can artificially improve performance over time), and education background [107] (poor performance for individuals who are less educated), potentially skewing results. Furthermore, the trajectory of dementia varies significantly among individuals due to the complex interplays of age, genetics, sex, and other comorbidities [108]. Some individuals may experience a gradual decline in cognition over many years, while others show rapid deterioration. Many longitudinal studies employ an "up-to-interval" method [75], classifying participants into CU, MCI, ADem, and non-ADem within a specified follow-up period. However, this approach often falls short in capturing the disease trajectory of individuals experiencing gradual cognitive decline. In addition, older participants are more likely to withdraw from the study due to their dependency on others (e.g., reduced mobility discourage their participation), leading to their disease trajectory not fully captured. Cohort study designs can be enhanced to improve data quality. Longitudinal study designs should consider incorporating more objective diagnostic criteria, such as expanding the use of Aβ PET scans, and integration of blood-based biomarkers, tau, and neuroinflammation markers, to enhance the assessments accuracy. Additionally, developing strategies to prolong study follow-up duration is crucial for capturing the full progression of disease states over time. Research funding bodies could play a crucial role in driving this progress by prioritizing investment and providing support to longitudinal studies.

Data standardization

The existing longitudinal datasets exhibit a lack of uniformity and standardized approach in sample/data collection and record format, making it difficult to validate and compare metrics like accuracy, sensitivity, and specificity between ML models that built on different datasets [109]. For example, although AIBL and ROSMAP collected depression related data, yet different scales were used—AIBL adapted the Hospital Anxiety and Depression Scale while ROSMAP used the Center for Epidemiological Studies Depression scale. The lack of uniformity in data collection could also be attributed to the intrinsic nature of the technology. For example, various platforms, techniques, and environmental factors could introduce biases and variabilities into omics dataset [110]. In addition, omics data is often noisy and sparse, especially when detecting molecules of low abundance, and therefore more prone to batch effect. Furthermore, different annotation systems or reference databases used to identify proteins, metabolites, and genes can lead to mismatches and inconsistencies. Also, different omics dataset may lack of common features due to experiment

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 17 of 21

set up. All these make it less practical to standardize the omics data.

To enhance the performance of ML models in dementia research, addressing variability in data collection methods is crucial. The Alzheimer's Dementia Onset and Progression in International Cohorts initiative [111] exemplifies the successful application of data harmonization, integrating data from five international dementia cohort studies, including the Adult Children Study, ADNI, AIBL, the Dominantly Inherited Alzheimer Network, and the National Alzheimer's Coordinating Center. Similar initiatives should be encouraged, as they are crucial for enhancing statistical power, and enabling more robust ML applications in dementia, leveraging the existing longitudinal datasets. In addition, publication of sample collection protocols, along with raising awareness of the requirements and benefits of data pooling for ML among biomedical and clinician scientists, could promote consistent data collection practices and enhance collaborative research efforts globally. Of paramount importance, inconsistencies in data formats can undermine the effectiveness of ML models. Advanced tools like 'dtool' provide practical solutions for standardizing data formats and enhancing quality by encapsulating data and metadata into consistent, unified dataset structures with readily accessible metadata for both the collective dataset and its individual files [112]. Data repositories could endorse guidelines that only accept datasets meeting standardized criteria.

Data generalizability

A longitudinal dataset may lack of generalizability. The study setting and enrolment criteria would exclude certain populations based on ethnicity, education level, socio-economic status, or comorbid conditions. For example, research studies might exclude participants with severe cardiovascular diseases or advanced diabetes, arguing that these conditions could confound the cognitive assessments used to diagnose and track ADem progression [113]. Moreover, studies that require participants to be English-speaking exclude individuals from a culturally and linguistically diverse background (e.g., the indigenous population in Australia, who have a higher risk of ADem). These exclusions can result in datasets that fail to fully represent the diverse population affected by dementia. The clinical application of ML models built from biased data will consequently be limited. Collaborative efforts between researchers, clinicians, and regulatory bodies are crucial in developing criteria that balance scientific rigor with practical feasibility. Furthermore, the major dementia longitudinal studies are often restricted to national boundaries, constraining their generalizability and the assessment of their performance in more border

real-world scenarios. Researchers are encouraged to employ multiple datasets, where the model is trained on one dataset (e.g., ADNI) and validated on another dataset (e.g., AIBL) [114] to address this challenge.

Computational and memory burden

Computational and memory burden is another technical challenge to ML-dementia, particularly as recent studies focus on high-dimensional longitudinal omics data. Advanced tools such as the versatile toolbox MEFISTO [115] and the PALMO platform [116] are now capable of modelling spatial and temporal omics data. These tools utilize high-performance computing resources and implement various optimization strategies to improve processing efficiency. However, the high computational and memory demands of these algorithms can limit their applicability in AD studies that involve large sample sizes. Furthermore, the high volume of data requires a robust data management solution. Distributed computing platforms, like Apache Hadoop [117], can be employed to efficiently handle, store, and share the large-scale data, facilitating collaborative efforts across different research groups and locations. However, these platforms are not always affordable, creating a technical barrier.

From bench to clinic

Artificial intelligence (AI), such as ML, has already demonstrated success in disease tracking, as evidenced by FDA-approved devices like Apple's Atrial Fibrillation History Feature [5]. While ML applications have yet to be implemented in dementia clinical practice, anticipated challenges must be considered for future implementation in dementia diagnosis and care.

Acceptance of ML tools by patients

The targeted population for ML-dementia tools is older adults, which raises questions about their readiness to accept these technological innovations [118]. Many older adults are not as technologically adept as younger generations, making it challenging for them to understand ML and its potential in diagnosing and managing diseases. This lack of understanding can result in low trust in ML-generated results, leading to hesitation in their use for healthcare purposes. Moreover, some ML tools collect data using wearable devices, raising privacy concerns among older adults who may be unsure how their data will be used. Furthermore, not all older adults want to receive predictions about their disease progression or early detection due to psychological fears and anxieties [119].

To address these challenges and improve acceptance among older adults, several steps should be taken. Increasing public awareness of ML and its benefits in

healthcare is crucial, as many people may not realize that AI/ML are already being used. Ensuring transparency in data usage and robust data security measures can help build trust, while offering a personalized approach where individuals can opt in or out of predictive analyses can promote autonomy [120]. Providing comprehensive psychological support can help individuals cope with the emotional impact of potential diagnoses and empower them to make informed decisions about their health and care plans. By addressing these concerns through patient education, demonstrating the reliability and benefits of ML tools, and ensuring robust data security measures, we can foster greater acceptance of ML-dementia tools among older adults.

Acceptance of ML tools by clinicians

Clinicians tend to prefer techniques that are transparent and interpretable, aligning with conventional clinical reasoning. One of the barriers for clinicians to trust and uptake the output of ML models is the opaque nature of these algorithms, often referred to as "black boxes." ML models can obscure the logic behind their complex decision-making processes, sometimes producing results that cannot be easily justified by existing biomedical knowledge. The "black box" nature of ML potentially erodes clinicians' trust, hindering the adoption of these models in clinical practice. In response to these challenges, there is an increasing focus on developing explainable AI techniques, such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) [121]. These methods aim to make the decision-making processes of ML models more transparent and understandable, thereby can potentially enhance trust among clinicians. Another significant challenge is that many clinicians have not received formal training in ML, which can hinder their ability to effectively use and explain these tools to patients [122]. Providing basic education about ML to clinicians and incorporating an AI/ML training component in medical school curriculum can enhance their ability to use innovative tools and communicate the benefits to patients. Of paramount importance, involving clinicians in the co-design of ML-dementia models can ensure AI/ML tools meet clinical needs and foster greater acceptance and integration into practice. Last but not least, some clinicians are hesitant to accept AI/ML tools due to concerns about job displacement [122]. However, it is essential to understand that AI/ML tools are designed to augment, not replace, the work of clinicians, similar to other diagnostic tests. Clinicians should be assured that their clinical judgment cannot be replaced by AI/ML and that the role of AI/ML in clinical practice should be clearly defined in relevant guidelines.

Ethics and regulatory considerations

The integration of AI/ML in healthcare brings forth numerous ethical and regulatory concerns that could potentially impede their implementation. Recently, the World Health Organization issued new guidance on the ethics and governance of AI technology applications in healthcare [123], emphasizing the need for AI/ML developers to prioritize ethical principles. To facilitate the potential implementation of AI/ML tools in dementia diagnosis and management, we also advocate for the development of local guidelines to fit the culture/religious needs. On the regulatory front, compliance with healthcare regulations is indispensable. Regulatory bodies, such as FDA, the European Medicines Agency, and the Therapeutic Goods administration (Australia), should get prepared for processing more applications for AI/ML medical devices in the future. A clear approach must be established for post-deployment continuous monitoring and reporting, to maintain their safety and effectiveness in the clinic [122]. More importantly, it is crucial that regulations should clearly define the responsibilities and accountabilities of AI/ML developers and healthcare providers for any errors generated by AI/ML tools. This includes specifying the extent of liability for developers in the event of AI/ML malfunction or incorrect predictions, as well as outlining the role of healthcare providers in interpreting AI/ML outputs before making clinical decisions. Regulations should also detail mechanisms for reporting and addressing errors, as well as protocols for updating and improving AI/ML tools from reported errors. An in-depth discussion on regulatory matters concerning ML/AI is outside the scope of this review. Regulatory bodies, clinicians, and public health experts are encouraged to work on regulatory matters to prepare our healthcare systems for the implementation of AI/ML tools.

Abbreviations

Aβ	Amyloid-beta
AD	Alzheimer's disease
ADem	Alzheimer's dementia
ADAS	Alzheimer's Disease Assessment Scale—Cognitive Subscale
ADNI	Alzheimer's Disease Neuroimaging Initiative
APOE	Apolipoprotein E
AI	Artificial intelligence
AIBL	Australian Imaging, Biomarker and Lifestyle (Study)
AUC	Area under the curve
CDR-SB	Clinical Dementia Rating—Sum of Boxes
CU	Cognitive unimpaired
MCI	Mild cognitive impairment
ML	Machine learning
MMSE	Mini-Mental State Examination
MRI	Magnetic resonance imaging
OASIS	Open Access Series of Imaging Studies
PET	Positron emission tomography
RL	Reinforcement learning
ROSMAP	Religious Orders Study/Memory and Aging Project
SVM	Support vector machine

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 19 of 21

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-024-01540-6.

Supplementary Material 1.

Authors' contributions
Conceptualization, Y.W., S.L., B.G., L.J. and Y.P.; literature review, Y.W. and S.L.; writing—original draft preparation, Y.W., S.L., Y.P.; writing—review and editing, A.S., A.H., C.C., B.G., C.L.M., L.J. and Y.P.; visualization, Y.W.; supervision, B.G., L.J. and Y.P.; project administration, L.J. and Y.P.; funding acquisition, Y.P. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors have consented for this review paper to be published.

Competing interests
The authors declare no competing interests.

References
1. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. Nat Rev Dis Primer. 2015;1:15056.
2. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina. 2020;56:455.
3. El Naqa I, Murphy MJ. "What is machine learning?" In: El Naqa I, Li R, Murphy MJ, editors. Machine Learning in Radiation Oncology. Cham: Springer; 2015. p. 3–11.
4. Imaging | GE HealthCare (Australia & New Zealand). Available from: https://www.gehealthcare.com.au/products/imaging. Cited 2024 Feb 4.
5. Apple Watch gets new heart health feature "AFib history". Available from: https://www.deccanherald.com/technology/apple-watch-gets-new-heart-health-feature-afib-history-1238219.html. Cited 2024 Feb 4.
6. Framingham Heart Study. Available from: https://www.framinghamheartstudy.org/. Cited 2024 May 21.
7. Home | National Institute on Aging: Baltimore Longitudinal Study of Aging. Available from: https://www.blsa.nih.gov/. Cited 2024 May 21.
8. Hofman A, Breteler MMB, Van Duijn CM, Krestin GP, Pols HA, Stricker BHCh, et al. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819–29.
9. Religious Orders Study/Memory and Aging Project (ROSMAP) - DSS NIAGADS. Available from: https://dss.niagads.org/cohorts/religious-orders-study-memory-and-aging-project-rosmap/. Cited 2024 Apr 24.
10. Yi Z. Introduction to the Chinese Longitudinal Healthy Longevity Survey (CLHLS). In: Yi Z, Poston DL, Vlosky DA, Gu D, editors. Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions. Dordrecht: Springer; 2008. cited 2024 Jul 23. p. 23–38.
11. National Alzheimer's Coordinating Center. Available from: https://naccdata.org/. Cited 2024 Apr 24.
12. Wisconsin Registry for Alzheimer's Prevention – UW–Madison. Available from: https://wrap.wisc.edu/. Cited 2024 May 21.
13. ADNI | Alzheimer's Disease Neuroimaging Initiative. Available from: https://adni.loni.usc.edu/. Cited 2024 Apr 24.
14. The AIBL study - aibl.org.au. Available from: https://aibl.org.au/. Cited 2024 Apr 24.
15. UK Biobank - UK Biobank. Available from: https://www.ukbiobank.ac.uk/. Cited 2024 Apr 24.
16. OASIS Brains - Open Access Series of Imaging Studies. Available from: https://www.oasis-brains.org/.Cited 2024 Apr 24.
17. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian longitudinal study on aging (CLSA). Can J Aging. 2009;28:221–9.
18. Le Duff F, Develay AE, Quetel J, Lafay P, Schueck S, Pradier C, et al. The 2008–2012 French Alzheimer plan: description of the national Alzheimer information system. J Alzheimers Dis. 2012;29:891–902.
19. Whelan BJ, Savva GM. Design and methodology of the Irish longitudinal study on ageing. J Am Geriatr Soc. 2013;61:S265–8.
20. Molinuevo JL, Gramunt N, Gispert JD, Fauria K, Esteller M, Minguillon C, et al. The ALFA project: a research platform to identify early pathophysiological features of Alzheimer's disease. Alzheimers Dement Transl Res Clin Interv. 2016;2:82–92.
21. Bos I, Vos S, Vandenberghe R, Scheltens P, Engelborghs S, Frisoni G, et al. The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. Alzheimers Res Ther. 2018;10:64.
22. Anti-Amyloid treatment in Asymptomatic Alzheimer's disease (A4). Available from: https://www.alzheimers.gov/clinical-trials/anti-amyloid-treatment-asymptomatic-alzheimers-disease-a4. Cited 2024 Apr 24.
23. Blennow K, Zetterberg H. Cerebrospinal fluid biomarkers for Alzheimer's disease. J Alzheimers Dis. 2009;18:413–7.
24. Yang J, Wu S, Yang J, Zhang Q, Dong X. Amyloid beta-correlated plasma metabolite dysregulation in Alzheimer's disease: an untargeted metabolism exploration using high-resolution mass spectrometry toward future clinical diagnosis. Front Aging Neurosci. 2023;15:1189659.
25. Howard J. New blood test that screens for Alzheimer's may be a step closer to reality, study suggests. CNN. 2024. Available from: https://www.cnn.com/2024/01/22/health/alzheimers-blood-test-screening-study/index.html. Cited 2024 Jun 26.
26. Fandos N, Pérez-Grijalba V, Pesini P, Olmos S, Bossa M, Villemagne VL, et al. Plasma amyloid β 42/40 ratios as biomarkers for amyloid β cerebral deposition in cognitively normal individuals. Alzheimers Dement Diagn Assess Dis Monit. 2017;8:179–87.
27. Janelidze S, Mattsson N, Palmqvist S, Smith R, Beach TG, Serrano GE, et al. Plasma p-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. Nat Med. 2020;26:379–86.
28. Dementia Panel Test - PreventionGenetics. Available from: https://www.preventiongenetics.com/testInfo?val=Dementia-Panel. Cited 2024 Jan 17.
29. Metabolomics Core Services and Fees | BCM. Available from: https://www.bcm.edu/research/atc-core-labs/metabolomics-core/services-and-fees. Cited 2024 Jun 26.
30. Pricing and Ordering - Commercial. Proteomics Int. Available from: https://www.proteomics.com.au/analytical-services/pricing-and-ordering/. Cited 2024 Jun 26.
31. Hansen N, Rauter C, Wiltfang J. Blood based biomarker for optimization of early and differential diagnosis of Alzheimer's dementia. Fortschr Neurol Psychiatr. 2022;90:326–35.
32. Guo Y, You J, Zhang Y, Liu W-S, Huang Y-Y, Zhang Y-R, et al. Plasma proteomic profiles predict future dementia in healthy adults. Nat Aging. 2024;4:247–60.
33. Sequencing - Genomics Research Centre. Available from: https://research.qut.edu.au/grc/diagnostic-testing/service-pricing/sequencing/. Cited 2024 Apr 24.
34. Hancock C, Bernal B, Medina C, Medina S. Cost analysis of diffusion tensor imaging and MR tractography of the brain. Open J Radiol. 2014;4:260–9.

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 20 of 21

35. Sperling R. The potential of functional MRI as a biomarker in early Alzheimer's disease. Neurobiol Aging. 2011;32:S37-43.

36. Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol. 2010;6:67–77.

37. Nordberg A, Rinne JO, Kadir A, Långström B. The use of PET in Alzheimer disease. Nat Rev Neurol. 2010;6:78–87.

38. De la Fuente GS, Ritchie CW, Luz S. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. J Alzheimers Dis. 2020;78:1547–74.

39. Standard Search | Medicare Benefits Schedule. Available from: https://www9.health.gov.au/mbs/search.cfm. Cited 2024 Jun 26.

40. Mintun MA, Lo AC, Duggan Evans C, Wessels AM, Ardayfio PA, Andersen SW, et al. Donanemab in early Alzheimer's disease. N Engl J Med. 2021;384:1691–704.

41. Van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early Alzheimer's disease. N Engl J Med. 2023;388:9–21.

42. Tan MS, Cheah P-L, Chin A-V, Looi L-M, Chang S-W. A review on omics-based biomarkers discovery for Alzheimer's disease from the bioinformatics perspectives: statistical approach vs machine learning approach. Comput Biol Med. 2021;139:104947.

43. Fisher CK, Smith AM, Walsh JR. Machine learning for comprehensive forecasting of Alzheimer's disease progression. Sci Rep. 2019;9:13622.

44. Önen Dumlu Z, Sayın S, Gürvit İH. Screening for preclinical Alzheimer's disease: deriving optimal policies using a partially observable Markov model. Health Care Manag Sci. 2023;26:1–20.

45. Wang M, Greenberg M, Forkert ND, Chekouo T, Afriyie G, Ismail Z, et al. Dementia risk prediction in individuals with mild cognitive impairment: a comparison of cox regression and machine learning models. BMC Med Res Methodol. 2022;22:284.

46. Quemy A. Two-stage optimization for machine learning workflow. Inf Syst. 2020;92:101483.

47. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). New Delhi: IEEE; 2016. p. 1310–5.

48. Hahne F, Huber W, Gentleman R, Falcon S, Gentleman R, Carey VJ. Unsupervised machine learning, in bioconductor case studies. J R Stat Soc Ser A Stat Soc. 2010;173:465–6.

49. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. J Artif Intell Res. 1996;4:237–85.

50. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. BioEssays. 2004;26:99–105.

51. Selkoe DJ. Translating cell biology into therapeutic advances in Alzheimer's disease. Nature. 1999;399:A23-31.

52. Duchesne S, Caroli A, Geroldi C, Barillot C, Frisoni GB, Collins DL. MRI-based automated computer classification of probable AD versus normal controls. IEEE Trans Med Imaging. 2008;27:509–20.

53. Zhang D, Shen D, Alzheimer's Disease Neuroimaging Initiative. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS ONE. 2012;7:e33182.

54. Khagi B, Kwon G-R. 3D CNN design for the classification of Alzheimer's disease using brain MRI and PET. IEEE. 2020;8:217830–47.

55. Ramzan F, Khan MUG, Rehmat A, Iqbal S, Saba T, Rehman A, et al. A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. J Med Syst. 2020;44:37.

56. Balaji P, Chaurasia MA, Bilfaqih SM, Muniasamy A, Alsid LEG. Hybridized deep learning approach for detecting Alzheimer's disease. Biomedicines. 2023;11:149.

57. Chu C, Wang YF, Wang Y, Fowler C, Masters CL, et al. Dementia severity age: a novel indicator to predict the onset of MCI and Alzheimer's dementia. https://doi.org/10.2139/ssrn.4845137.

58. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14:535–62.

59. Sancesario GM, Bernardini S. Alzheimer's disease in the omics era. Clin Biochem. 2018;59:9–16.

60. Weintraub S, Wicklund AH, Salmon DP. The neuropsychological profile of Alzheimer disease. Cold Spring Harb Perspect Med. 2012;2:a006171.

61. Richards M, Folstein M, Albert M, Miller L, Bylsma F, Lafleche G, et al. Multicenter study of predictors of disease course in Alzheimer disease (the "predictors study"). II. Neurological, psychiatric, and demographic influences on baseline measures of disease severity. Alzheimer Dis Assoc Disord. 1993;7:22–32.

62. O'Bryant SE. Staging dementia using clinical dementia rating scale sum of boxes scores: a Texas Alzheimer's research consortium study. Arch Neurol. 2008;65:1091.

63. Perneczky R, Wagenpfeil S, Komossa K, Grimmer T, Diehl J, Kurz A. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. Am J Geriatr Psychiatry. 2006;14:139–44.

64. Kasula BY. A machine learning approach for differential diagnosis and prognostic prediction in Alzheimer's disease. Int J Sustain Dev Comput Sci. 2023;5:1–8.

65. Nemoto K, Sakaguchi H, Kasai W, Hotta M, Kamei R, Noguchi T, et al. Differentiating dementia with Lewy bodies and Alzheimer's disease by deep learning to structural MRI. J Neuroimaging. 2021;31:579–87.

66. Nguyen H-D, Clément M, Planche V, Mansencal B, Coupé P. Deep grading for MRI-based differential diagnosis of Alzheimer's disease and Frontotemporal dementia. Artif Intell Med. 2023;144:102636.

67. Qiang Y-X, You J, He X-Y, Guo Y, Deng Y-T, Gao P-Y, et al. Plasma metabolic profiles predict future dementia and dementia subtypes: a prospective analysis of 274,160 participants. Alzheimers Res Ther. 2024;16:16.

68. Marzban EN, Eldeib AM, Yassine IA, Kadah YM, Initiative ADN. Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks. PLoS ONE. 2020;15:e0230409.

69. Wang J, Wei R, Xie G, Arnold M, Kueider-Paisley A, Louie G, et al. Peripheral serum metabolomic profiles inform central cognitive impairment. Sci Rep. 2020;10:14059.

70. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci Rep. 2021;11:3254.

71. Naz S, Ashraf A, Zaib A. Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. Multimed Syst. 2022;28:85–94.

72. Rye I, Vik A, Kocinski M, Lundervold AS, Lundervold AJ. Predicting conversion to Alzheimer's disease in individuals with mild cognitive impairment using clinically transferable features. Sci Rep. 2022;12:15566.

73. Hashmi A, Barukab O. Dementia classification using deep reinforcement learning for early diagnosis. Appl Sci. 2023;13:1464.

74. Mahendran N, Vincent PMDR. Deep belief network-based approach for detecting Alzheimer's disease using the multi-omics data. Comput Struct Biotechnol J. 2023;21:1651–60.

75. Beltrán JF, Wahba BM, Hose N, Shasha D, Kline RP, For the Alzheimer's Disease Neuroimaging Initiative. Inexpensive, non-invasive biomarkers predict Alzheimer transition using machine learning analysis of the Alzheimer's Disease Neuroimaging (ADNI) database. PLOS ONE. 2020;15:e0235663.

76. Jiang S, Xie Y, Colditz GA. Functional ensemble survival tree: dynamic prediction of Alzheimer's disease progression accommodating multiple time-varying covariates. J R Stat Soc Ser C Appl Stat. 2021;70:66–79.

77. Kwak S, Oh DJ, Jeon Y-J, Oh DY, Park SM, Kim H, et al. Utility of machine learning approach with neuropsychological tests in predicting functional impairment of Alzheimer's disease. J Alzheimers Dis. 2022;85:1357–72.

78. Lian C, Liu M, Wang L, Shen D. Multi-task weakly-supervised attention network for dementia status estimation with structural MRI. IEEE Trans Neural Netw Learn Syst. 2021;33:4056–68.

79. Mofrad SA, Lundervold AJ, Vik A, Lundervold AS. Cognitive and MRI trajectories for prediction of Alzheimer's disease. Sci Rep. 2021;11:2122.

80. Saboo K, Choudhary A, Cao Y, Worrell G, Jones D, Iyer R. Reinforcement learning based disease progression model for Alzheimer's disease. Adv Neural Inf Process Syst. 2021;34:20903–15.

81. Mukherji D, Mukherji M, Mukherji N, Alzheimer's Disease Neuroimaging Initiative. Early detection of Alzheimer's disease using neuropsychological tests: a predict–diagnose approach using neural networks. Brain Inform. 2022;9:23.

82. Bucholc M, Titarenko S, Ding X, Canavan C, Chen T. A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia. Expert Syst Appl. 2023;217:119541.

Wang *et al. Alzheimer's Research & Therapy*   (2024) 16:175

Page 21 of 21

83. Zou H, Zeng D, Xiao L, Luo S. Bayesian inference and dynamic prediction for multivariate longitudinal and survival data. Ann Appl Stat. 2023;17:2574–95.

84. Palmqvist S, Insel PS, Zetterberg H, Blennow K, Brix B, Stomrud E, et al. Accurate risk estimation of β-amyloid positivity to identify prodromal Alzheimer's disease: cross-validation study of practical algorithms. Alzheimers Dement. 2019;15:194–204.

85. Langford O, Raman R, Sperling RA, Cummings J, Sun C-K, Jimenez-Maggiora G, et al. Predicting amyloid burden to accelerate recruitment of secondary prevention clinical trials. J Prev Alzheimers Dis. 2020;7:213–8.

86. Shan G, Bernick C, Caldwell JZK, Ritter A. Machine learning methods to predict amyloid positivity using domain scores from cognitive tests. Sci Rep. 2021;11:4822.

87. Janelidze S, Teunissen CE, Zetterberg H, Allué JA, Sarasa L, Eichenlaub U, et al. Head-to-head comparison of 8 plasma amyloid-β 42/40 assays in Alzheimer disease. JAMA Neurol. 2021;78:1375.

88. Lew CO, Zhou L, Mazurowski MA, Doraiswamy PM, Petrella JR, Initiative ADN. MRI-based deep learning assessment of amyloid, tau, and neurodegeneration biomarker status across the Alzheimer disease spectrum. Radiology. 2023;309:e222441.

89. Zhang Y, Ghose U, Buckley NJ, Engelborghs S, Sleegers K, Frisoni GB, et al. Predicting AT(N) pathologies in Alzheimer's disease from blood-based proteomic data using neural networks. Front Aging Neurosci. 2022;14:1040001.

90. Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. Alzheimer Dis Assoc Disord. 2002;16:203–12.

91. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. J Neuropathol Exp Neurol. 2012;71:266–73.

92. Castellazzi G, Cuzzoni MG, Cotta Ramusino M, Martinelli D, Denaro F, Ricciardi A, et al. A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features. Front Neuroinformatics. 2020;14:25.

93. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris AB, Alzakari N, et al. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Appl Sci. 2021;11:796.

94. Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, Hosseinzadeh M, et al. Machine learning (ML) in medicine: review, applications, and challenges. Mathematics. 2021;9:2970.

95. Cascarano A, Mur-Petit J, Hernández-González J, Camacho M, De Toro EN, Gkontra P, et al. Machine and deep learning for longitudinal biomedical data: a review of methods and applications. Artif Intell Rev. 2023;56:1711–71.

96. Molenberghs G, Kenward M. Missing data in clinical studies. Chichester: Wiley; 2007.

97. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30:377–99.

98. Mowbray FI, Fox-Wasylyshyn SM, El-Masri MM. Univariate outliers: a conceptual overview for the nurse researcher. Can J Nurs Res. 2019;51:31–7.

99. Dubey R, Zhou J, Wang Y, Thompson PM, Ye J, Initiative ADN. Analysis of sampling techniques for imbalanced data: An n= 648 ADNI study. Neuroimage. 2014;87:220–41.

100. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

101. Fowler C, Rainey-Smith SR, Bird S, Bomke J, Bourgeat P, Brown BM, et al. Fifteen years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. J Alzheimers Dis Rep. 2021;5:443–68.

102. Casaletto KB, Heaton RK. Neuropsychological assessment: past and future. J Int Neuropsychol Soc. 2017;23:778–90.

103. Sperling RA, Karlawish J, Johnson KA. Preclinical Alzheimer disease—the challenges ahead. Nat Rev Neurol. 2013;9:54–8.

104. Edmonds EC, Delano-Wood L, Galasko DR, Salmon DP, Bondi MW. Subjective cognitive complaints contribute to misdiagnosis of mild cognitive impairment. J Int Neuropsychol Soc. 2014;20:836–47.

105. Skillbäck T, Farahmand BY, Rosén C, Mattsson N, Nägga K, Kilander L, et al. Cerebrospinal fluid tau and amyloid-β $_{1-42}$ in patients with dementia. Brain. 2015;138:2716–31.

106. Duff K, Lyketsos CG, Beglinger LJ, Chelune G, Moser DJ, Arndt S, et al. Practice effects predict cognitive outcome in amnestic mild cognitive impairment. Am J Geriatr Psychiatry. 2011;19:932–9.

107. Loewenstein DA, Arguelles T, Barker WW, Duara R. A comparative analysis of neuropsychological test performance of -speaking and -speaking patients with Alzheimer's disease. J Gerontol. 1993;48:P142–9.

108. Ma L, Tan ECK, Bush AI, Masters CL, Goudey B, Jin L, et al. Elucidating the Link Between Anxiety/Depression and Alzheimer's Dementia in the Australian Imaging Biomarkers and Lifestyle (AIBL) Study. J Epidemiol Glob Hea. 2024;4364 (ahead of print).

109. Salthouse TA. Within-cohort age-related differences in cognitive functioning. Psychol Sci. 2013;24:123–30.

110. Yu Y, Zhang N, Mai Y, Ren L, Chen Q, Cao Z, et al. Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. Genome Biol. 2023;24:201.

111. Bourgeat P, Doré V, Rowe CC, Benzinger T, Tosun D, Goyal MS, et al. A universal neocortical mask for Centiloid quantification. Alzheimers Dement Diagn Assess Dis Monit. 2023;15:e12457.

112. Olsson TSG, Hartley M. Lightweight data management with dtool. PeerJ. 2019;7:e6562.

113. Irie F, Fitzpatrick AL, Lopez OL, Kuller LH, Peila R, Newman AB, et al. Enhanced risk for Alzheimer disease in persons with type 2 diabetes and APOE ε4: the cardiovascular health study cognition study. Arch Neurol. 2008;65:89–93.

114. Shishegar R, Cox T, Rolls D, Bourgeat P, Doré V, Lamb F, et al. Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI. Sci Rep. 2021;11:23788.

115. Velten B, Braunger JM, Argelaguet R, Arnol D, Wirbel J, Bredikhin D, et al. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. Nat Methods. 2022;19:179–86.

116. Vasaikar SV, Savage AK, Gong Q, Swanson E, Talla A, Lord C, et al. PALMO: a comprehensive platform for analyzing longitudinal multi-omics data. Nat Commn. 2023;14:1684.

117. Apache Hadoop. Available from: https://hadoop.apache.org/. Cited 2024 Apr 24.

118. Jauk S, Kramer D, Avian A, Berghold A, Leodolter W, Schulz S. Technology acceptance of a machine learning algorithm predicting delirium in a clinical setting: a mixed-methods study. J Med Syst. 2021;45:48.

119. Anxiety and older adults: overcoming worry and fear - Afmerican Association for Geriatric Psychiatry. https://www.aagponline.org/. Available from: https://www.aagponline.org/patient-article/anxiety-and-older-adults-overcoming-worry-and-fear/. Cited 2024 Jun 25.

120. Saunders S, Gomes-Osman J, Jannati A, Ciesla M, Banks R, Showalter J, et al. Towards a lifelong personalized brain health program: empowering individuals to define, pursue, and monitor meaningful outcomes. Front Neurol. 2024;15:1387206.

121. Peng J, Zou K, Zhou M, Teng Y, Zhu X, Zhang F, et al. An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. J Med Syst. 2021;45:61.

122. Nair M, Andersson J, Nygren JM, Lundgren LE. Barriers and enablers for implementation of an artificial intelligence–based decision support tool to reduce the risk of readmission of patients with heart failure: Stakeholder interviews. JMIR Form Res. 2023;7:e47335.

123. Organization WH. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. 2024. 2024.

Publisher's Note