

<https://doi.org/10.1038/s41746-025-01640-z>

# Alzheimer's disease digital biomarkers multidimensional landscape and AI model scoping review

Check for updates

Wenhai Qi<sup>1,13</sup>, Xiaohong Zhu<sup>1,13</sup>, Bin Wang<sup>1,13</sup>, Yankai Shi<sup>1</sup>, Chaoqun Dong<sup>1</sup>, Shiyiing Shen<sup>1</sup>, Jiaqi Li<sup>1</sup>, Kun Zhang<sup>2</sup>, Yunfan He<sup>3</sup>, Mengjiao Zhao<sup>4</sup>, Shiyuan Yao<sup>1</sup>, Yongze Dong<sup>5</sup>, Huajuan Shen<sup>5</sup>, Junling Kang<sup>6</sup>, Xiaodong Lu<sup>7</sup>, Guowei Jiang<sup>8</sup>, Lizzy M. M. Boots<sup>8</sup>, Heming Fu<sup>9</sup>, Li Pan<sup>10</sup>, Hongkai Chen<sup>11</sup>, Zhenyu Yan<sup>11</sup>, Guoliang Xing<sup>11</sup>✉ & Shihua Cao<sup>1,12</sup>✉

As digital biomarkers gain traction in Alzheimer's disease (AD) diagnosis, understanding recent advancements is crucial. This review conducts a bibliometric analysis of 431 studies from five online databases: Web of Science, PubMed, Embase, IEEE Xplore, and CINAHL, and provides a scoping review of 86 artificial intelligence (AI) models. Research in this field is supported by 224 grants across 54 disciplines and 1403 institutions in 44 countries, with 2571 contributing researchers. Key focuses include motor activity, neurocognitive tests, eye tracking, and speech analysis. Classical machine learning models dominate AI research, though many lack performance reporting. Of 21 AD-focused models, the average AUC is 0.887, while 45 models for mild cognitive impairment show an average AUC of 0.821. Notably, only 2 studies incorporated external validation, and 3 studies performed model calibration. This review highlights the progress and challenges of integrating digital biomarkers into clinical practice.

As the global population continues to age, the incidence and severity of Alzheimer's disease (AD) have steadily increased, posing a significant public health challenge and disease burden worldwide<sup>1</sup>. According to the World Health Organization, more than 55 million people are currently living with dementia<sup>2</sup>, with AD accounting for 60%-70% of all dementia cases<sup>3</sup>. The main symptoms of AD include cognitive dysfunction, memory loss, and mood fluctuations<sup>4</sup>. Due to its complex etiology and unclear pathophysiological mechanisms, there are currently no targeted treatments or drugs capable of fully reversing the disease progression, which places substantial economic and healthcare burdens on both society and patients' families<sup>5</sup>. This therapeutic challenge underscores the critical importance of early detection and diagnosis, which can extend the window for intervention, delay symptom progression, optimize care planning, reduce caregiver burden, and ultimately preserve patient quality of life for longer periods.

Currently, the clinical diagnosis of AD primarily relies on neuropsychological assessments and the traditional biomarker-based diagnostic methods defined by the NIA-AA's ATN framework<sup>6</sup>. Although these conventional approaches are widely accepted, they are often costly, difficult to scale, and may involve invasive or inconvenient procedures, making frequent testing challenging. For instance, measuring "A" markers, which reflect β-amyloid protein levels, typically requires a lumbar puncture to assess cerebrospinal fluid Aβ42 concentrations<sup>7</sup> or amyloid PET scans<sup>8</sup>. Additionally, while neuropsychological assessments are simple and quick, they heavily rely on patient self-reports, introducing subjectivity and variability depending on the evaluator. Moreover, these assessments are usually conducted at specific time points, making them susceptible to various factors (e.g., patient comorbidities, medication use, or motivation), which can result in misdiagnosis or underdiagnosis. Given these limitations, many researchers are exploring alternative and emerging diagnostic methods to

<sup>1</sup>School of Nursing, Hangzhou Normal University, Hangzhou, China. <sup>2</sup>School of Nursing, Fujian Medical University, Fujian, China. <sup>3</sup>School of International Relations and Public Affairs, Fudan University, Shanghai, China. <sup>4</sup>School of Nursing, Zhejiang Chinese Medical University, Hangzhou, China. <sup>5</sup>Nursing Department, Zhejiang Provincial People's Hospital, Hangzhou, China. <sup>6</sup>Department of Neurology, The Third Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China.

<sup>7</sup>Department of Neurology, Affiliated Hospital of Hangzhou Normal University, Hangzhou, China. <sup>8</sup>Department of Psychiatry and Neuropsychology and Alzheimer Center Limburg, School for Mental Health and Neuroscience Maastricht University, Maastricht, The Netherlands. <sup>9</sup>Department of Electrical and Computer Engineering, Stony Brook University, New York, USA. <sup>10</sup>Department of Pathology, the University of Hong Kong, Hong Kong SAR, China. <sup>11</sup>Department of Electrical and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China. <sup>12</sup>Key Laboratory of Cognitive Disorder Assessment Technology, Hangzhou, China. <sup>13</sup>These authors contributed equally: Wenhai Qi, Xiaohong Zhu, Bin Wang. ✉ e-mail: [glxing@ie.cuhk.edu.hk](mailto:glxing@ie.cuhk.edu.hk); [csh@hznu.edu.cn](mailto:csh@hznu.edu.cn)

supplement the existing diagnostic toolkit. For instance, ocular biomarkers<sup>9</sup>, other fluid-based markers<sup>10</sup>, and blood-based biomarkers<sup>11</sup> have been investigated. While these approaches have reduced testing costs and minimized invasive procedures to some extent, they still fall short of meeting the requirements for repeated testing or longitudinal monitoring and continue to face long-standing challenges in these areas.

The exponential growth in digital biomarker studies has created an urgent need for systematic analysis of research trends, methodological approaches, and translational progress in this rapidly evolving field. The U.S. Food and Drug Administration (FDA) defines digital biomarkers as “a characteristic or set of characteristics collected through digital health technologies, which serve as indicators of normal biological processes, pathogenic processes, or responses to exposure or interventions, including therapeutic interventions”<sup>12</sup>. In AD research, digital biomarkers typically refer to objective, quantifiable physiological and behavioral data collected through digital devices such as sensors, wearables, and implantable devices<sup>13</sup>. Examples include gait parameters measured via wearable devices<sup>14,15</sup>, eye movement parameters collected by eye-tracking devices<sup>16</sup>, and speech features recorded through microphones<sup>17</sup>. These measurement methods are not only objective and ecologically valid, but also enable seamless data collection during daily activities, enabling real-time monitoring of subtle changes in health status<sup>18</sup>. Furthermore, digital biomarkers support longitudinal data collection, making continuous tracking of patients’ health status possible<sup>19</sup>. This capability is crucial for early prediction and intervention in disease progression, ultimately improving treatment outcomes and enhancing quality of life. Therefore, digital biomarkers hold significant potential for the diagnosis and management of AD and are poised to become essential tools for disease monitoring and treatment in the future.

With the rapid development of digital biomarkers in AD, a comprehensive overview of the multidimensional landscape in this field is becoming increasingly essential to understand emerging trends. While several reviews have explored the potential applications of digital biomarkers in AD, their scope and focus diverge from that of our study. For instance, one review analyzed dementia-related digital biomarker phenotypes derived from mobile and wearable devices, highlighting the potential of these technologies<sup>20</sup>. Similarly, another scoping review concentrates on the use of digital biomarker in non-clinical, home-based settings for monitoring and follow-up in mild cognitive impairment (MCI) or early-stage AD<sup>18</sup>. These reviews are more context-specific and do not address the developmental variations across different types of digital biomarkers. Furthermore, as these reviews were published earlier, they now require updating to reflect recent advances. With the growing integration of artificial intelligence (AI) in healthcare, it is crucial to explore the potential synergies between AI and digital biomarkers; however, existing reviews predominantly focus on AI’s applications in traditional biomarkers<sup>21,22</sup>. This study aims to address these gaps by systematically reviewing and analyzing the current landscape of digital biomarkers in AD, with a particular emphasis on recent advancements in AI and interdisciplinary collaboration. Through this comprehensive analysis, we aim to: (1) identify emerging research frontiers and unexplored opportunities; (2) uncover methodological strengths and limitations in current approaches; (3) characterize successful models of interdisciplinary and industry-academic collaboration; and (4) provide evidence-based recommendations for accelerating translation of digital biomarkers into real-world clinical practice. Table 1 highlights the main differences between our study and previous reviews in the field.

## Protocol and registration

This study adheres to the PRISMA-ScR guidelines for scoping reviews and reporting the search strategy<sup>23</sup>. The protocol was pre-registered through the Open Science Framework (<https://doi.org/10.17605/OSF.IO/6DK5U>). The checklist for this study can be found in Supplementary Table 1.

## Ethical considerations

Ethics committee permission was not required, as this study was a retrospective analysis of the existing published studies.

## Results

### Number of articles included in the analysis

Initially, a total of 24,257 records were retrieved. After a series of filtering steps, 16,205 studies remained. Finally, 431 studies were selected for bibliometric analysis, of which 15 were added through reference list searches. After evaluating the full text of the 431 records, 86 studies were included in our scoping review, with 12 being added following a second-round search update. The specific retrieval and screening process is illustrated in Fig. 1a, and Fig. 1b provides a more detailed overview of our research content.

### The annual trends of publications

Conducting productivity analysis in a research field helps in understanding the dynamics and emerging trends within that field. The earliest related study in this field was published in 2004, and after 20 years of development, the cumulative output has reached 431 publications. In 2017, the annual output exceeded 20 publications for the first time (22/431, 5.10%), and in both 2022 and 2023 it peaked at 77 publications (77/431, 17.9%). The red dashed line in Fig. 2a represents the fitted trend line, showing an overall upward trajectory ( $R^2 = 0.87$ , indicating a good model fit). Joinpoint analysis identified significant turning points in publication volume in the years 2013, 2019, and 2022. The slope for the period 2004–2013 (slope1) was 0.76, for 2013–2019 (slope2) was 3.35, and for 2019–2022 (slope3) was 16.22. The slope from 2022 to 2024, due to the retrieval cutoff date (May 2024), did not cover the full year of 2024, resulting in a slope of -13.90. Notably, the differences in slopes between slope1 and slope2, slope2 and slope3, as well as slope3 and slope4 were all statistically significant ( $p < 0.05$ ), indicating notable changes in the growth trends at these time points, as shown in Fig. 2b.

Based on publication volume and slope changes, the research output in this field can be divided into three distinct phases. The first stage (2004–2012) had a total of 26 publications (6.03% of the total), with a compound annual growth rate (CAGR) of 25.10%. The second stage (2013–2018) saw 91 publications (21.11% of the total), with a CAGR of 25.59%. In the third stage (2019–2024), although the data for 2024 is not yet complete, the trend line in Fig. 2a predicts continued growth in output for 2024. During this stage, 314 publications (72.85% of the total), with a CAGR of 27.34% from 2019 to 2023.

### Institutional analysis

Conducting institutional output and collaboration analysis aids in dissecting the structure of the research field. A total of 912 institutions have participated in research on AD digital biomarkers, collectively publishing 1,403 papers. Among these institutions, 489 (53.6%) are universities, contributing 864 papers (61.6%); 209 (22.9%) are hospitals, producing 268 papers (19.1%); 134 (14.7%) are research institutes or government entities, publishing 174 papers (12.4%); and 80 (8.8%) are companies, publishing 97 papers (6.9%), as shown in Figs. 3a and 3b. Oregon Health & Science University in the United States has the highest output, publishing 13 papers (0.93%). This institution also has the highest citation count, with 492 citations (1.98%), and an average of 37.85 citations per paper. The top 10 institutions by output are predominantly universities, with nearly half located in the United States, as detailed in Table 2.

A total of 236 institutions published more than two papers (25.87%), with a combined output of 727 papers (51.82% of the total publications). These institutions have formed close collaborations within several cooperative clusters, particularly centered around high-output universities. In terms of the timeline, institutional collaboration has been primarily concentrated since 2019, as shown in Fig. 3c, d.

### Country analysis

Analyzing national output and collaboration patterns provides a macro-level understanding of the global progress in AD digital biomarkers, the research disparities between countries, and emerging collaboration trends. A total of 44 countries have contributed to research publications in this field. The top 10 most productive countries (with ties allowed, resulting in the

**Table 1 | Highlights the key differences between our bibliometric and scoping review study and previous research in the field, indicating how our study builds upon and updates earlier work**

Previous reviews (Year)	Scope	Comparative contribution of our review
Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. (2019) <sup>20</sup>	This review explores how digital biomarkers obtained through mobile and wearable devices can assist in the early detection and monitoring of AD, particularly by utilizing data collected from multiple sensors to identify early signs of the disease.	Our review provides an overview of the broad range of AD digital biomarker research, including details on data collection devices, tasks, and other relevant aspects. Additionally, we discuss the trends in the popularity of various types of digital biomarkers.
Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. (2019) <sup>18</sup>	This review evaluates the application of digital biomarker technologies in home settings for the early detection and monitoring of MCI and AD, exploring the potential and prospects of various technologies.	Our review covers a variety of digital biomarker technologies, extending beyond home-based applications. Additionally, we review six studies on home-based monitoring digital biomarkers that are based on AI technologies.
Artificial intelligence for dementia—Applied models and digital health. (2023) <sup>190</sup>	This review explores how artificial intelligence and digital health technologies, through machine learning and digital biomarkers, play a crucial role in risk prediction, diagnosis, and monitoring disease progression in AD.	Building on this research, our review specifically examines 86 studies utilizing artificial intelligence and digital biomarkers, offering multifaceted insights into AI models.
Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. (2023) <sup>221</sup>	The primary contribution of this review lies in exploring the applications of AI and machine learning in identifying biomarkers for AD.	This review specifically focuses on the research of AI models for digital biomarkers.
Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions. (2022) <sup>21</sup>	This review focuses on how AI aids in the early detection of dementia, including AD, through methods such as computerized cognitive testing, motion analysis, speech and language testing, and brain imaging interpretation.	Our review not only covers AI research in areas such as motion analysis and speech analysis but also provides a more detailed overview of various types of digital biomarkers.
A review of artificial intelligence methods for Alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis. (2024) <sup>22</sup>	This review explores the applications of AI and machine learning in early diagnosis and monitoring of AD by integrating neuroimaging and sensor data, with a particular focus on the review of imaging content.	Our review focuses solely on digital biomarkers, providing a systematic overview of AI applications in various types of digital biomarkers.
Predictive Accuracy of Digital Biomarker Technologies for Detection of Mild Cognitive Impairment and Pre-Frailty Amongst Older Adults: A Systematic Review and Meta-Analysis. (2022) <sup>222</sup>	This review summarizes the predictive accuracy of digital biomarker technologies in the early detection of MCI and pre-frailty through systematic evaluation and meta-analysis.	Our review places greater emphasis on the research of digital biomarkers for AD and MCI, providing a comprehensive review of 86 AI models, including their performance.
The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. (2022) <sup>164</sup>	This review extracted data from other systematic reviews on the performance of AI models in diagnosing mental disorders, including AD and MCI, summarizing their accuracy, sensitivity, and specificity. However, this study focuses on research utilizing neuroimaging and psychological data.	Our review focuses on 86 studies of different types of digital biomarker AI research, providing detailed information including performance, data collection tasks, devices, model validation, code availability, and so on.
Mapping Knowledge Landscapes and Emerging Trends in AI for Dementia Biomarkers: Bibliometric and Visualization Analysis. (2024) <sup>163</sup>	This review uses bibliometric analysis to assess the application of AI in dementia biomarker research, exploring the current trends and future directions of combining various biomarkers, such as imaging, genetics, blood, and digital biomarkers, with AI algorithms.	Our review further explores the multidimensional landscape of digital biomarkers in AD, uncovering the emerging trends of various digital biomarkers. Additionally, it offers a comprehensive review of 86 AI models in this field.

inclusion of 11 countries) have collectively contributed 450 studies (74.75% of the total). These countries also account for 8,358 citations (70.96%) and 627 institutions (68.75%). Among these high-output countries, 5 are in Europe, 3 in Asia, 2 in the Americas, and 1 in Oceania. The United States ranks first in publication output, citation count, collaborative countries, number of institutions, and gross domestic product (GDP). Notably, all of these high-output countries are ranked among the top 15 by GDP, as detailed in Table 3.

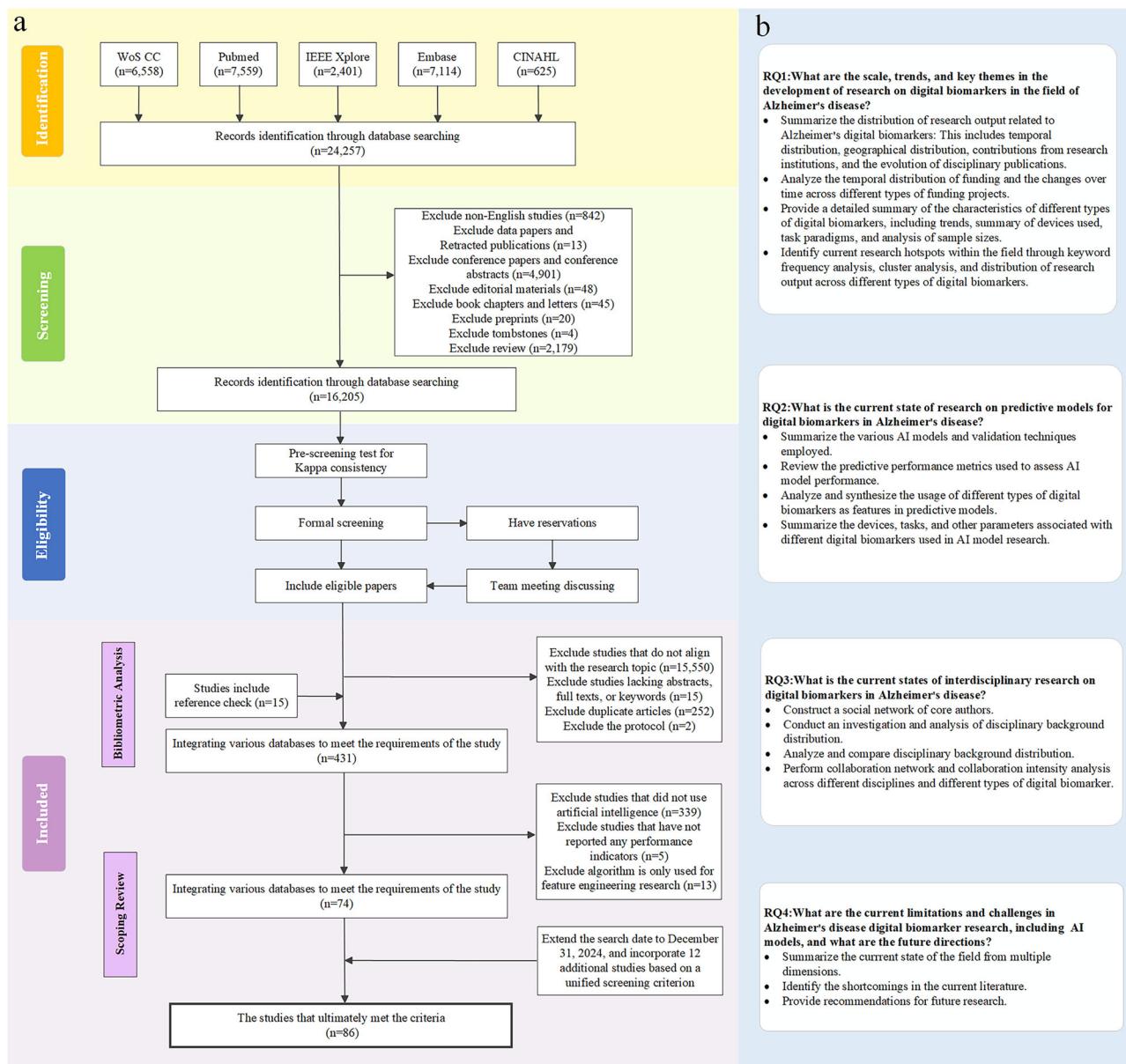
There are notable differences in the timeline of relevant research among these high-output countries. The United States was the first to initiate research on AD digital biomarkers and also has the longest research span. Japan and Australia gradually initiated relevant research since 2007, while China entered the field in 2014. By May 2024, it had already risen to second place in global output. Since 2018, nearly all high-output countries have consistently published research each year, as shown in Fig. 4a. Globally, 20 European countries have participated in digital biomarker research for AD diagnosis, with their output significantly surpassing that of most other regions. In contrast, African countries have exhibited comparatively less interest in this topic, with only two countries contributing a small number of studies, as detailed in Fig. 4b.

Regarding global cross-national collaboration, the observed frequency and intensity of collaboration are below anticipated levels. The United States

has the highest frequency of cross-national collaborations, with 93 spanning 28 countries (28/44, 63.64%). The UK follows with 49 collaborations across 21 countries. On the other hand, Asian countries like China, Japan, and South Korea have fewer cross-national collaborations, with 15, 15, and 5 collaborations, respectively. These collaborations span fewer countries: 4 (4/44, 9.09%), 9 (9/44, 20.45%), and 5 (5/44, 11.36%) countries, respectively, as shown in Fig. 4c.

### Disciplinary publication patterns

Analyzing publication patterns within a discipline can reveal the research cycles and emerging trends in the field. To date, journals from 54 disciplines have published on this topic. Specifically, between 2004 and 2012, 20 disciplines were involved, with publications primarily concentrated in the fields of neuroscience, clinical neurology, psychiatry, and geriatrics, as shown in Fig. 5a. From 2013 to 2018, the number of disciplines grew to 35, with neuroscience, clinical neurology, and geriatrics remaining the dominant fields. However, more engineering and interdisciplinary journals, such as those in medical informatics, computer science, mathematics, and computational biology, began to accept relevant research, as shown in Fig. 5b. Between 2019 and 2024, the number of participating disciplines further expanded to 46. While neuroscience and clinical neurology remain continue to be core disciplines, medical informatics and health care sciences and



**Fig. 1 | Literature retrieval and inclusion process flowchart.** **a** Flow diagram illustrating the PRISMA approach for the identification, screening, and selection of studies. **b** Specific research content.

services have gradually emerged as key areas for publication, as shown in Fig. 5c.

Figure 6 shows that among the top 10 disciplines with the most publications on digital biomarkers, Neurosciences, Clinical Neurology, and Geriatrics & Gerontology are the highest-yielding fields. Notably, Neurosciences has consistently accounted for over 20% of the publication output of the top 10 disciplines each year. Clinical Neurology and Geriatrics & Gerontology have also consistently contributed more than 10% of the publication output each year, and these three fields have maintained continuous publication activity over the past 20 years. Psychiatry has also demonstrated steady publication activity since 2010, though its research output has remained relatively low. It is also noteworthy that over the past five years, Medical Informatics has steadily contributed about 10% of the top 10 disciplines' publication output, indicating some potential for growth in this field.

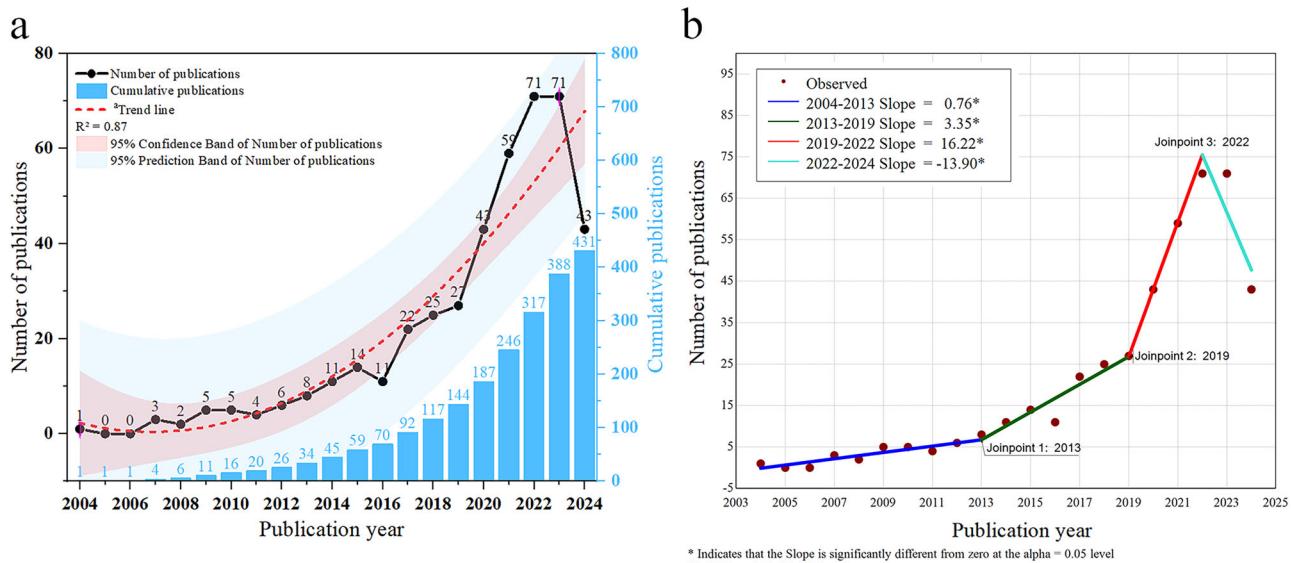
### Funding analysis

An analysis of funding sources reveals the distribution trends of research grants. A total of 350 studies (350/431, 81.21%) have received funding, with

a cumulative 1,345 funding instances, averaging 3.84 funding sources per funded study. These funding projects were supported by 539 different sources. Among the top ten funding departments or agencies, three belong to U.S. government bodies, with the National Institutes of Health being the most frequent funder, providing 159 grants (11.82%), as shown in Table 4.

From 2004 to 2011, the annual funding instances was relatively low, with fewer than 10 per year. Since 2019, the number of annual funding instances has increased significantly, reaching a peak of 256 in 2022 (19.03%, 256/1,345). The fitted trend line for annual funded projects indicates a growing trend in funding for this field ( $R^2 = 0.76$ ). Additionally, the average number of funding occurrences per year in this field exhibits a fluctuating growth trend. Before 2016, the average number of funding instances fluctuated significantly, peaking at 5.67 in 2012 before quickly dropping to 2.38. However, over the past 20 years, the average number of funding instances per study has shown an overall upward trend, stabilizing around 3.5 after 2020, as shown in Fig. 7a.

We categorized the funding sources into government funding, non-profit organizations and foundation funding, corporate funding,



international organizations funding, university and research institution funding, and personal funding. Government departments were the primary source of funding, with 224 projects (41.6%), providing 888 funding instances (66.0%), averaging 3.96 instances per project. Nonprofit organizations and foundations followed, with 135 different projects (25.1%) providing 221 instances (16.4%), averaging 1.64 instances per project. Notably, although universities and research institutions are the main research entities, their funding was lower than expected, with only 82 different projects and 92 funding instances, averaging 1.12 instances per project. Additionally, we found a minimal number of personal funding instances in this field, as shown in Fig. 7b, c.

### Keyword analysis

Keyword analysis reveals the research trends, hot topics, and technological advancements in the field of digital biomarkers. After data cleaning, a total of 897 keywords were obtained, appearing 2332 times in total. According to Price's Law<sup>24</sup>, keywords with a frequency of nine occurrences or more are considered high-frequency keywords in this study. A total of 33 high-frequency keywords were identified, appearing 1,146 times in total, which accounts for 49.14% of all keyword occurrences. The top three most frequent keywords are all related to AD: "Alzheimer's disease" (165 occurrences, 7.08%), "Mild Cognitive Impairment" (164 occurrences, 7.03%), and "Dementia" (122 occurrences, 5.23%). Following these were "Gait analysis" (76 occurrences, 3.26%) and "Machine learning" (49 occurrences, 2.10%). The distribution of high-frequency keywords is provided in Supplementary Table 2. In terms of maturity, the number of keywords began to increase gradually in 2016, with almost all high-frequency keywords showing progressive maturation. By around 2020, most began transitioning to orange or red shades, indicating a significant increase in research activity. Notably, the term "digital biomarkers" as a standardized keyword did not appear until 2020, despite related digital biomarkers such as gait analysis, eye movements, and smart homes emerging in earlier periods, as shown in Fig. 8a.

The co-occurrence and clustering of the keywords resulted in five major clusters:

#1 Red Cluster (Eye Movement and Cognitive Tracking Technologies): includes keywords such as eye tracking, eye movements, reading, anti-saccade, oculomotor behavior, and attention.

#2 Green Cluster (Gait Monitoring and Analysis Technologies): includes keywords such as foot, doppler radar, feature extraction, sensors, machine learning, task analysis, monitoring, and early detection.

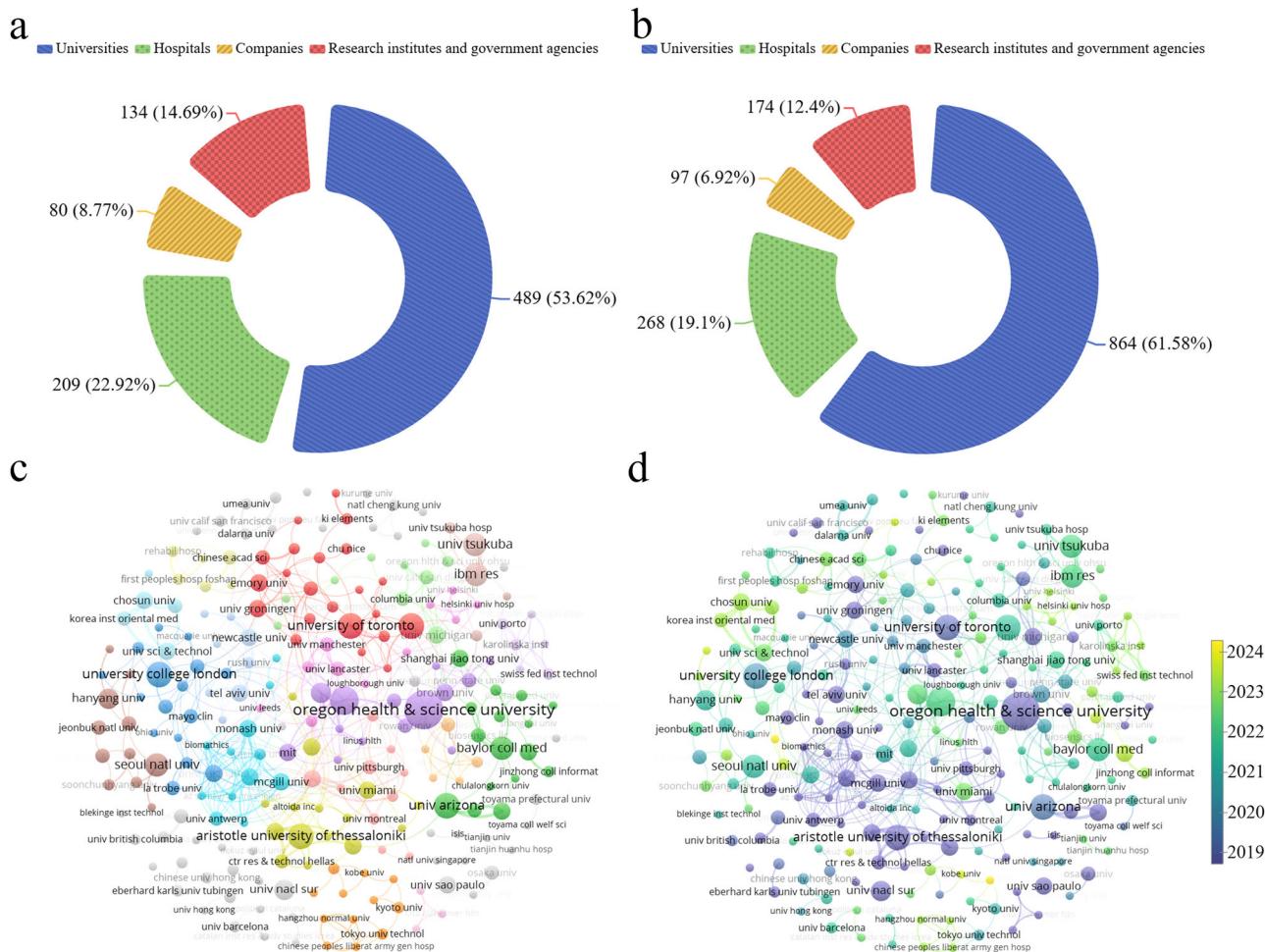
#3 Blue Cluster (Home Activity Behavior and Monitoring Technologies): includes keywords such as smart homes, remote monitoring, wearable devices, and technology.

#4 Light Blue Cluster (Cognitive Aging and AI-Assisted Behavioral Assessment Technologies): includes keywords such as kinematics, trajectory, assessment, naturalistic driving, artificial intelligence, cognitive test, and cognitive aging.

#5 Yellow Cluster (Digital Speech and Cognitive Analysis Technologies): includes keywords such as speech, voice, digital technology, episodic memory, language, and natural language processing. These clusters are shown in Fig. 8b.

### Trends in various types of digital biomarkers

Analyzing the output trends of different types of digital biomarkers helps reveal the development trends and potential of research and applications. The 431 studies covered 11 different types of digital biomarkers, with significant differences in the number and trends of studies across these categories. Research on limb movement digital biomarkers was the most prevalent, totaling 134 studies (31.1%), with a rapid increase after 2015, peaking at 25 studies in 2022. The second most researched category was digital assessments using mobile or dedicated ICT devices, with 120 studies (27.8%). This category saw a rapid increase after 2016 and remained steady with over 15 studies annually after 2021. Eye movement biomarkers and speech biomarkers also showed considerable research activity, accounting for 12.07% and 7.89% of the total, respectively. Speech biomarkers experienced rapid growth after 2019, while eye movement biomarkers exhibited a fluctuating upward trend. Home activity biomarkers and multi-modal biomarkers were less represented, with 27 and 24 studies, respectively. Multi-modal biomarkers have shown slow but fluctuating growth since 2018. Other categories, such as natural driving behavior, biofeedback or physiological signals, and sleep patterns, had relatively fewer studies, but their diversity highlights the broad and evolving nature of current digital health research. Overall, research on digital biomarkers has significantly increased in recent years, particularly in the areas of limb movement and eye movement



**Fig. 3 | Institutional Collaboration in Alzheimer's Disease Digital Biomarkers Diagnostic Research.** **a** Distribution of institution types involved in Alzheimer's disease digital biomarkers diagnostic research. **b** Output by institution type in Alzheimer's disease digital biomarkers diagnostic research. **c** Network diagram of institutional collaboration. Color coding is used to display clusters, with institutions

within the same cluster sharing the same color. The size of the circles increases with the number of publications. **d** Evolution of institutional collaborations over time. Color coding is used to represent the average time for constructing institutional collaboration networks. The size of the circles increases with the number of publications.

**Table 2 | Basic Information on the Top Ten Most Productive Institutions in Alzheimer's Disease Digital Biomarkers Diagnostic Research (Including 11 Institutions). (\*CPP: citations per paper)**

Rank	Organization	Output (N = 1403, n%)	Citations (N = 24,908, n%)	*CPP	Country
1	Oregon Health & Science University	13 (0.93)	492 (1.98)	37.85	United States
2	Boston University	9 (0.64)	57 (0.23)	6.33	United States
3	Aristotle University of Thessaloniki	8 (0.57)	254 (1.02)	31.75	Greece
4	University of Arizona	8 (0.57)	192 (0.77)	24.00	United States
5	University College London	8 (0.57)	241 (0.97)	30.13	United Kingdom
6	University of Toronto	8 (0.57)	190 (0.76)	23.75	Canada
7	Washington University	8 (0.57)	163 (0.65)	20.38	United States
8	Baylor College of Medicine	7 (0.50)	54 (0.22)	7.71	United States
9	IBM Research	7 (0.50)	52 (0.21)	7.43	Not available
10	Seoul National University	7 (0.50)	56 (0.22)	8.00	South Korea
11	University of Tsukuba	7 (0.50)	53 (0.21)	7.57	Japan

biomarkers. With technological advancements and the growing accessibility of devices, these biomarkers are expected to play an increasingly important role in disease monitoring and early diagnosis in the future, as shown in Fig. 9.

#### Sample size analysis

An analysis of the sample size provides insight into the scale of research on different digital biomarkers. As shown in the violin plot, most studies across all types of digital biomarkers tend to have relatively small sample sizes.

**Table 3 | Basic Information of the Top Ten Most Productive Countries in Alzheimer's Disease Digital Biomarkers Diagnostic Research (Including 11 Countries)**

Rank	Country	Output (N = 602, n%)	Citations (N = 11,779, n%)	Organization (N = 912, n%)	Partner countries (N = 44, n%)	<sup>a</sup> 2024GDP Rank	United Nations region
1	United States	135 (22.43)	3,296 (27.98)	164 (17.98)	28 (63.64)	1	Americas
2	China	60 (9.97)	481 (4.08)	96 (10.53)	4 (9.09)	2	Asia
3	United Kingdom	41 (6.81)	788 (6.69)	36 (3.95)	21 (47.73)	6	Europe
4	Japan	41 (6.81)	492 (4.18)	63 (6.91)	9 (20.45)	4	Asia
5	South Korea	36 (5.98)	225 (1.91)	48 (5.26)	5 (11.36)	14	Asia
6	Canada	30 (4.98)	923 (7.84)	48 (5.26)	13 (29.55)	10	Americas
7	Spain	28 (4.65)	491 (4.17)	49 (5.37)	18 (40.91)	15	Europe
8	France	25 (4.15)	707 (6.00)	37 (4.06)	16 (36.36)	7	Europe
9	Germany	20 (3.32)	399 (3.39)	32 (3.51)	10 (22.73)	3	Europe
10	Australia	17 (2.82)	308 (2.61)	14 (1.54)	11 (25.00)	13	Oceania
10	Italy	17 (2.82)	248 (2.11)	40 (4.39)	16 (36.36)	9	Europe

<sup>a</sup>2024GDP Rank: Gross Domestic Product.

However, studies involving mobile devices or dedicated ICT devices, as well as physical movement biomarkers, have lower medians but still offer greater potential for larger sample sizes, with some studies involving over 2,000 participants. The sample sizes in studies on multimodal biomarkers and non-specialized ICT show considerable variability, with some studies involving more than 1,000 participants, suggesting the potential for expanding sample sizes. In contrast, studies on sleep and natural driving behaviors typically have smaller sample sizes, which may be related to the greater difficulty in collecting such data on a large scale. These differences suggest that there may be significant variations in the application scenarios and data collection methods for different types of digital biomarkers. There is still potential for improvement in both sample sizes and data quality across various study types, as shown in Fig. 10.

#### The devices and paradigms commonly used in the collection of different digital biomarkers

There are differences in the collection devices used for various digital biomarkers. In the collection of physical activity data, different types of movements, such as finger movements, gait, and upper limb activities, tend to rely on specific devices. Gait monitoring is often performed during single- or dual-task gait tests, where spatiotemporal features, variability characteristics, and other data are collected using wearable devices and electronic road systems. Additionally, an increasing number of technological devices have been introduced into gait research, including cameras, radar, force plates, motion capture systems, and pressure-sensitive shoes. These devices provide multidimensional data beyond traditional spatiotemporal gait measures, such as postural features, radar-based time and frequency domain data, and pressure characteristics<sup>25,26</sup>. Finger movements are often more refined compared to other types of movements. These are typically measured using digitizers and digital pens, with tasks such as drawing a digital clock or performing the Trail Making Test to quantify motion trajectories and pressure parameters<sup>27,28</sup>. Upper limb movements usually rely on wearable technologies to capture features in activities such as elbow flexion-extension or drumming, for instance, average arm speed and elevation angles<sup>29</sup>. Eye movement biomarkers are primarily collected through eye trackers during tasks involving fixation and saccades to capture eye movement characteristics. However, traditional eye tracking devices suffer from limited portability and high costs, which restricts their widespread use. Researchers are developing more portable solutions, such as integrating eye trackers into the VIVE Pro Eye headset to record eye movement features<sup>30</sup>. Other devices, such as eye-tracking glasses and portable eye trackers, are also employed for precise data collection<sup>31,32</sup>. Speech biomarkers are collected via microphones during tasks such as picture descriptions and spontaneous

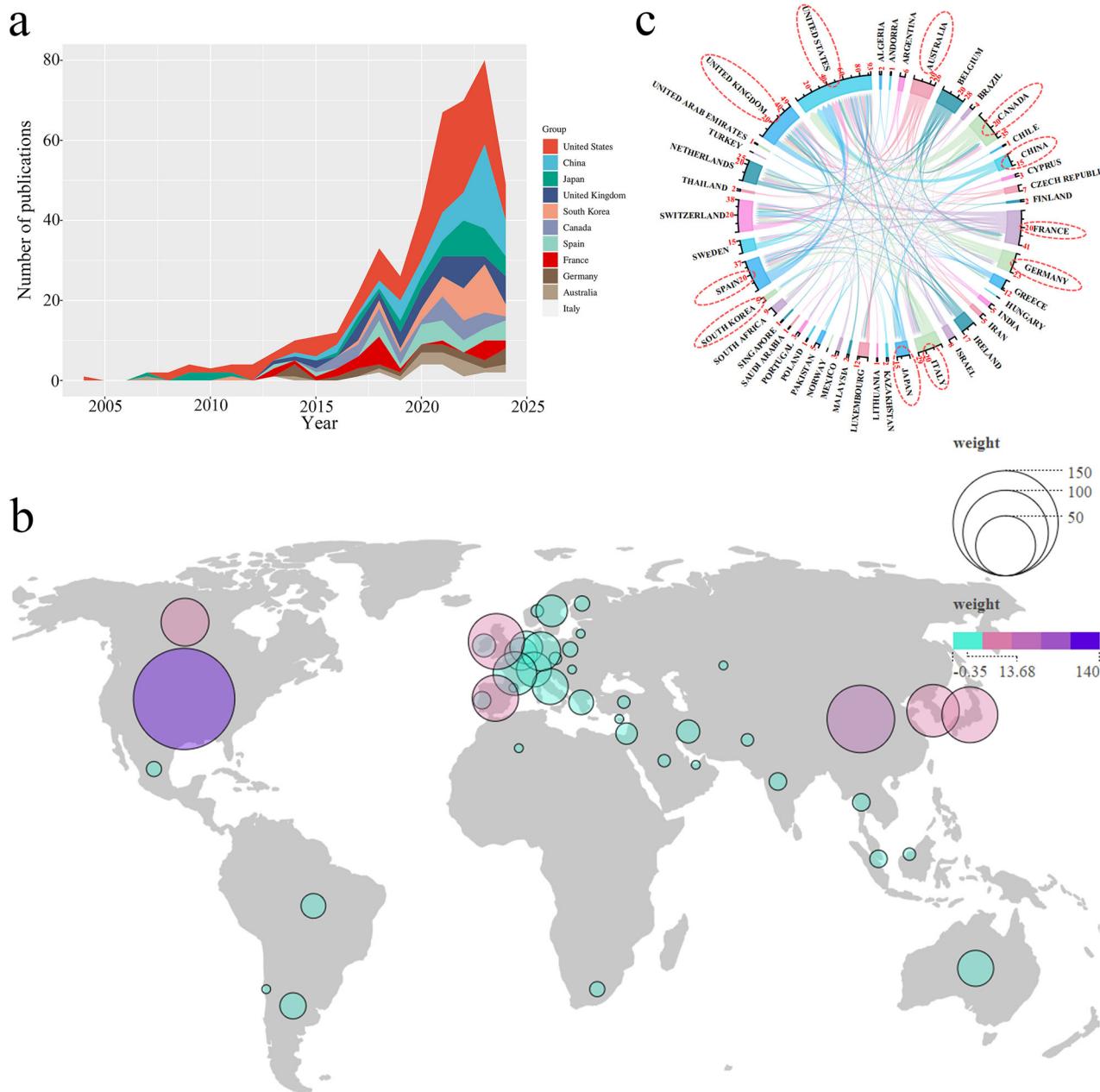
speech, and analyzed using natural language processing techniques. These technologies extract auditory and linguistic features, generating structured data to aid in the diagnosis of AD<sup>21</sup>. In sleep biomarker research, while polysomnography remains the gold standard for assessing sleep physiology, it is unsuitable for long-term, non-invasive, and naturalistic early AD detection or preventive studies. More practical alternatives, such as wristbands and activity monitors, have proven effective as monitoring tools<sup>33</sup>. It is worth noting that tests based on information and communication technology or mobile devices mainly rely on smartphones, tablets, computers, and virtual reality devices for data collection. Although the variety of devices is relatively limited, they encompass a broad range of emerging measurement paradigms and variants, offering flexibility to accommodate different testing needs<sup>34–36</sup>. In-home activity monitoring currently mainly relies on embedded sensors to capture patients' behaviors in their home environment. Devices used in driving behavior and physiological signal studies are relatively uniform: driving behavior is mostly recorded using GPS loggers, with some studies employing camera systems, while physiological signals are predominantly captured using portable EEG systems to record electroencephalographic data. Common devices used for different types of measurements and the associated tasks are outlined in Fig. 11.

#### Author analysis

Collaboration network analysis can reveal key information such as core authors and collaboration patterns within the academic community. A total of 2571 researchers were identified, who collectively collaborated on 3185 studies, with an average of 5.84 co-authors per study. Only 3 studies were independently completed by a single researcher. According to Price's Law, the publication threshold for core authors is approximately 3 papers. A total of 121 authors (4.71%) met this criterion, collectively contributing 460 papers (460/3,185, 14.44%). However, this falls short of the Price's Law requirement, which stipulates that core authors should account for over 50% of the total publications. Overall, the co-occurrence network among core authors shows relatively independent clusters with limited connections between them, indicating strong collaboration within clusters but minimal collaboration across clusters, as shown in Fig. 12a. In terms of collaboration timing, the cooperation network among core authors began to form between 2019 and 2024, with the peak of collaboration activity occurring between 2021 and 2022. Although some earlier collaborations exist, these primarily involved highly productive authors, as shown in Fig. 12b.

#### Changes in multidisciplinary participation

We summarized the disciplinary backgrounds of the authors involved in the research and analyzed the number of different disciplines participated in



**Fig. 4 | Productivity and Collaboration Among Nations in Alzheimer's Disease Digital Biomarkers Diagnostic Research.** **a** Temporal distribution of output by high-producing countries. **b** Chord diagram of international collaborations among countries. **c** Geographic distribution map of publication by country in Alzheimer's

disease digital biomarkers diagnostic research. The colors representing countries/regions have no specific meaning; only the thickness of the lines between them is significant, indicating the frequency of collaboration between different countries. The thickness of the lines corresponds to the values on their respective axes.

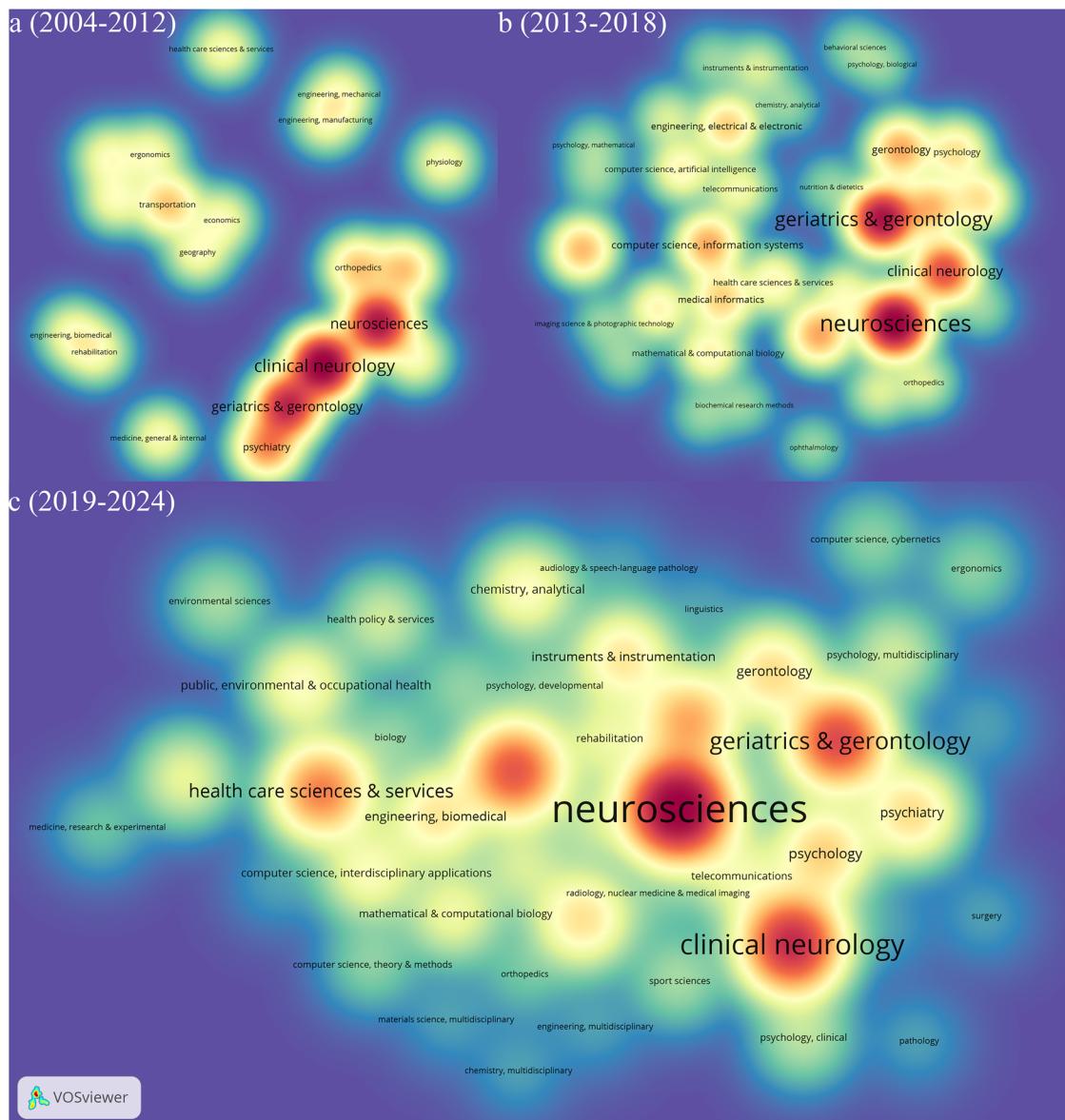
each study. Overall, medicine disciplines had higher participation than engineering-related ones. Specifically, neurology had the highest participations, with 1417 instances (1417/3,185, 44.49%), followed by other medical disciplines with 480 instances (480/3,185, 15.07%). Computer science or communication engineering accounted for 367 (367/3,185, 11.52%). Geriatrics had the fewest participations, with only 60 instances (60/3,185, 1.88%). From the perspective of participation rate, neurology had the highest annual average participation rate at 0.60, followed by other medical disciplines at 0.31, and computer and communication engineering at 0.25.

In terms of yearly changes, the participation rate for medical-related disciplines fluctuated significantly. Neurology consistently had the highest participation rate, staying above 0.5 each year, with fluctuations stabilizing after 2018. Other medical disciplines showed greater variation in participation rates and have yet to establish a stable trend. Geriatrics, psychology,

and psychiatry had smaller fluctuations, but their annual participation rates remained low, with psychology exceeding 20% only in 2011 and psychiatry surpassing 20% in 2022 and 2024, as shown in Fig. 13a. Participation in engineering and other disciplines also fluctuated, particularly before 2019. However, after 2020, the fluctuations in participation rates for these disciplines began to decrease, showing an upward trend, as shown in Fig. 13b.

#### Differences in disciplinary participation in various types of digital biomarkers for Alzheimer's diagnosis

There are notable differences in disciplinary participation across different types of digital biomarkers for Alzheimer's diagnosis. Neurologists have extensively participated in research on all types of digital biomarkers. Professionals from other medical fields, biomedical or medical engineering, have a broad presence across different studies. Researchers from computer



**Fig. 5 | Heatmaps of Disciplinary Publications in Alzheimer’s Disease Digital Biomarkers Diagnostic Research Over Three Time Periods. a 2004–2012. b 2013–2018. c 2019–2024.** Color indicates the popularity of a discipline during the given time period, with redder colors representing higher popularity and dominance.

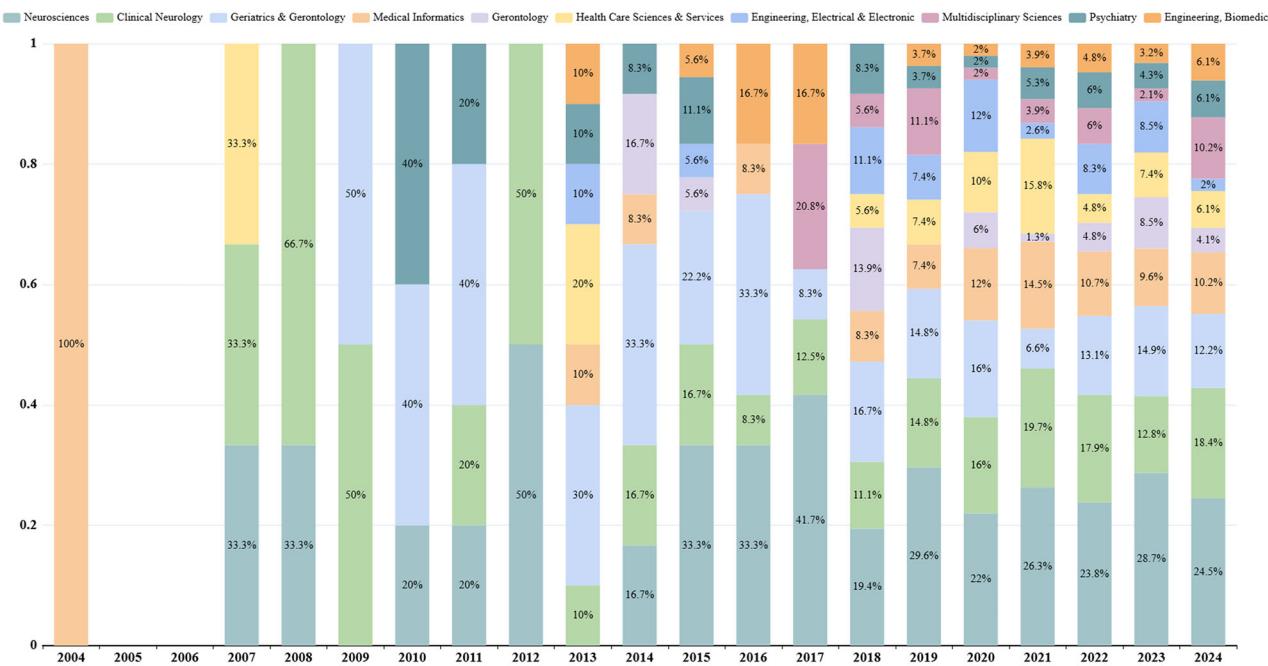
or information engineering, and other engineering fields, also had high coverage, participating in 10 out of 11 types of studies (90.9%). In contrast, professionals from geriatrics and psychiatry were involved in fewer types of research, with limited participation in areas such as natural driving, sleep, and physiological signals. From the perspective of the different types of biomarker research, 7 out of the 11 types of digital biomarker studies showed incomplete disciplinary participation. Home activity and multi-modal biomarker research had relatively high participation rates (8/9, 88.89%), while sleep and other types of research had the lowest disciplinary participation rates (4/9, 44.44%). Comparatively, studies on limb movement, eye movement, speech, and digital testing using mobile or ICT devices were the most comprehensively covered by various disciplines, as shown in Fig. 14.

#### Interdisciplinary collaboration analysis

Interdisciplinary collaboration analysis can effectively reveal how these collaborations drive digital biomarker research and optimize future research directions. We analyzed the interdisciplinary collaboration patterns between various fields. Overall, psychiatry and geriatrics showed limited collaboration with other disciplines. Bioinformatics or

medical engineering mainly collaborated with neurology and computer or communication engineering, but had fewer partnerships with other disciplines. Psychology exhibited more collaboration with computer or communication engineering and neurology, as well as other medical disciplines. Neurology maintained strong collaborative relationships across disciplines, especially with computer or communication engineering and other engineering fields. Computer or communication engineering primarily collaborated with biomedical or medical engineering, neurology, and psychology, with less collaboration with other disciplines. Other engineering fields tended to collaborate with neurology, while other medical disciplines were more inclined to work with psychiatry, geriatrics, psychology, neurology, and other medical fields, but showed weaker collaboration with engineering-related fields, as illustrated in Fig. 15.

In the collaborative network analysis of the four most widely studied biomarkers, the field of neurology continues to demonstrate significant collaborative interest, although its primary collaborators differ slightly. In research on motor biomarkers, neurology predominantly collaborates with other medical disciplines. In contrast,



**Fig. 6 | Temporal Patterns of Disciplinary Publications in Alzheimer’s Disease Digital Biomarkers Diagnostic Research.** The different colors of the bars represent different disciplines, and the length of each bar indicates the proportion of output from a particular discipline during the specified period.

**Table 4 | Top Ten Funding Sources for Alzheimer’s Disease Digital Biomarkers Diagnostic Research**

Rank	Grant	Frequency (N = 1345, n%)	Country
1	National Institutes of Health (NIH)	159 (11.82)	United States
2	National Institute on Aging (NIA)	87 (6.47)	United States
3	Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (JSPS KAKENHI)	34 (2.53)	Japan
4	Engineering and Physical Sciences Research Council (EPSRC)	20 (1.49)	United Kingdom
5	Canadian Institute Of Health And Research (CIHR)	17 (1.26)	Canada
6	National Science Foundation (NSF)	16 (1.19)	United States
7	Alzheimer’s Association	15 (1.12)	Not available
8	National Natural Science Foundation of China (NSFC)	13 (0.97)	China
9	National Key R&D Program Of China	12 (0.89)	China
10	Alzheimer’s Research U.K.	12 (0.89)	United Kingdom

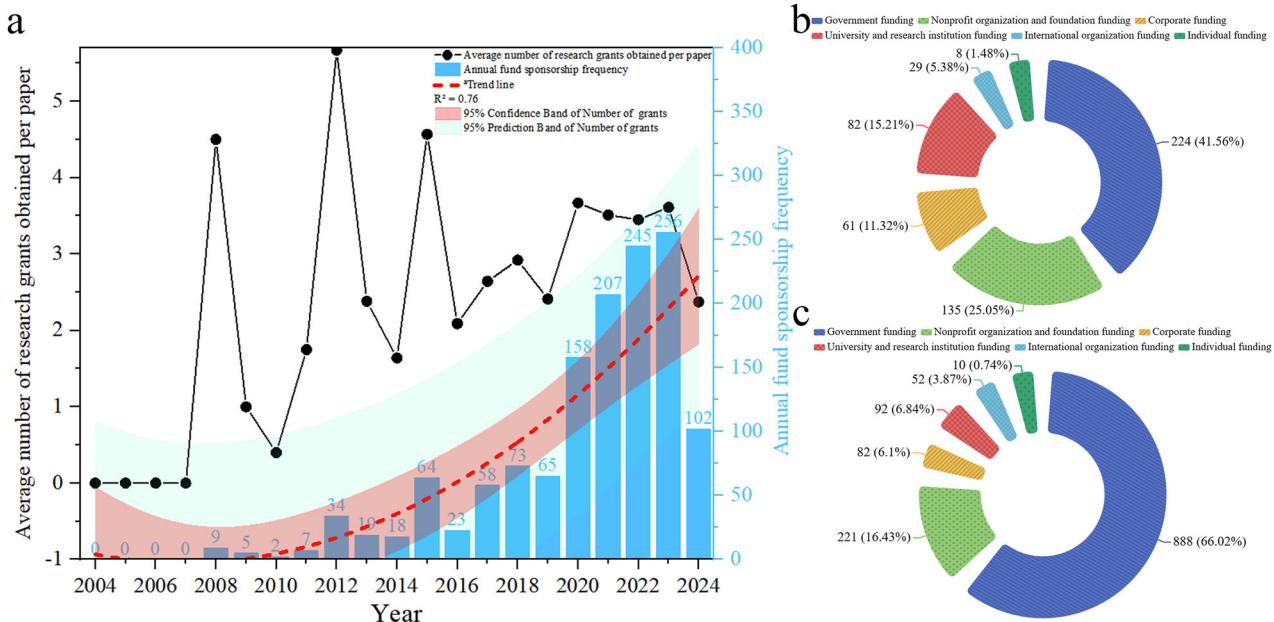
studies on digital detection biomarkers and oculomotor biomarkers show stronger collaboration with fields such as computer science or communication engineering, as well as other medical areas. In the case of language biomarkers, despite the generally limited strong collaboration across disciplines, neurology maintains a close partnership with computer science or information engineering. For further details, please refer to Supplementary Figure 1.

#### Scope definition of artificial intelligence model research

The methods used for build AI models with digital biomarkers share many similarities with those for traditional biomarkers, particularly in model training and validation. However, the key difference lies in data collection methods and the data types of data involved. Traditional biomarkers rely on medical imaging (e.g., MRI or PET), blood tests, and cerebrospinal fluid analysis, whereas digital biomarkers capture behavioral and physiological data through wearable devices, smartphones, tablets, and other digital tools in daily life or specific task paradigms<sup>37,38</sup>. These data collection methods introduce the challenge of handling high-dimensional and unstructured data. For instance, gait data generates multiple data points per second, speech analysis

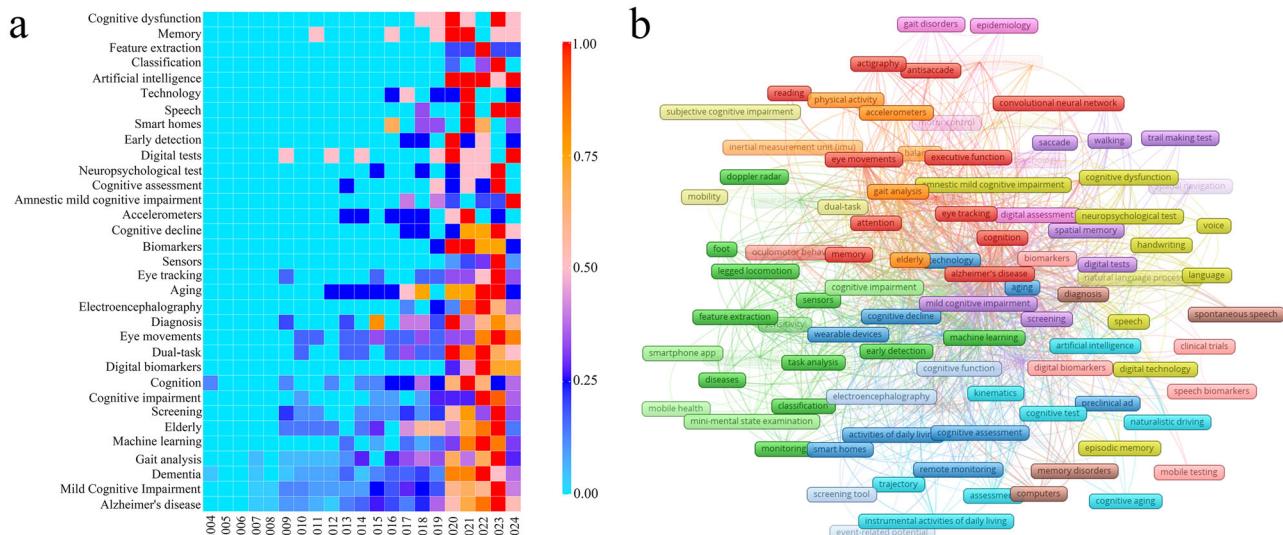
involves spectral features, and handwriting trajectories include various parameters such as speed, pressure, and direction. Traditional statistical methods may struggle to fully extract meaningful information from this data. Thus, AI techniques can process large-scale, complex data, facilitating early diagnosis, personalized treatment, and continuous disease monitoring. The specific method for constructing the digital biomarker AI model is shown in Fig. 16.

The following sections provide a comprehensive overview of the results from our scope-defined review, organized around several key themes identified during the analysis. We first describe the data used in these studies, covering aspects such as data patterns, resources, sample size, data imbalance, and handling of missing data. Next, we discuss the AI-based models and their application methods. This is followed by a detailed presentation of the validation procedures, performance metrics, and comparisons across the studies. We also summarize the distribution of features across different types of digital biomarkers, highlighting their collection methods and the task paradigms employed in AI models. Lastly, we examine the reporting standards and reproducibility of the included studies to evaluate the transparency and reliability of the research.



**Fig. 7 | Types and Quantities of Funding in Alzheimer's Disease Digital Biomarkers Diagnostic Research.** **a** Funding trends in Alzheimer's disease digital biomarkers diagnostic research. The red dashed line represents the trend line

(<sup>a</sup>Trend line:  $y = 0.7712x^2 - 6.0744x + 9.2722$ ), **b**) Distribution of funding types.  
**c** Quantities of each type of funding.



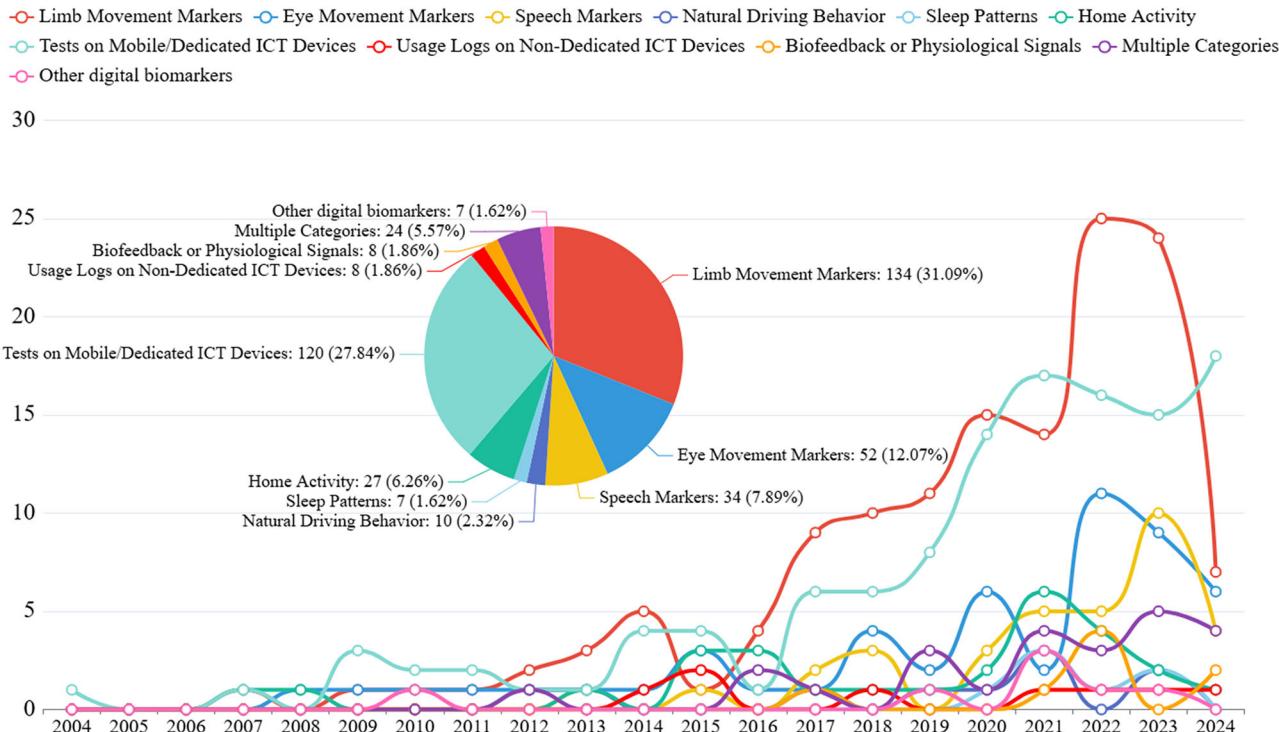
**Fig. 8 | Keywords Analysis in Alzheimer's disease Digital Biomarkers Diagnostic Research.** **a** Temporal heatmap of high-frequency keywords. Color represents the proportion of keyword frequency for that year relative to the total frequency of the

keyword. The more frequent the keyword, the redder the color, indicating a more mature topic. **b** Clustering diagram of keywords co-occurrence. Color coding is used to display clusters, with keywords within the same cluster sharing the same color.

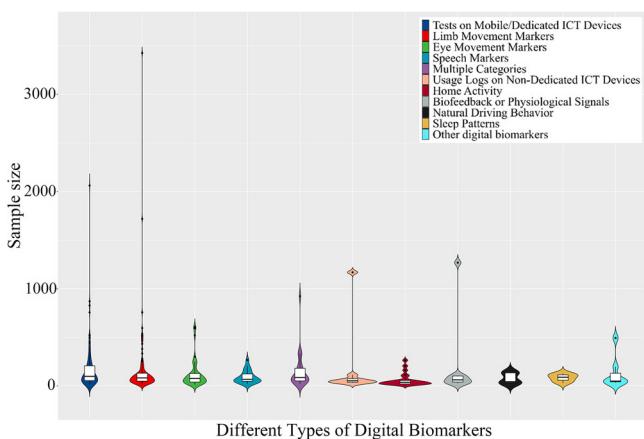
## Distribution of basic research information

A total of 86 studies were included in the analysis. Among these, 36 studies involved AD patients, with a total of 1663 cases, and sample sizes ranging from 5 to 166 participants. 63 studies included patients with MCI, totaling 3869 cases, with sample sizes ranging from 5 to 403 participants. Additionally, some studies included patients with other diseases closely related to AD: 3 studies on dementia with Lewy bodies with 100 participants<sup>27,39,40</sup>; 1 study on vascular dementia with 27 participants<sup>41</sup>; 1 study on subjective cognitive impairment with 56 participants<sup>42</sup>; 1 study on subjective memory complaints with 9 participants<sup>43</sup>; 1 study on PreMCI with 20 participants<sup>44</sup>; 1 study on PreAD with 64 participants<sup>45</sup>; and 1 study

on mixed dementia with 38 participants<sup>42</sup>. In addition, 9 studies reported dementia patients without specifying the exact subtype. Given that AD accounts for the majority of dementia subtypes, and considering that some researchers use the term "dementia" to broadly refer to AD, as well as the representativeness of the digital devices used in these studies, we did not exclude these dementia-focused studies<sup>46–54</sup>. One study<sup>47</sup> mentioned that its dementia group included both AD and Parkinson's disease patients, but did not specify the proportion of each. Furthermore, one study on suspected dementia patients<sup>55</sup> and another that predicted dementia risk in 18 elderly participants<sup>56</sup>, both of which were also included in the analysis.



**Fig. 9 | Production, Distribution and Trends of Various Types of Alzheimer's Disease Digital Biomarkers.** The pie chart is used to represent the proportion of research on different types of digital biomarkers, while the line chart illustrates the trend of changes in various digital biomarkers over time.



**Fig. 10 | Sample Size Distribution in Alzheimer's Disease Digital Biomarkers**

**Research.** The violin plot displays the distribution of sample sizes for research on each type of digital biomarker. The different colors of the violins represent distinct categories of digital biomarker research. The body of the violin represents the primary distribution range of the research, with wider sections indicating a higher number of studies and narrower sections representing fewer studies.

In terms of study design, only 9 studies were prospective studies<sup>45,49,57–63</sup> and 2 were case-control studies<sup>64,65</sup>, and the remaining studies were cross-sectional in design. The publication dates of these studies ranged from 2011 to 2024, with the majority ( $n = 74$ ) published in the past five years. Trendline analysis indicates a clear trajectory in the number of studies, reflecting the increasing interest in using AI models for diagnosing and predicting AD through digital biomarkers, as shown in Fig. 17a. From a geographical perspective, these studies were conducted across 19 different countries in Asia, Europe, Africa, North America, and South America. China contributed the highest number of studies ( $n = 23$ ), followed by the United States ( $n = 16$ ) and Japan ( $n = 9$ ), as shown in Fig. 17b.

### Distribution of digital biomarker research types

Among the included studies, there were:

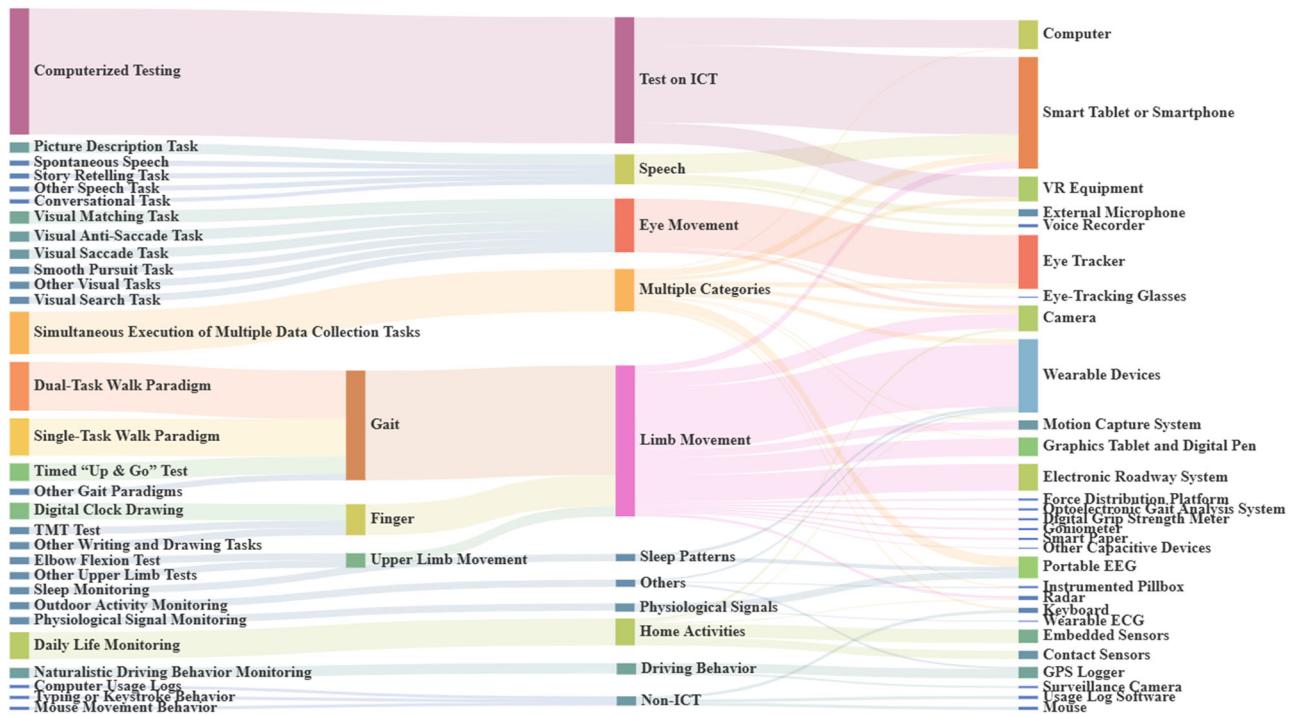
- 13 studies based on gait digital biomarkers<sup>26,46–49,66–73</sup>
- 10 studies based on manual digital biomarkers<sup>27,41,57,58,74–79</sup>
- 11 studies based on eye movement digital biomarkers<sup>39,40,80–88</sup>
- 12 studies based on speech digital biomarkers<sup>42,50,51,59,60,64,65,89–93</sup>
- 13 studies based on ICT device-based digital testing biomarkers<sup>55,59,62,63,104</sup>
- 13 studies based on multi-type digital biomarkers<sup>43,44,61,105–114</sup>
- 6 studies based on home activity digital biomarkers<sup>54–56,62,63,115</sup>
- 5 studies based on physiological signals<sup>116–120</sup>
- 1 study based on non-ICT or dedicated device testing<sup>121</sup>
- 1 study based on other biomarkers<sup>122</sup>
- and 1 study based on driving behavior<sup>45</sup>

We have summarized the basic information and main findings from these studies, which are detailed in Supplementary Tables 3 to 13.

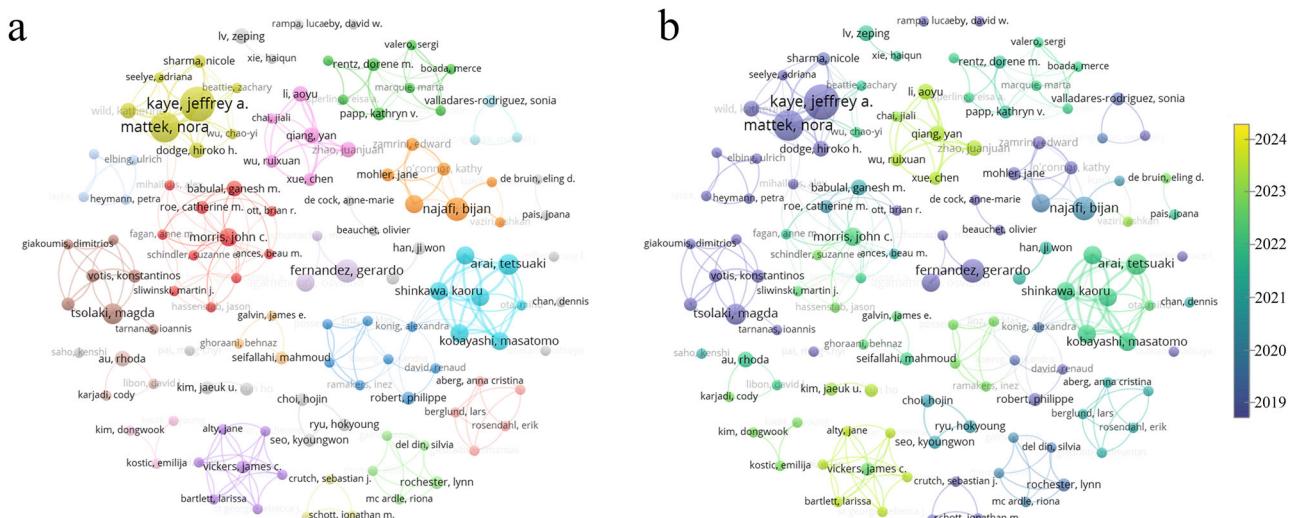
### Specific algorithm usage

In a review of 86 studies on digital biomarkers for AD, researchers employed various machine learning algorithms to enhance the accuracy of disease prediction and classification. Each algorithm differs in its data processing capabilities and model complexity, making the selection of an appropriate classifier crucial to the reliability of study outcomes. Details on the types of algorithms used and the distribution of optimal models can be found in Supplementary Table 14.

Statistical analysis revealed that support vector machines (SVM) were the most commonly used algorithm, appearing in 49 instances. Logistic regression (LR) and random forests (RF) were also widely used, with 38 and 32 instances, respectively. In contrast, simpler models, such as decision trees (DT) and naive Bayes (NB), were used less frequently, with 14 and 7 instances, respectively. Although these simpler models offer advantages in terms of computational cost and interpretability, they are less effectively in handling the high-dimensional, complex data typical of AD, which accounts for their limited usage. Additionally, neural networks (NNs), with their



**Fig. 11 | A Sankey diagram of the devices and paradigms commonly used in the collection of different digital biomarkers.** The left column shows the data collection paradigms, the middle column represents different types of digital biomarkers, and the right column lists the devices used for data collection.



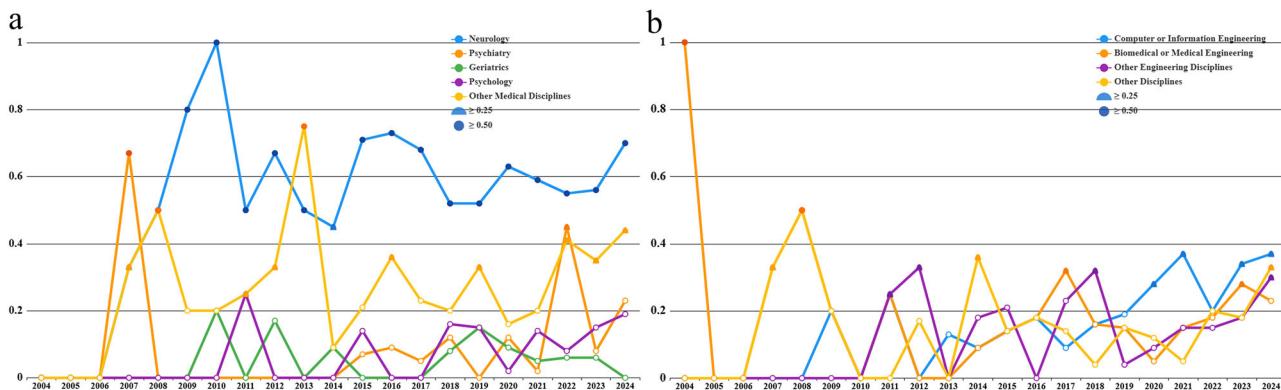
**Fig. 12 | Core Author Collaboration in Alzheimer's Disease Digital Biomarkers Diagnostic Research.** **a** Network diagram of core author collaboration. Color coding is used to display clusters, with researchers within the same cluster sharing the same color. The size of the circles increases with the number of publications. **b** Average

timeline of core author collaboration initiatives. Color coding is used to represent the average time for constructing researcher collaboration networks. The size of the circles increases with the number of publications.

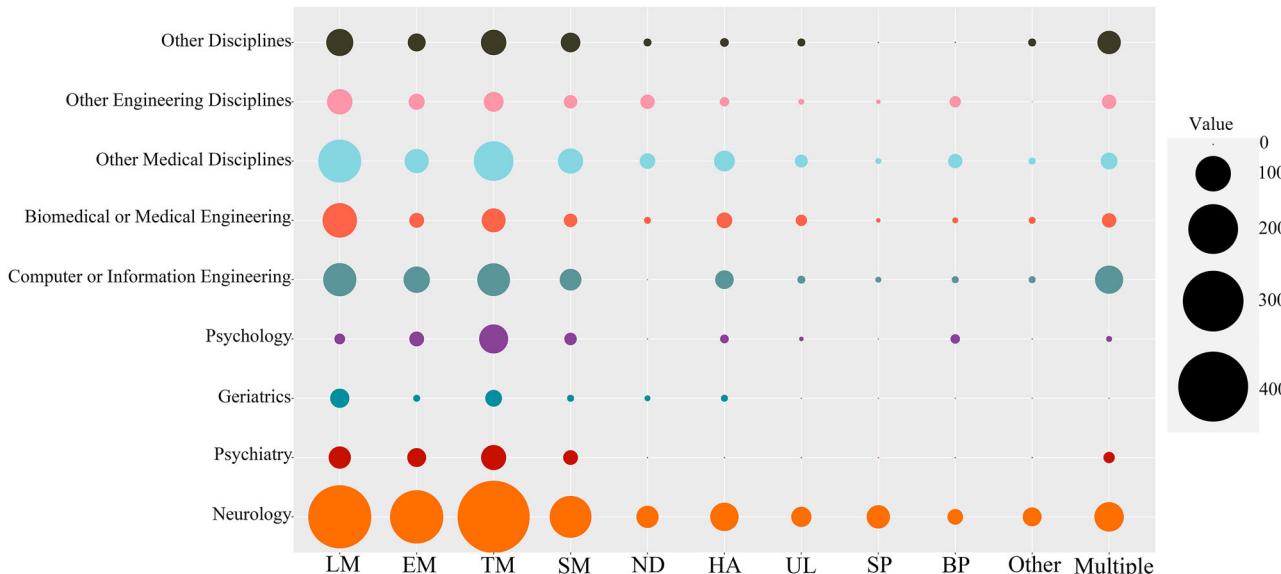
robust capacity for nonlinear feature extraction, were applied 27 times. Ensemble learning models, such as gradient boosting and XGBoost, were used in 11 and 9 times, respectively, demonstrating their potential for handling nonlinearity and high-dimensional data. Further details are provided in Fig. 18a.

Overall, the selection of algorithms in AD digital biomarker research is diverse, reflecting various efforts to enhance model accuracy and robustness. Specifically, SVM has been the most widely applied algorithm across different types of digital biomarker studies. For instance, in speech biomarkers research, SVM has been used 9

times, demonstrating its effectiveness in handling high-dimensional audio data, such as frequency, pitch, and rhythm. By maximizing the decision boundary and selecting the optimal hyperplane, SVM offers robust classification capabilities. LR and RF have also been frequently employed. As a linear model, logistic regression, is simple yet offers strong interpretability. In contrast, random forests improve model robustness and noise resistance by integrating multiple decision trees, making them particularly well-suited for large feature spaces and complex data distributions. Neural networks, while excelling in processing large-scale data, are sometimes limited by their high



**Fig. 13 | Participation Rate of Disciplinary Backgrounds in Alzheimer's Disease Digital Biomarkers Research. a** Annual participation rate changes in medical-related disciplines. **b** Annual participation rate changes in engineering-related disciplines.



**Fig. 14 | Distribution of participants by disciplinary backgrounds in various types of Alzheimer's disease digital biomarkers research.** LM limb movement, EM eye movement, TM Test on mobile or ICT devices, SM Speech markers, ND Natural driving, HA Home activity, UL non-dedicated ICT biomarkers, SP Sleep pattern, BP

Biofeedback or physiological signal, Other Other biomarkers, Multiple Multiple biomarkers.) The size of the circles represents the frequency, with larger circles indicating higher frequencies. Different colors represent different disciplines.

computational complexity and reliance on large amounts of labeled data. Further details are provided in Fig. 18b.

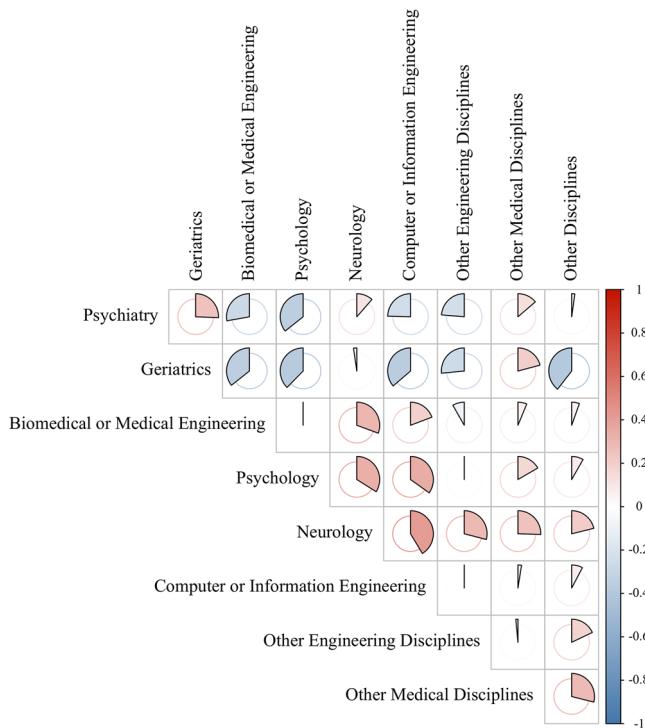
Of the 86 studies included, 44 compared the performance of two or more algorithms. Among them, 17 studies identified SVM as the best classifier, followed by neural networks in 8 studies and logistic regression in 6 studies. Further details are provided in Fig. 18c. Several studies also conducted in-depth comparisons of multiple model architectures. For example, in eye-tracking research, 12 convolutional neural network (CNN) models were compared, with the MC-CNN ultimately selected for classification<sup>80</sup>. In physiological signal studies, 7 models were compared, with the K-nearest neighbors (KNN) model performing the best<sup>116</sup>. For studies focusing on home activity, 8 models were evaluated, and the deep neural network (DNN) achieved the highest performance<sup>56</sup>.

Notably, in specific categories of digital biomarker research, SVM has consistently demonstrated superior performance across multiple domains. For instance, in studies on speech, gait, and eye-tracking, SVM was repeatedly identified as the best model (Fig. 18d). Moreover, some studies have optimized model performance through innovative architecture designs. For example, in eye-tracking research, a deep learning model based

on a nested autoencoder (NeAE-Eye) was proposed, which effectively leveraged eye-tracking data for AD diagnosis<sup>83</sup>. In gait analysis, an adaptive neuro-fuzzy inference system (ANFIS), combining artificial neural networks and fuzzy logic, was employed for classification and prediction<sup>68</sup>. The selection and design of these models various strategies aimed at improving disease prediction accuracy and advancing early diagnosis and personalized treatment.

#### Feature distribution of digital biomarkers

In the process of feature selection, the lack of unified standards and the variety of feature types generated during measurement remain significant challenges in AD digital biomarker research. Similarly, differences between digital biomarker types, measurement devices, and paradigms contribute to variations in data collection, further complicating the research process. All studies reported the types of features they employed. In gait studies, the most frequently used features were spatiotemporal features, employed in 12 out of 13 studies, followed by gait variability features, used in 6 studies. In the 10 manual digital biomarker studies, trajectory and temporal features were most commonly used, appearing in 8 studies, followed by pressure features,



**Fig. 15 | Heatmap of Interdisciplinary Collaboration Among Disciplines.** The size and color of the sector areas represent the strength of collaboration between different disciplines. The redder the color and the larger the sector area, the stronger the collaboration between disciplines.

used in 7 studies. In the 11 eye movement studies, fixation features were most prevalent, found in 8 studies. In the 12 speech biomarker studies, acoustic features were included in 11 studies. Among the 13 studies involving mobile or specialized ICT device testing, task count features were used in 10 studies. In the 5 biofeedback or physiological signal studies, frequency-domain features were used in all of them. Among the 13 multi-type digital biomarker studies, EEG features were included in 6 different studies. In the 6 home activity studies, all employed spatial and activity pattern features. Additionally, studies on natural driving behavior, non-ICT device testing, and other types of digital biomarkers utilized their respective unique features, as shown in Fig. 19 and Supplementary Table 15.

### The performance of AI model

Due to the objective differences in research methods and feature types, we did not directly compare the performance across studies. Instead we presented and described the performance distribution in various tasks of digital biomarker research.

A total of 21 studies conducted binary classification between AD and healthy controls, including 2 from gait studies<sup>69,71</sup>, 6 from manual biomarker studies<sup>27,57,74,77–79</sup>, 7 from eye movement biomarker studies<sup>39,40,80,81,83,85,87</sup>, 1 from ICT device-based digital testing<sup>96</sup>, 2 from multi-type biomarker research<sup>111,114</sup>, 1 from Biofeedback research<sup>120</sup>, 1 from home activity biomarker research<sup>115</sup>, and 1 from other research<sup>122</sup>. In terms of performance, the overall Area Under the Curve (AUC) ranged from 0.76 to 1, with an average AUC of 0.887. Accuracy (ACC) ranged from 0.73 to 1, with an average ACC of 0.911. Sensitivity (SEN) ranged from 0.71 to 1, with an average SEN of 0.909. Specificity (SPE) ranged from 0.57 to 1, with an average SPE of 0.889.

A total of 45 studies performed binary classification between MCI and HC, including 7 from gait studies<sup>47,48,66–68,72,73</sup>, 7 from manual biomarker studies<sup>57,58,74–78</sup>, 1 from eye movement studies<sup>82</sup>, 8 from speech biomarker studies<sup>50,60,64,65,89,91–93</sup>, 6 from ICT device testing<sup>94,96–98,103,104</sup>, 10 from multi-type biomarker studies<sup>61,105–113</sup>, 2 from physiological signal studies<sup>118,119</sup>, 3

from home activity studies<sup>54,62,63</sup>, and 1 from non-ICT device research<sup>121</sup>. In terms of performance, the overall AUC ranged from 0.62 to 0.97, with an average AUC of 0.821. ACC ranged from 0.53 to 1, with an average ACC of 0.825. SEN ranged from 0.46 to 1, with an average SEN of 0.817. SPE ranged from 0.68 to 0.99, with an average SPE of 0.825, as shown in Fig. 20.

Surprisingly, only 7 studies have performed ternary classification between AD, MCI, and HC, coming from gait research<sup>70</sup>, Multiple categories<sup>111</sup>, manual research<sup>74</sup>, eye movement research<sup>84,88</sup>, and ICT device-based digital testing<sup>95,102</sup>. Additionally, 3 studies performed binary classification between AD and MCI, 2 of which were manual biomarker studies<sup>57,78</sup> and 1 was a speech biomarker study<sup>42</sup>. Detailed performance metrics for each study are provided in Supplementary Table 16.

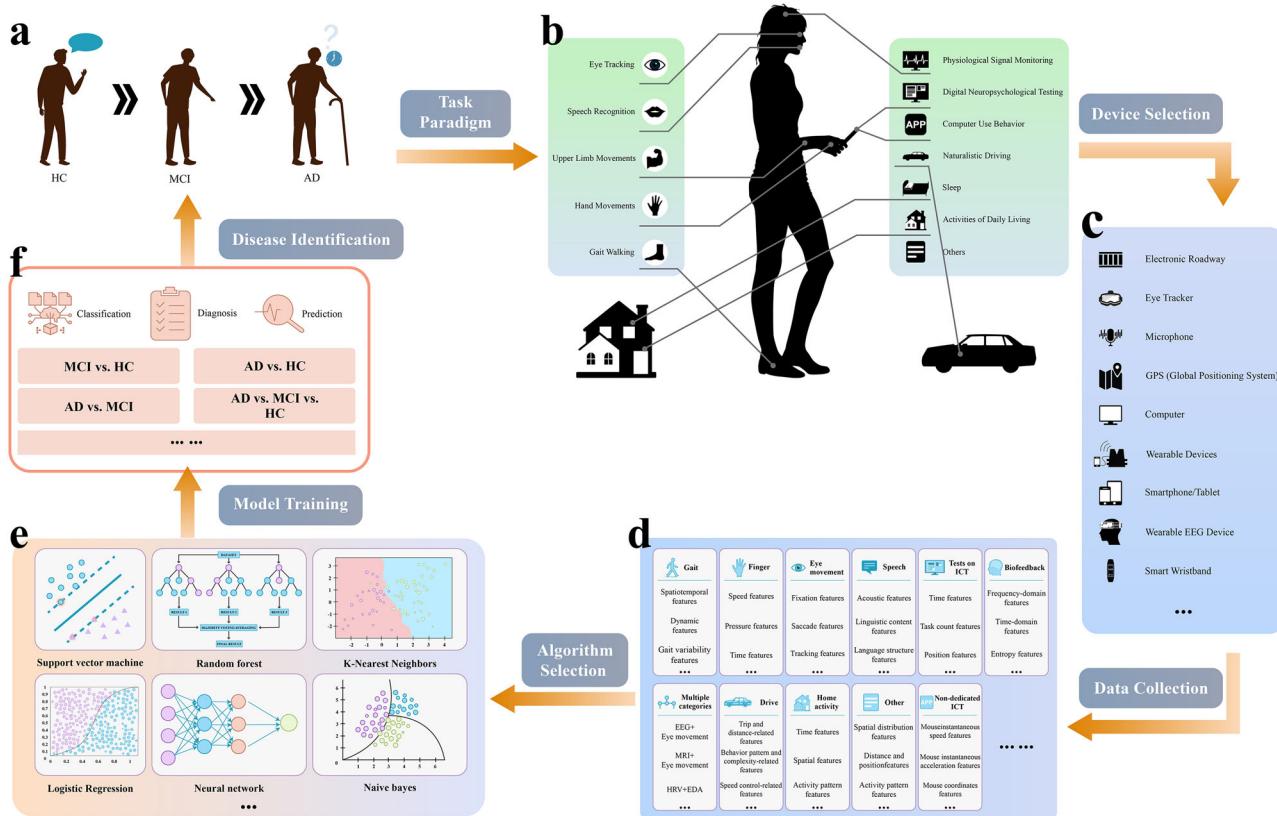
In studies using different digital biomarkers to distinguish between AD and HC, two gait studies<sup>69,71</sup> did not report AUC metrics, but their ACC, SPE, and SEN were all greater than 0.9. Among the six manual biomarker studies<sup>27,57,74,77,78</sup>, five studies reported ACC greater than 0.9<sup>57,74,77–79</sup>, while one study reported an ACC of 0.76<sup>27</sup>. Three studies<sup>77,79</sup> did not report AUC metrics, and two studies<sup>27,77</sup> did not report SEN and SPE metrics. In the seven eye movement biomarker studies, four studies<sup>80,81,83,85</sup> reported ACC, all greater than 0.8. AUC values were reported in four studies, ranging from 0.76 to 0.99<sup>39,40,80,87</sup>. Although various digital biomarkers have demonstrated relatively high accuracy in distinguishing AD from HC, the reporting standards across studies are inconsistent, particularly concerning key metrics such as the AUC, SEN, and SPE. The performance metrics for the remaining categories of digital biomarker studies are shown in Fig. 21.

In studies distinguishing MCI from HC, all seven gait biomarker studies reported specific ACC values<sup>47,48,66–68,72,73</sup>, with all ACCs values greater than 0.7. One study, which used a Kinect-V.2 camera to capture gait metrics and combined with an Adaptive Neuro-Fuzzy Inference System, achieved an ACC of 0.93<sup>68</sup>, while the lowest two studies reported ACCs of only 0.72 and 0.7<sup>66,67</sup>. Only three studies reported specific AUC values, ranging from 0.83 to 0.89<sup>48,66,73</sup>. Six studies reported specific SEN values, all above 0.8<sup>47,48,67,68,72,73</sup>. Among the seven manual biomarker studies, six studies reported specific ACC values, all above 0.7<sup>57,58,74,76–78</sup>. All speech biomarker studies reported specific ACC values, with a wide range. The highest ACC reached 0.95<sup>60</sup>, while the lowest was only 0.53<sup>91</sup>, with the remaining studies reporting ACCs above 0.6. In the ten multi-type biomarker studies<sup>61,105–113</sup>, all studies reported specific ACC values, all of which showed high accuracy, with ACCs above 0.8, as shown in Fig. 21. These results suggest that gait, manual, vocal, and multi-type biomarkers exhibit a certain level of accuracy in distinguishing MCI from HC.

### Sample balance and missing data handling

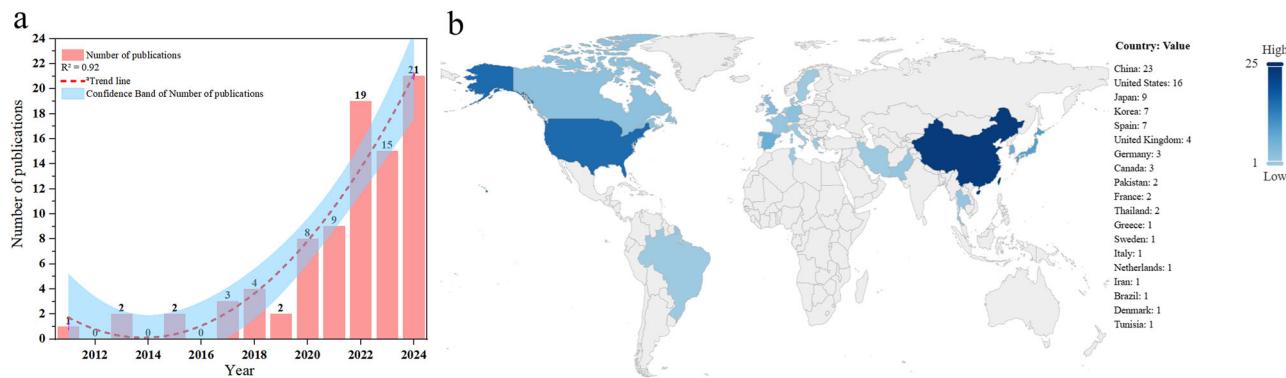
Differences in task accessibility between groups can lead to sample imbalance during data collection. This imbalance can prevent models from effectively learning the characteristics of all groups during training, ultimately impacting model generalization and distorting evaluation metrics. We included 12 studies that reported and employed sample balancing techniques. Five studies<sup>66,73,76,82,103</sup> used methods based on Synthetic Minority Oversampling Techniques (SMOTE), with two studies employing variants: SVMSMOTE<sup>66</sup> and BorderlineSMOTE<sup>103</sup>. Four studies utilized data augmentation methods<sup>79,80,96,121</sup>, one used stratified sampling<sup>27</sup>, one applied cost-sensitive learning<sup>54</sup>, and one study mentioned resampling for sample balancing without specifying the method used<sup>46</sup>.

Due to differences in task accessibility and execution difficulty, missing data is a common challenge in research and can significantly impact study outcomes and conclusions. Despite its importance, only 14 studies reported their handling of missing data. Three studies used listwise deletion<sup>43,55,56</sup>, one of the simplest and most direct methods for addressing missing data, but it can limit the available data for model development and introduce bias if the remaining sample is not representative. Eight studies used mean imputation<sup>52,73,98,99,107,110,111,114</sup>, two studies used Multiple Imputation by Chained Equations (MICE)<sup>27,41</sup>, and one study used zero imputation<sup>65</sup>. Details information on sample balancing and missing data handling methods can be found in Supplementary Table 17.



**Fig. 16 | The Process of Developing AI Models for Alzheimer’s Disease Digital Biomarkers.** **a** Recruiting Alzheimer’s disease patient samples to form the study cohort. **b** Collecting data from multiple devices, including wearables, smartphones, and others. **c** Using sensors, mobile applications, and other tools to acquire various

digital biomarker data. **d** Extracting and analyzing feature data such as movement, speech, eye movement, and physiological signals. **e** Selecting appropriate machine learning and AI algorithms to train predictive models. **f** Applying the models for disease classification and prediction, aiding in early detection and management.



**Fig. 17 | Spatial-temporal distribution of research output.** a Distribution of research years in AI model-based studies on digital biomarkers for Alzheimer's disease diagnosis and prediction. The red dashed line represents the trend line

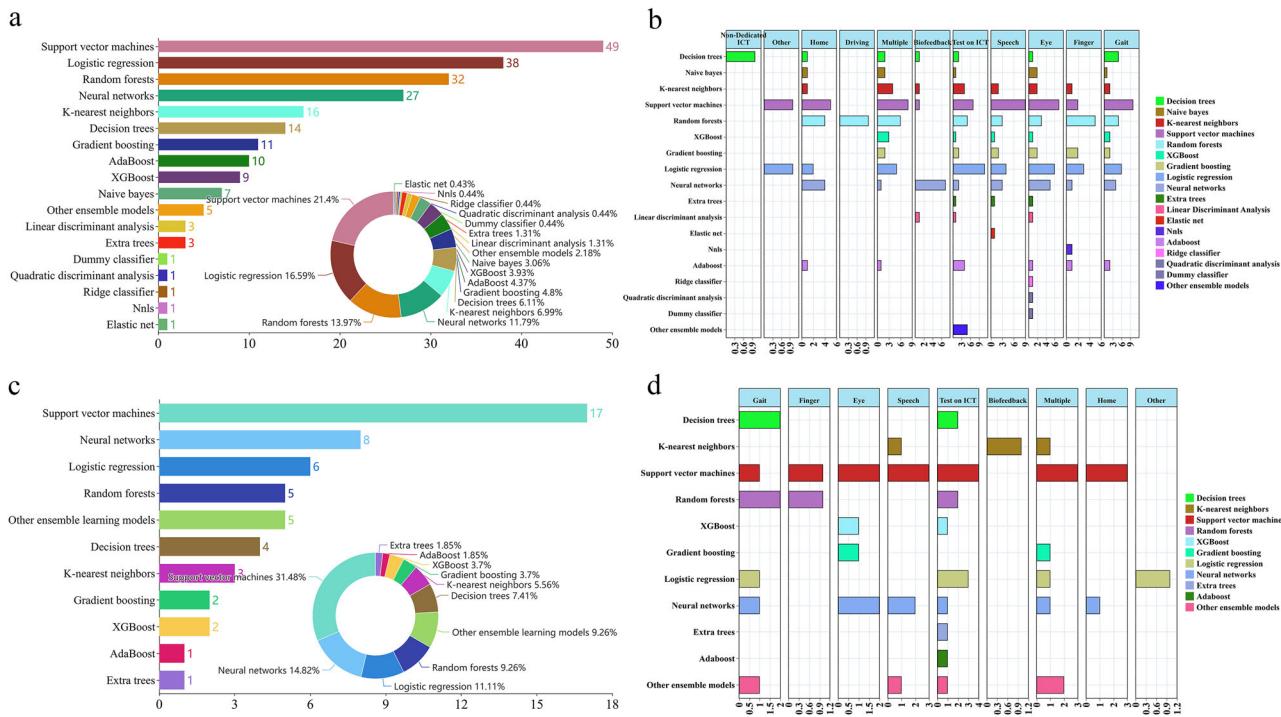
(<sup>a</sup>Trend line:  $y = 0.2026x^2 - 1.549x + 3.0714$ ). **b** Regional distribution of AI model-based studies on digital biomarkers for Alzheimer's disease diagnosis and prediction.

## Fusion of multiple data modalities

Different types of digital biomarkers often complement each other in research. Information fusion strategies in machine learning are generally categorized into four types: early fusion, mid-level fusion, late fusion, and hybrid fusion. Early fusion involves integrating all modalities at the initial stage and inputting the combined data into a single model for training<sup>123</sup>. Mid-level fusion extracts features progressively by using the output of one model as the input for another, allowing for iterative feature extraction and optimization<sup>123</sup>. Late fusion models each modality separately, subsequently integrating the results, often

through weighted averaging or voting mechanisms to generate the final predictions.<sup>123</sup>

Research on various digital biomarkers has demonstrated different fusion strategies: four studies employed early fusion strategies<sup>108,111,113,114</sup>; four studies utilized mid-level fusion<sup>61,105,106,109</sup>; three studies applied late fusion<sup>43,110,112</sup>; and two studies adopted hybrid fusion<sup>44,107</sup>. From the perspective of feature sources, six studies utilized EEG signals collected via portable EEG devices<sup>106–108,110,112,113</sup>. Only one study incorporated traditional imaging biomarkers into the analysis<sup>105</sup>. Furthermore, Yasunori Yamada's study<sup>111</sup> developed a multimodal model based on three prevalent digital



**Fig. 18 | Distribution of AI Model Algorithms Used in Alzheimer's Digital Biomarkers.** **a** Overall Use of AI model algorithms. **b** Distribution of AI model algorithm use across different types of digital biomarkers. **c** Use of AI model

biomarkers: gait, manual movement, and speech. In contrast, Aoyu Li<sup>108</sup> integrated physiological signals, such as electrodermal activity, heart rate variability, and EEG, with cognitive test data during digital cognitive assessments to conduct a multidimensional analysis.

In these studies, the performance of unimodal and multimodal combinations of biomarkers was thoroughly compared, with results consistently demonstrating that multimodal models outperformed unimodal models overall. For example, Se Young Kim's multimodal model<sup>109</sup>, which combined eye-tracking and manual data, improved accuracy by 13.3% compared to the unimodal model using only manual data. Similarly, Yasunori Yamada's study<sup>111</sup> showed that a model integrating manual movement, gait, and speech biomarkers achieved an 11.1% increase in accuracy in a ternary classification task compared to a model using only speech data. These findings indicate that integrating information from multiple data sources enhances diagnostic accuracy. For a detailed comparison of the best-performing unimodal and multimodal models, can be found in Supplementary Table 18.

### Validation of AI models

An effective predictive model is characterized by its ability to accurately estimate individual risk—meaning that the predicted outcomes align closely with actual outcomes, reliably distinguishing between high-risk and low-risk individuals (discrimination), and performing well across different populations<sup>124</sup>. Calibration and discrimination can be evaluated through internal validation (using the same dataset as the one used to develop the model) or external validation (using a different dataset). External validation is generally preferred as it more comprehensively assesses a model's generalizability<sup>125</sup>. However, in AD digital biomarker studies, 7 studies did not explicitly state whether they performed internal or external validation, and only 2 studies conducted both internal and external validation<sup>52,57</sup>. One manual biomarker study used temporal validation, testing the model with follow-up data collected one year later<sup>57</sup>. Another study used the external dataset from the ADNI-3 cohort for validation<sup>52</sup>. In terms of internal validation, most studies employed validation methods appropriate for small sample sizes. Specifically, 38 studies used k-fold cross-validation, while

4 studies adopted the more rigorous nested cross-validation, and 2 studies utilized stratified cross-validation. Additionally, 24 studies opted for leave-one-out cross-validation, and 2 studies employed leave-two-out cross-validation. In contrast, other validation methods, such as the hold-out method, are used less frequently. Moreover, model calibration is crucial for evaluating predictive performance, but only 4 studies<sup>44,52,87,100</sup> used calibration plots or the Hosmer-Lemeshow test to assess model calibration. Detailed information on validation methods and calibration approaches used in each study can be found in Supplementary Table 17.

### Reproducibility and reporting standards

Transparency and reproducibility are fundamental pillars of robust scientific research. This necessitates adherence to the Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines in studies involving AI predictive models, as well as the publication of code to enhance reproducibility<sup>126</sup>. However, among the 86 studies examined, only one explicitly acknowledged adherence to the TRIPOD guidelines<sup>40</sup>, and two studies made their research code publicly available<sup>40,120</sup>. This omission undoubtedly undermines the credibility and applicability of AI predictive models in the medical field. Further details can be found in Supplementary Table 17.

### Description of data missing in device collection

Data quality is critical in the development and application of AI models, especially during the training and inference stages. Studies have shown that in 26 reports on data collection, all provided specific descriptions of issues related to data missing. These included 3 gait studies, 3 manual biomarker studies, 4 eye movement studies, 5 speech studies, 4 tests on ICT or mobile devices, 2 physiological signal studies, 2 multi-type biomarker studies, 2 home activity studies, and 1 other type of study. The number of excluded samples due to missing data ranged from 1 to 474 individuals. The excluded sample size accounted for 0.69% to 58.52% of the total sample in each study.

In terms of reasons for data exclusion, 8 studies reported issues with technical devices (e.g., equipment malfunction), 6 studies cited patient or participant-related reasons (e.g., inability to complete tasks, refusal to



**Fig. 19 | Sunburst Chart Showing The Distribution of Specific Features Used in Various Digital Biomarker Studies.** Pie segments of the same color represent the features used in research on the same type of digital biomarker. The size of the segments indicates the frequency of feature usage, with larger areas reflecting more frequent usage.

participate in measurements), 2 studies faced problems with patient inclusion or exclusion during the research design, 1 study had issues related to data management or uploading, and 6 studies had two or more factors contributing to data loss and sample exclusion. These challenges further reduced the available sample sizes, highlighting areas that require closer attention in future research. Further details can be found in Supplementary Table 19.

## **Gait biomarker collection devices, specific tasks, and feature descriptions**

Various technical methods, such as wearable devices, depth cameras, and pressure sensors, have been used to capture subtle gait variations, revealing differences in patients' movement patterns during static and dynamic tasks. These findings further support the idea that gait analysis is an effective tool for assessing cognitive function. Among the 13 studies on gait digital

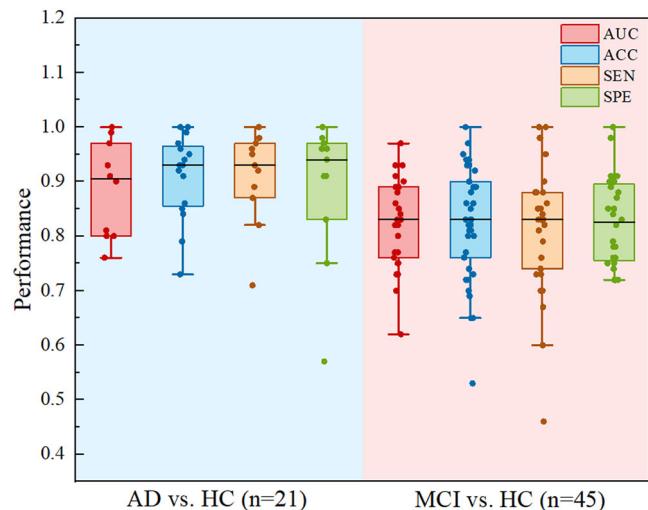
biomarkers, 4 used wearable devices for data collection<sup>47,49,66,67</sup>, 3 used electronic walkways<sup>26,48,70</sup>, 5 used the Kinect camera<sup>68,69,71-73</sup>, and 1 study employed a Doppler radar system<sup>46</sup>.

In terms of task selection and execution, 3 studies used simple single-task 10-meter walking tests<sup>46,71,72</sup>, while 7 studies employed both single-task and dual-task walking tests<sup>26,48,66–68,70,73</sup>. It's important to note that although the single-task and dual-task tests aim to analyze cognitive load and behavioral performance during walking, there are subtle differences in task design. For instance, in addition to the common task of walking while counting backward, there are more complex dual tasks, such as walking while answering questions and performing a chip exchange<sup>66</sup>, or simpler tasks like a 10-meter walk<sup>71</sup> versus designs with obstacles. Overall, AD and MCI patients with cognitive decline exhibit different performances in terms of walking spatial, temporal, and velocity characteristics, as well as variability in these features.

## Finger movement biomarker collection devices, specific tasks, and feature descriptions

In the 10 manual biomarker studies, various drawing and writing tasks were performed, including the Trail Making Test (TMT)<sup>27,57,74</sup>, Clock Drawing Test (CDT)<sup>27,41,74,75,78</sup>, writing a sequence of ‘l’ letters<sup>79</sup>, pentagon copying tasks, and sentence writing tests<sup>27,44</sup>. These tasks are primarily used to assess individuals’ executive function, visuospatial abilities, and writing skills, which are often impaired to varying degrees in AD and MCI patients. Additionally, devices such as digital tablets and pens, touchscreen systems, and electromagnetic tablets can capture not only static results of writing and drawing but also subtle changes during the dynamic process.

Overall, cognitively impaired patients exhibit significantly different performances in these tasks compared to healthy controls. For instance, differences have been observed in drawing speed<sup>17</sup>, pressure fluctuations<sup>74</sup>, speed variability<sup>74</sup>, pause time<sup>75</sup>, and smoothness<sup>76</sup>.



**Fig. 20 | The Box plot of Binary Classification Performance Based on AD and MCI AI Models.** Each point in the box plot represents an individual study. The horizontal line inside the box indicates the median of the data, while the whiskers represent the range of the maximum and minimum values. Data points outside the whiskers are considered outliers.

## Eye movement biomarker collection devices, specific tasks, and feature descriptions

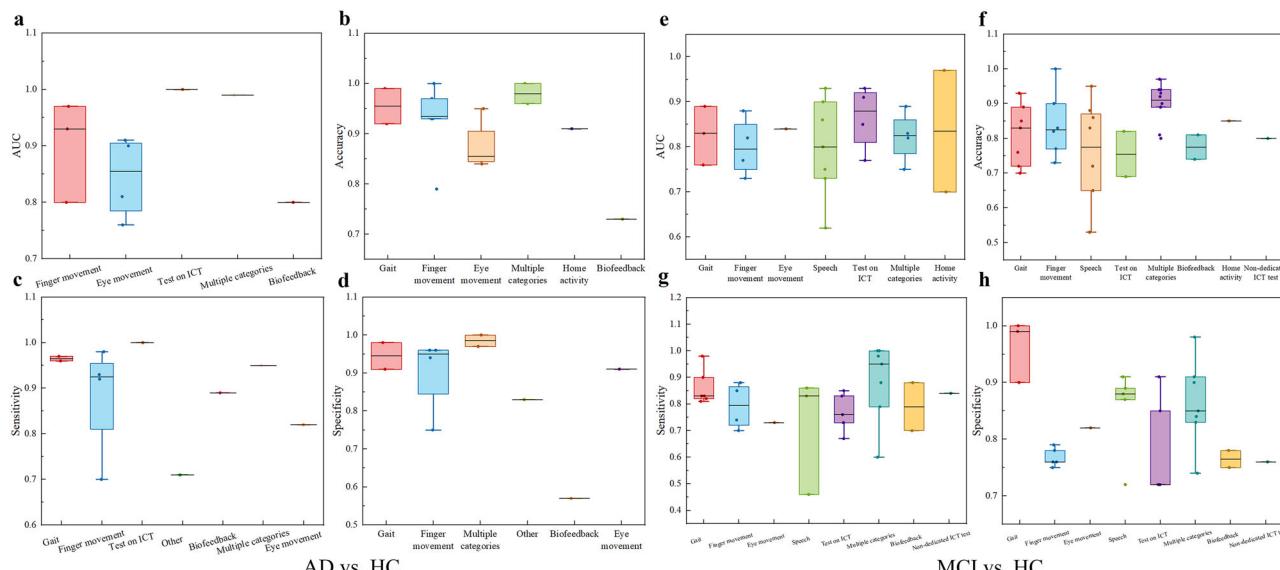
The 11 eye movement studies included various types of systems, ranging from custom-built setups<sup>80,81,83</sup> to commercial devices<sup>39,40,82,84–88</sup>, and examined different visual tasks, particularly those assessing patients’ visual attention and eye movement control. Saccade, fixation, and visual matching tasks were the primary paradigms. Overall, these studies indicated that AD and MCI patients exhibited significant cognitive deficits during visual tasks. For example, in visual matching and visual search tasks, patients’ fixation locations<sup>80</sup>, fixation durations<sup>80</sup>, and visual exploration levels differed markedly from those of healthy controls, especially in tasks requiring precise fixation and quick responses.

These differences were not only reflected in fixation duration and eye movement paths but also in more subtle features such as visual attention heatmaps<sup>83</sup>, eye movement speed<sup>81</sup>, and fixation stability<sup>85</sup>. Additionally, using data from various eye-tracking devices, such as custom non-invasive eye-tracking systems, the Tobii Pro Spectrum system, and the Eyelink II head-mounted eye tracker, researchers found that AD patients showed higher error rates and lower accuracy in error inhibition<sup>82</sup>, error correction<sup>82</sup>, and fixation precision<sup>81,85</sup>. Quantitative light reflex pupillometry has shown that AD patients have significantly lower average and maximal pupil constriction velocities than healthy controls<sup>40</sup>. Recent studies also indicate that eye-tracking data from tablet devices provides diagnostic performance comparable to traditional commercial systems. For example, Qinjie Li et al. used a Xiaomi Mi 5 Pro tablet to capture eye movement data (latency, accuracy, and duration) and developed a logistic regression model that differentiated AD patients from healthy individuals with 82% sensitivity and 91% specificity<sup>87</sup>.

## Speech biomarker collection devices, specific tasks, and feature descriptions

Twelve studies utilized speech recognition and recording devices to assess language and cognitive functions. These studies employed various technological platforms, including automatic speech recognition systems, digital tablets, and wearable devices, covering a range of task paradigms from picture description and verbal fluency tests to conversation tasks and memory tests. Overall, MCI patients exhibited significant impairments in language expression and cognitive function, particularly in tasks involving verbal fluency, complex language generation, and memory.

These differences were evident not only in the complexity of language content, such as the number of semantic clusters and word repetition



**Fig. 21 | Box plot of AI Model Performance for Binary Classification Based on AD or MCI versus HC.** **a** AUC for AD and HC Classification, **(b)** ACC, **(c)** SEN, **(d)** SPE. **e** AUC for MCI and HC classification, **(f)** ACC, **(g)** SEN, **(h)** SPE.

frequency<sup>50</sup>, but also in acoustic features like pitch<sup>51</sup>, jitter<sup>89</sup>, and spectral characteristics<sup>90,92</sup>. For instance, in picture description tasks conducted using the AcceXible platform, patients' vocal spectral features, such as F3 bandwidth and Hammarberg index, showed noticeable abnormalities<sup>90</sup>. These features reflect the stability and consistency of vocal production during language generation.

### **ICT-based testing biomarker collection devices, specific tasks, and feature descriptions**

In 13 studies using touchscreen devices, virtual reality (VR) equipment, and computer technology to assess cognitive function in AD and MCI patients, cognitive function was evaluated across multiple dimensions, including memory, attention, executive function, and spatial navigation ability. These digitized tests were able to distinguish different levels of cognitive impairment. For example, in the digital Chinese Neuropsychological Consensus Battery (CNCB) test conducted on touchscreen devices, patients exhibited significant cognitive deficits in word recall, task completion time, and accuracy in reverse sequence memory<sup>52</sup>. In the dual-task balance ball test, patients exhibit increased cognitive load when facing multi-tasking, as evidenced by higher dual-task costs and reduced time spent in the inner circle, reflecting difficulties in both balance control and attention allocation<sup>102</sup>. Similarly, in the VR virtual supermarket task, patients' navigation trajectories, task execution time, and behavioral data showed abnormalities<sup>96,97,103</sup>, reflecting difficulties in spatial memory and task planning.

### **Biofeedback and physiological signal biomarker collection devices, specific tasks, and feature descriptions**

Studies on biofeedback and physiological signals are relatively scarce, with only five studies focusing on this area. These studies provided profound insights into patients' neurophysiological characteristics by analyzing brain electrical activity and heart rate variability. For example, in a computer-based cognitive task using the Active Two BioSemi system, patients' EEG signals displayed distinct phase characteristics, frequency band features, and time-domain properties<sup>116</sup>, which may be associated with cognitive decline. Additionally, in resting-state EEG activity recorded using the Bitbrain multifunctional wireless wearable system, patients exhibited abnormalities in relative power characteristics, spectral entropy, and Hjorth complexity<sup>117</sup>, which may reflect brain function dysregulation.

### **Multi-type biomarker collection devices, specific tasks, and feature descriptions**

Thirteen studies on multi-type biomarkers (multimodal digital biomarkers) combined virtual reality (VR), EEG, eye-tracking, and other sensing devices to assess AD and MCI. These studies aimed to gain deeper insights into and detect changes in cognitive function by analyzing physiological signals, behavioral data, and language features. In studies combining EEG and eye-tracking, patients displayed abnormalities in event-related potentials (ERP), power spectral density, and eye movement characteristics during resting-state and delayed matching tasks<sup>105</sup>, further indicating deficits in visual processing and cognitive control. Similarly, in a language description task using wearable EEG devices (such as the MUSE 2) combined with VR equipment (such as the Oculus Quest 2), significant abnormalities were revealed in both speech generation and brain electrical activity<sup>107</sup>. In a CERAD cognitive task, individuals with MCI exhibited reduced brain electrical activity and heart rate variability (HRV), which may indicate dysregulation of autonomic nervous function<sup>113</sup>. The multimodal evaluation approach, which integrates multiple digital and sensor devices, offers a more comprehensive understanding of the cognitive changes in AD and MCI patients.

### **Home activity biomarker collection devices, specific tasks, and feature descriptions**

Six studies on home life monitoring used various sensors, such as door sensors<sup>55,63,115</sup>, infrared motion sensors<sup>54,56,62,63</sup>, location sensors<sup>55</sup>, and sleep

sensors<sup>55</sup>, to capture behavioral patterns and activity features, aiming to gain deeper insights into changes in cognitive function. Using wireless motion sensor nodes and passive infrared sensors, researchers analyzed spatial movement features, activity intensity, and activity pattern characteristics, further revealing the reduced complexity of movement and decreased activity intensity in patients performing daily tasks<sup>56</sup>.

Similar studies employed more complex sensor combinations, such as vibration sensors, temperature and humidity sensors, and Lidar sensors, to analyze behavioral patterns, frequency and duration of daily activities, and nighttime activity patterns<sup>115</sup>. These detailed monitoring data helped uncover the patterns and behavioral abnormalities in patients' activities throughout different times of the day.

Furthermore, studies utilized complexity analysis methods based on sensor data, such as loop complexity, entropy, room transition features, and fractal indices<sup>54</sup>. By calculating these features, researchers were able to quantify changes in behavioral patterns, particularly in terms of environmental transitions and the complexity of activities.

In addition, several studies analyzed GPS data to capture outdoor activity features, driving behaviors, and interactions with non-dedicated ICT devices, providing further insights into patients' digital interaction behaviors. For specific details regarding the data acquisition devices and task paradigms, please refer to Supplementary Tables 20 to 30.

## **Discussion**

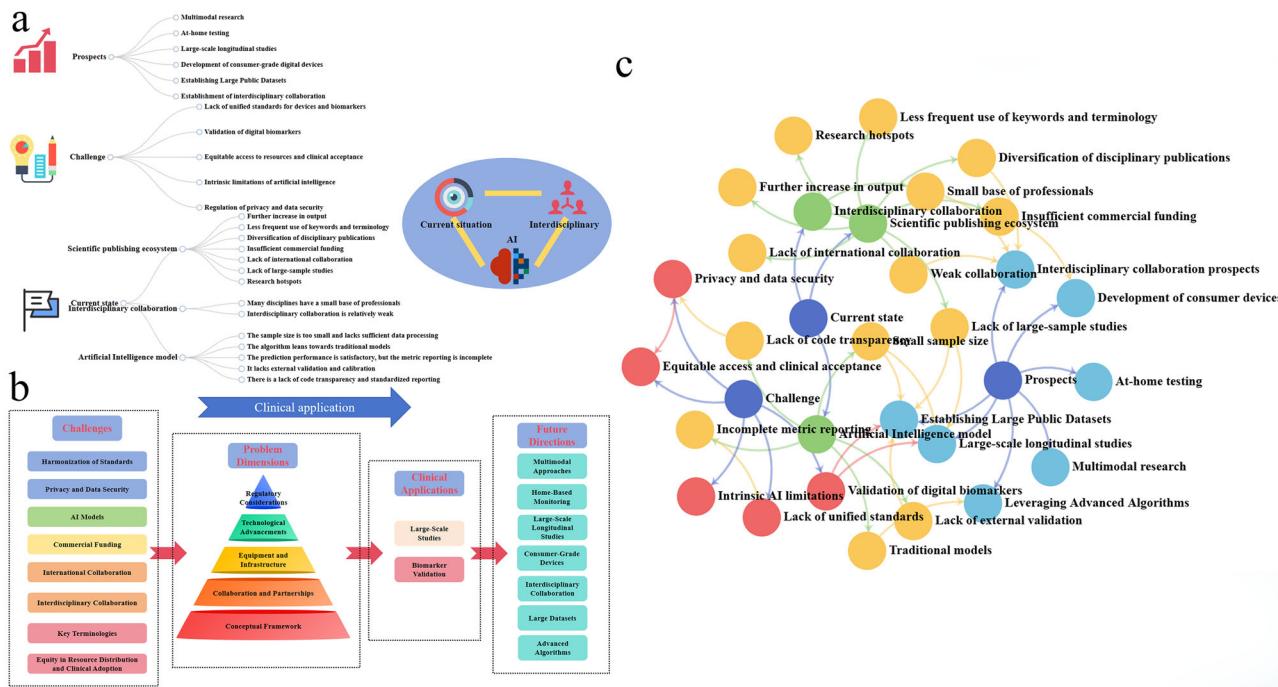
Through an in-depth analysis of the scientific output, interdisciplinary collaboration, and key advancements in AI model research within the field of AD digital biomarkers, we gain a comprehensive understanding of the current landscape, revealing several challenges and limitations. To structure the discussion more systematically, we categorize these issues into three major areas: the current state of the field, key challenges and obstacles, and emerging directions for future development. Detailed analysis is presented in Fig. 22a.

To effectively address these challenges, the key dimensions of clinical implementation can be conceptualized as a pyramid structure. At the base lies the conceptual framework, followed by partnerships, equipment and infrastructure, technological advancements, with regulatory requirements positioned at the apex. This structure illustrates that successful clinical application of digital biomarkers requires coordinated efforts across multiple levels. Standardization and privacy/data security present regulatory challenges, while AI models primarily face technical obstacles. The development of devices is influenced by commercial funding, and international and interdisciplinary collaborations drive scientific progress in this field. Terminology usage and clinical acceptance shape the conceptual framework of digital biomarkers. Additionally, large-scale studies and further validation are necessary for clinical adoption, both of which are significantly affected by the aforementioned five dimensions. The overall challenge framework at this stage is detailed in Fig. 22b.

Figure 22c illustrates the interrelationship between the key challenges, current status, and future outlook in the field of AD digital biomarkers, highlighting the complex, intertwined network among these elements. For instance, both AI model research and other types of studies face the challenge of limited sample sizes, which constrains progress. The training of AI models, in particular, demands larger datasets and more extensive data for robust training. Barriers within the scientific publishing ecosystem have resulted in fragmented research outputs, further hindering the overall advancement of the field. Nevertheless, future research directions are becoming clearer, encompassing home-based testing, the development of consumer-grade devices, the implementation of large-scale longitudinal studies, and the application of advanced algorithms. In the following sections, we will delve deeper into our key findings.

Currently, the body of research on digital biomarkers in the diagnosis and assessment of AD is steadily growing, indicating a rapidly evolving field. Specifically, a notable inflection point occurred in 2019, likely driven by technological advancements and shifts in conceptual frameworks.

- Technological Advancements:



**Fig. 22 | Discussion on Digital Biomarker Research for Alzheimer’s Disease.** **a** The current multidimensional landscape of digital biomarkers in Alzheimer’s disease. **b** The five key dimensions of the challenges in clinical implementation of digital biomarkers. **c** The relationship between the main challenges, current status,

and future prospects in the field of digital biomarkers for AD. The color of the circles represents the hierarchical relationships of the topics, while the arrows indicate the connections and directional relationships between all topics.

1. The widespread adoption of big data and cloud computing technologies has enabled researchers to process and store large-scale datasets, significantly expanding the application and in-depth study of digital biomarkers. 2. Progress in machine learning and AI has made handling high-dimensional data more efficient and feasible<sup>27</sup>.
  3. Continuous advancements in digital devices, such as wearable devices and smartphones, have provided higher-quality and more granular data sources<sup>128</sup>.
- **Conceptual Shifts:**  
The COVID-19 pandemic has accelerated the adoption of digital diagnostics for neurodegenerative diseases, greatly increasing the acceptance of digital technologies and devices<sup>129</sup>. Compared to traditional diagnostic approaches, cost-effective and non-invasive methods are gaining widespread attention<sup>130</sup>, further fueling the surge in research on digital biomarkers. More research outcomes in this area are anticipated in the near future.

The use of keywords reflects researchers’ understanding of the field. Our research indicates that only 34 studies, accounting for 7.89% of the total output, used “digital biomarkers” as a keyword. This term only began gaining broader recognition among researchers after 2020, suggesting that the awareness of this terminology in the field remains limited. Christian et al.<sup>131</sup> also pointed out that many related studies did not explicitly use this term, possibly due to inadequate dissemination or confusion with other concepts, such as “clinical outcome assessments”<sup>12</sup>. To enhance the reproducibility and impact of research, it is essential to clarify and promote standardized terminology<sup>132</sup>. The DACIA framework (Data, Aggregation, Contextualization, Interpretation, and Action)<sup>133</sup> offers a structured approach, facilitating the systematic collection and analysis of data by researchers. Additionally, categorizing digital biomarker studies into four dimensions—population, devices, tasks, and data—helps to clarify the research process. Through interdisciplinary training and educational resources, such as online courses, workshops, and seminars, researchers can further improve their understanding and use of these terms. In summary,

while the application of digital biomarkers in AD research is steadily increasing, the standardization and dissemination of terminology still require further improvement. By establishing a structured framework and promoting widespread use of these terms, future research will achieve greater reproducibility and international impact.

Overall, funding for digital biomarker research in AD has shown a fluctuating upward trend, with each study receiving an average of about 3.5 grants. However, the majority of this funding comes from government agencies, while support from private companies or consortia remains relatively limited. In traditional biomarker research and drug development, corporate funding has played a pivotal role in driving both research and commercialization<sup>134</sup>. Digital biomarker research heavily relies on various digital tools, such as the Eyelink eye-tracking device<sup>135</sup> and the GaitRite electronic walkway system<sup>136</sup>, which are typically developed by companies. Currently, the FDA is promoting the SaMD (Software as a Medical Device) initiative to streamline and regulate the approval processes for medical software, while encouraging companies to develop more advanced and reliable digital biomarker devices<sup>137</sup>. A successful example is Altoida, a U.S.-based company whose predictive machine learning algorithm was granted breakthrough device designation by the FDA in 2021. The company received significant corporate funding to further its research on predicting the progression from mild cognitive impairment to AD<sup>135</sup>. Therefore, increasing financial support from corporations and consortia, whether through direct investment or collaborations with research institutions, will provide more opportunities for the development and adoption of digital biomarkers, ultimately advancing innovation and progress in the field.

The United States is undoubtedly a core contributor to the field of digital biomarkers in AD. However, international collaboration remains relatively limited in other high-output countries, particularly China, South Korea, and Japan, where research tends to focus on internal collaborations. In contrast, European countries demonstrate a stronger willingness for cross-border cooperation, although this is often confined to neighboring nations. This trend benefits from the knowledge spillover effect facilitated by geographic proximity, which promotes innovation and the flow of tacit

knowledge<sup>138</sup>. Given the heterogeneity of dementia, there is a pressing need for multicenter collaborative studies to address its complexity. Establishing larger international collaboration platforms, providing funding and policy support, and sharing data and resources would further promote global cross-border cooperation. This would not only improve research quality and efficiency but also offer more treatment hope for dementia patients. For instance, the U.S.-led Framingham Heart Study, which collects data from diverse racial groups<sup>139</sup>, and the ADNI project, centered around traditional biomarkers<sup>140</sup>, have already fostered international cooperation on a global scale. Future efforts could look to replicate these models to establish more international collaborative projects.

Digital biomarkers have the advantages of being cost-effective, non-invasive, and easily repeatable, offering great potential for large-scale studies. However, to date, only a limited number of large-sample studies have been conducted, primarily in gait analysis and digital testing, falling short of expectations. The primary limitations may not stem from technical issues, as the large-scale analytical capabilities of machine learning and deep learning have been well demonstrated in traditional biomarker research. The limiting factors may include insufficient funding, limited access to equipment, patient acceptance, and the complexity of testing protocols. For example, home activity monitoring requires the installation of embedded sensors and cameras<sup>141</sup>, while natural driving behavior studies necessitate onboard devices<sup>142</sup>, both of which could raise concerns about intrusiveness. Additionally, long-term longitudinal studies are prone to external disruptions<sup>143</sup>. Despite the increasing prevalence of digital devices, activity monitoring, which closely reflects daily life, still faces challenges such as computer literacy and device maintenance (e.g., recharging), which impact the feasibility of large-scale studies. Regardless of the type of digital biomarker, large-sample validation is essential. Only through comparative studies with traditional biomarkers can digital biomarkers be widely applied in clinical screening and diagnosis.

Research on digital biomarkers for AD primarily focuses on motor activity, speech, eye movement, and digital assessments. These biomarkers, which are closely tied to daily life functions, utilize established technologies and algorithms such as gait analysis, speech recognition, and eye-tracking to capture early AD-related behavioral changes. Notable features, including gait abnormalities<sup>144,145</sup>, speech fluency changes<sup>146</sup>, and slowed eye movements<sup>147</sup>, have shown high correlation with early cognitive decline associated with AD. However, their specificity is limited, as similar abnormalities are seen in conditions like Parkinson's disease, amyotrophic lateral sclerosis, and general frailty in the elderly, and other diseases<sup>148–150</sup>, which reduces their unique diagnostic value for AD. Additionally, environmental factors (e.g., lighting, posture) and cultural differences (e.g., language) can influence performance, impacting both data accuracy and model generalizability. Speech biomarkers face significant challenges in cross-cultural and linguistic adaptation, and the dynamic nature of individual speech further complicates data analysis<sup>151,152</sup>. Future research should focus on enhancing the specificity of these biomarkers, developing adaptive models that are robust across cultural and environmental contexts, and designing personalized biomarkers to increase their clinical applicability. Other digital biomarkers also hold valuable potential. Studies on driving behavior have shown that abnormal cerebrospinal fluid Aβ42/Aβ40 ratios are associated with poor driving performance<sup>142</sup>, supporting the potential of driving behavior as a biomarker for Alzheimer's disease. Simple, everyday tasks such as keyboard typing and computer access logs may also reflect cognitive decline without the need for specialized equipment<sup>153,154</sup>. Monitoring of physiological signals and sleep patterns currently relies heavily on laboratory-based polysomnography (PSG) and clinical EEG<sup>118</sup>. However, emerging technologies such as wearable EEG devices and activity trackers offer new possibilities for monitoring in everyday settings<sup>155</sup>. These diverse digital biomarkers expand the toolkit for early detection of AD.

Research on digital biomarkers for AD has gained significant attention from influential scholars, such as Morris, John C., and Kaye, Jeffrey A., but interdisciplinary collaboration remains insufficient. The field's highly interdisciplinary nature, spanning sensor technology, computational

science, and neuroscience, requires researchers from diverse backgrounds to advance. Despite recent efforts, participation from related disciplines like psychiatry, psychology, and gerontology remains low. Attracting more interdisciplinary researchers is crucial, as demonstrated by initiatives like the U.S. FDA's digital health expert network<sup>156</sup> and the Global Ataxia Initiative's consensus on smartphone sensor evaluation standards<sup>157</sup>, both of which have successfully fostered collaboration. Educational programs, like those at Stanford University, emphasize the importance of interdisciplinary teams<sup>158</sup>, which could be replicated to further engage researchers from diverse fields, such as neuroscience and engineering. Clear role definitions across disciplines—where engineering drives technological innovation and neurology addresses diagnostic challenges—are essential for effective collaboration<sup>159,160</sup>. A shift to open research models and cross-disciplinary partnerships, supported by both government and private sectors, will foster the integration of science and technology, enhancing research quality and advancing AD biomarker development.

Many studies suffer from small sample sizes, particularly in AD research, where some sample sizes were fewer than 20 participants<sup>56,96,99,115</sup>. Small sample sizes limit the generalizability and fit of AI models, often leading to imbalanced performance across different datasets. Furthermore, the high heterogeneity of the data exacerbate these challenges, especially when AD and MCI share overlapping behavioral features, making the issue of noise particularly pronounced. Improper data handling, such as neglecting sample imbalances or ignoring methods for missing data imputation, can result in biased or misleading performance metrics<sup>161</sup>. However, only a few studies have provided detailed reports on how missing values and sample imbalances were handled. This lack of transparency affects the reliability and reproducibility of the results. To address these issues, future research should focus on comparing different imputation methods and providing detailed reports on how data imbalances and missing values are handled. This will not only improve model accuracy and stability but also promote the broader application of AI models in AD research.

Although machine learning is inherently an iterative process, and comparing multiple algorithms can help generate more optimized predictive models<sup>162</sup>, this approach is still underutilized in existing research. Only half of the studies compared multiple algorithms and selected the optimal model, while the rest tested only one algorithm. Classical machine learning models are widely used. However, in traditional biomarker research, deep learning algorithms, such as neural networks, have gradually become mainstream<sup>163</sup>. Deep learning models can automatically extract features from raw data without the need for manual selection. This capability enables the models to capture more complex patterns, particularly when dealing with high-dimensional data. Additionally, the application of transfer learning can further enhance training efficiency. Future research should more extensively test a variety of algorithms, especially deep learning models, to improve prediction accuracy and advance the field of digital biomarker research.

Many studies fail to sufficiently report commonly used evaluation metrics, making it more difficult to assess the generalizability and practical applicability of AI models. Our analysis reveals that the average AUC, accuracy, sensitivity, and specificity for AD vs. HC classification are generally higher than those for MCI vs. HC classification. This is likely due to the more pronounced differences between AD and HC, whereas distinguishing between MCI and HC is more challenging<sup>164</sup>. However, identifying the optimal model remains challenging, as the type of features and their combinations significantly impact model performance. Additionally, variations in data collection devices may lead to differing results. Although most models report an AUC greater than 0.75, indicating good predictive performance<sup>165</sup>, future research should systematically report these evaluation metrics, with at least specific values for AUC, ACC, SEN, and SPE, to provide a comprehensive perspective for comparing and improving model performance.

The lack of external validation is a major barrier to the application of artificial intelligence models. Only two studies have conducted external

validation, raising concerns about the generalizability of these models and limiting their potential for clinical use. The reproducibility crisis in precision psychiatry further underscores the importance of external validation. Therefore, validating AI models across different settings and populations is critical<sup>166</sup>. In addition, insufficient model calibration is another prominent issue. Calibration is a key step in ensuring that predicted probabilities align with actual outcomes. This is particularly important in clinical applications, where uncalibrated models may increase patient risk. Thus, future research should focus on the development and reporting of model calibration methods to ensure accuracy and reliability in clinical settings.

The lack of code transparency and adherence to reporting standards is a significant issue in current artificial intelligence model research. Many studies do not follow established reporting guidelines, such as the TRIPOD statement, compromising transparency and comprehensibility. Future studies should prioritize compliance with these standards to improve research quality. Additionally, the open availability of code and data is essential for ensuring research reproducibility. However, most studies have not made their model code and data publicly accessible, limiting the ability of other researchers to independently validate and replicate the models. To promote reproducibility, future work should place greater emphasis on making model code and data publicly available, facilitating validation and application in real-world settings. For example, Lukas et al.<sup>167</sup> have developed and released an open-source Python package called SciKit Digital Health, which provides a range of algorithms for digital biomarker feature extraction. The open access to this code is expected to facilitate advancements in early diagnosis, personalized treatment, and continuous monitoring of AD.

In fact, the clinical application of digital biomarkers in AD still faces several challenges that are unlikely to be resolved in the short term. These include regulatory concerns around privacy and data security, the establishment of unified standards, validation of digital biomarkers, issues of resource equity and clinical acceptance, and the inherent limitations of artificial intelligence.

With the widespread use of digital devices and software, data privacy and security concerns have grown increasingly important. Both devices and software must implement stringent privacy protection measures, and researchers and developers must comply with relevant privacy regulations. Early diagnosis of AD involves personal health, behavioral, and biometric data, and any data breaches could severely compromise privacy<sup>168</sup>. Therefore, technologies such as data anonymization and encryption, including federated learning, are crucial<sup>169</sup>. During the data collection process, it is essential to ensure informed consent from participants, particularly early-stage AD patients who may not fully understand the novel data collection methods used in digital biomarkers<sup>170</sup>. Additionally, traditional privacy regulations, such as HIPAA, face challenges in non-traditional clinical environments<sup>171</sup>. Technology companies often manage data sharing through end-user license agreements, which may no longer be applicable, raising ethical concerns<sup>172</sup>. This involves different considerations from various stakeholders, including patients, physicians, governments, and corporations. As technology evolves, relevant laws and regulations must be continuously updated to address emerging privacy protection challenges.

While various digital biomarkers have demonstrated good sensitivity and accuracy in research, the field lacks clear feature definitions and unified diagnostic standards. For instance, it remains challenging to determine the extent of behavioral differences and ranges that constitute reliable quantitative results due to the influence of multiple factors. With the ongoing emergence of new devices, the absence of international guidelines for device selection and usage has impeded the further development of AD monitoring. Additionally, discrepancies in data measurement and non-uniform formats across devices have created obstacles for data transmission and sharing. Establishing standardized data transmission protocols among similar biomarkers is therefore crucial. In terms of data processing, the lack of standardized methods for data collection, processing, and evaluation has increased the difficulty of validating biomarker effectiveness<sup>173</sup>. For example, the standardization of measurement methods and the precise placement of

wearable devices are expected to significantly improve the accuracy of gait analysis<sup>174</sup>. Hence, developing a comprehensive set of standards that covers AD digital biomarker selection, data processing workflows, and effectiveness evaluation is urgently needed to advance research in this field.

The validation of digital biomarkers is a critical prerequisite for their integration into clinical practice. To confirm their clinical effectiveness, extensive experimentation is required, along with correlation analysis with traditional biomarkers to ensure consistency with clinical indicators. Digital biomarkers should not only serve as a supplement or replacement for existing biomarkers but, in some cases, may become new standards for disease monitoring, prediction, and therapeutic evaluation. A major challenge in the validation process stems from the increasing granularity of features. The complexity and multidimensionality of the data demand additional efforts to establish links between digital biomarkers and traditional ones, often requiring multiple experimental validations. This validation process is akin to the standard reference validation of in vitro diagnostic devices and must be given significant attention to ensure the reliability and feasibility of digital biomarkers in clinical practice.

The diversity of devices introduces challenges related to resource equity. Device accessibility is notably unequal, particularly among AD patients and the elderly, where differences in digital literacy and technological acceptance exacerbate this inequality<sup>170</sup>. Additionally, clinical acceptance of digital biomarkers also presents a key challenge to their widespread adoption. Many healthcare professionals are not yet familiar with these new technologies, and using digital biomarkers may increase their workload<sup>175,176</sup>. Systematic training and integration initiatives will be critical in improving clinical acceptance. Moreover, data monopolies within the digital health sector contribute to inequality<sup>177</sup>. Some companies enforce closed data management for wearable devices, restricting external researchers' access and hindering academic research and medical innovation<sup>178</sup>. Moving forward, it will be essential to address resource equity in these three areas.

AI has brought opportunities but still faces several limitations in digital biomarkers. First, inconsistencies in data quality and quantity, especially variations in data from different devices, limit the generalizability of AI models. Additionally, the "black box" nature of AI remains a significant issue. While models can provide accurate predictions, the lack of transparency in the decision-making process hinders the establishment of trust in clinical applications<sup>163</sup>. Due to limited sample sizes and the need for high-quality data, AI models may underperform in certain groups if the training data for those populations is insufficient or not representative—an issue currently present in this field. Of course, this issue is not limited to AD; similar challenges are also observed in Parkinson's disease<sup>179</sup>. It is anticipated that with the establishment of high-quality datasets, this issue is expected to be mitigated. Moreover, accountability in the use of AI in medicine poses another major challenge. As AI ethics evolve, accountability issues are expected to be gradually addressed, although it will still take some time<sup>170</sup>.

Additionally, we believe that future research in this field can largely advance in seven specific directions: multimodal studies, home-based testing, large-sample longitudinal research, development of consumer-grade digital devices, construction of interdisciplinary collaboration frameworks, establishment of large open datasets, and development of advanced algorithms and systems. The integration of multimodal data is emerging as a key trend in the early diagnosis and assessment of AD and related dementias (ADRD). International initiatives such as the Early Detection of Neurodegenerative Diseases (EDoN) are developing digital toolkits based on multimodal data to identify early biomarkers of AD and ADRD<sup>180</sup>. Data fusion technologies will enable the combination of diverse modalities, building more precise predictive models, thus facilitating personalized treatment and early intervention<sup>15</sup>. Future research must focus on validating the clinical applicability of these methods and optimizing their use across different stages of disease progression.

At present, most assessments are still conducted in clinical settings, making them susceptible to the laboratory effect and training effect, which

diminish the objectivity of the results. This may be influenced by factors such as data collection methods and privacy concerns. For example, the “first-night effect” is commonly observed in sleep studies<sup>181</sup>, while the laboratory effect frequently affects gait research<sup>182</sup>. Home-based assessments can mitigate these biases by increasing the frequency of evaluations through ecological momentary assessments conducted via smartphones. Additionally, embedded passive sensors enable data collection from populations that are more difficult to access<sup>183</sup>. In fact, this approach promotes a shift from active testing to passive monitoring, potentially fostering multimodal research, such as the integration of sleep, gait, and other metrics into comprehensive monitoring.

Large-scale longitudinal studies face numerous challenges in utilizing digital biomarkers for AD detection. These challenges arise not only from the limitations of devices and technologies but also from issues related to data integration and privacy protection. However, some mobile device-based digital assessments have shown significant potential<sup>184</sup>, particularly in terms of assessment frequency and data diversity. Future research should prioritize the exploration of such tools that are scalable and feasible for widespread use, as they are likely to offer the first viable solutions for clinical evaluation and pave the way for further biomarker research<sup>185</sup>. Nevertheless, this does not suggest neglecting other forms of digital biomarker studies, as the ultimate goal for all biomarkers is to transcend the limitations of traditional approaches and advance towards large-scale, longitudinal assessments.

According to a market report by BIS Research, the global digital biomarkers market generated \$524.6 million in revenue in 2018, and is projected to exceed \$5.64 billion by 2025<sup>186</sup>, demonstrating immense market potential. In Alzheimer’s research, commonly used devices such as electronic gait mats and eye trackers are expensive and primarily confined to laboratory settings. Moving forward, efforts should focus on developing affordable, unobtrusive devices that can be integrated into clinical practice<sup>15</sup>. However, using commercial devices presents challenges related to data sharing and privacy<sup>187</sup>, and may lead to region-specific ethical issues, such as the commercial use of health data by insurance companies. Additionally, older adults’ acceptance of new devices is a critical factor<sup>188</sup>. Involving multidisciplinary teams and patient participation will be key to advancing this field.

The interdisciplinary nature of digital biomarker research underscores the importance of establishing stable and active collaboration networks. In the future, exploring how to build robust cross-disciplinary collaboration models in various types of digital biomarker research will become a key issue.

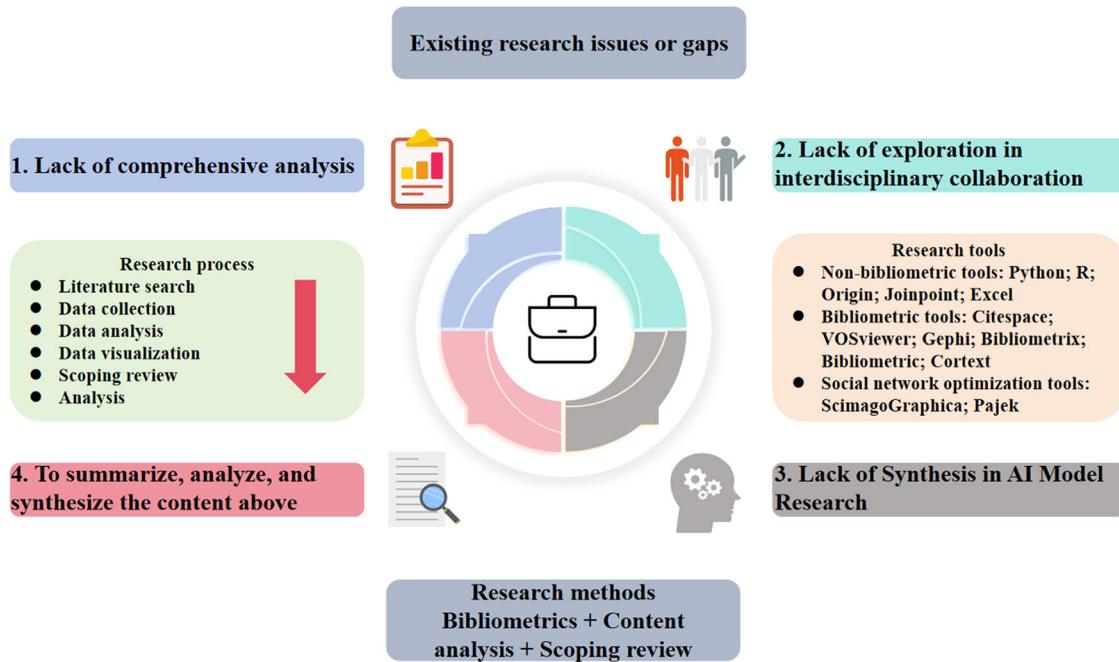
The establishment of large-scale public datasets facilitates resource sharing and provides research opportunities, particularly in resource-limited regions. For example, the RADAR-AD sub-study uses smart home sensor data to monitor activities of daily living, generating high ecological validity datasets that offer new insights into functional, behavioral, and perceptual decline in Alzheimer’s disease, supporting risk stratification analysis<sup>189</sup>. The latest ADNI initiative, ADNI4, one of the largest multi-center datasets in AD research, aims to recruit at least 20,000 participants through an online portal for long-term assessments, including the Novoic Storyteller test<sup>190</sup>. Similarly, the DPUK clinical studies and Great Minds registry plan to enroll up to 3 million participants for smartphone and web-based cognitive assessments, with data feeding back into the DPUK data sharing platform<sup>191,192</sup>. It is also worth noting that the AD-CLIP dataset, which focuses on behavioral data, not only provides valuable data but also employs depth camera technology to ensure privacy protection<sup>193</sup>. These initiatives not only bridge resource gaps but also support large-scale research on digital biomarkers. Moreover, the open-access nature of such efforts may provide greater opportunities for external validation of AI models.

In AI models for digital biomarkers, the development of novel algorithmic architectures and systems not only enhances personalized predictions but also facilitates the discovery of new digital biomarkers. For instance, the iSleep system, which monitors sleep via smartphones<sup>194</sup>, the EmoMarker system, designed to capture emotional binary digital

biomarkers in dementia patients<sup>195</sup>, and DeepHeart, a deep learning method for accurate heart rate estimation from photoplethysmography signals<sup>196</sup>, are key innovations. These advancements enable the quantification and analysis of physiological and behavioral features that were traditionally difficult to capture, significantly expanding the scope of digital biomarkers. This is precisely what current AD biomarker research lacks. In the future, the development of new algorithmic frameworks and systems for various types of biomarkers will provide forward-looking insights for personalized diagnosis and intervention in AD patients.

Moreover, research on data augmentation and synthesis using Large Language Models (LLMs) has emerged as a promising approach for generating large open datasets for AD patients. For example, SHADE-AD<sup>197</sup> leverages LLMs to learn activity features of AD patients from real-world data, enabling the creation of synthesized datasets to address challenges such as data imbalance. Additionally, researchers are increasingly focusing on multi-modal sensor systems to detect dyadic digital biomarkers related to the living conditions of AD patients, such as family expressed emotion—a quantifiable measure of the family environment in terms of hostility, criticism, and distancing<sup>195</sup>. Furthermore, LLMs’ advanced understanding and reasoning capabilities are being utilized to develop interactive in-home healthcare systems. CaringFM utilizes privacy-protecting sensors to deploy LLM at home provide general health suggestions and personalized medical information to elderly with AD and other chronic diseases<sup>198</sup>. Advanced methods from other diseases are also worth considering, such as the integration of radiofrequency (RF) technology and AI for new monitoring approaches in Parkinson’s disease (PD). This method has proven effective in gait analysis, enabling the extraction and analysis of gait velocity in PD patients. This helps assess the severity of the disease, its progression, and the patient’s response to medication<sup>199</sup>. In the field of sleep monitoring, this approach allows for non-contact extraction of the patient’s respiratory signals and accurately evaluates sleep stages and respiratory events<sup>200</sup>. Furthermore, By analyzing nocturnal breathing patterns, it has successfully and preliminarily differentiated AD from PD and effectively assessed disease progression<sup>201</sup>. Despite these promising advancements, future research must focus on validating the clinical applicability of these methods and optimizing their use across different stages of disease progression.

This study represents the first comprehensive analysis of research on digital biomarkers in AD diagnosis using biometric methods, with a focus on research patterns and the development of interdisciplinary collaboration. Although significant progress has been made in this field, many challenges remain. Based on these findings, we provide practical recommendations for future research. Additionally, with the advancements in AI algorithms, our understanding of AD diagnosis and monitoring is being redefined. This review aims to fill existing gaps in the literature by systematically summarizing and analyzing the latest developments in Alzheimer’s digital biomarkers, particularly the application of AI models. However, while we discuss the potential of these technologies, our study intentionally focuses on digital biomarker technologies and methodologies, which may not fully encompass all emerging technologies. For example, while technologies such as radio frequency and large language models show promise, they are still in the early stages of application and require further empirical studies to validate their efficacy and reliability. In fact, our study also has limitations. First, we analyzed only English-language publications, potentially overlooking high-quality research in other languages. Second, the lack of standardized classification criteria for digital biomarkers introduces a degree of subjectivity in the categorization process, despite consultation with interdisciplinary experts. Third, this study primarily focuses on the application of “digital biomarkers” in AD. However, this more specific definition may limit the discussion of broader, potentially relevant “measures.” Lastly, while we focused on AI model characteristics, we did not conduct a deep evaluation of methodological quality. Publication bias may have resulted in an overestimation of the benefits of AI models in risk prediction. Additionally, the heterogeneity of the included studies complicates direct comparison of results.



**Fig. 23 | Flowchart and analytical content of the bibliometrics and content analysis method used in this study.** The image illustrates the three main issues addressed by the study, the analytical workflow, and the tools and methods employed.

## Methods

This study is divided into three main parts: The first part uses bibliometrics and content analysis to examine the current state of the field from various dimensions, including research output, countries, and institutions. The second part employs bibliometric methods to explore the characteristics of researchers in the field, thereby identifying the status and trends of interdisciplinary collaboration. These analyses follow the bibliometric framework proposed by Cobo et al.<sup>202</sup>. The third part builds upon the hotspot topics identified in the first section, conducting an in-depth scoping review with a focus on paradigms, tasks, features, algorithms, and performance of machine learning models for different digital biomarkers. The overall study workflow is illustrated in Fig. 23.

### Analysis tool

Integrating information from various databases, such as data from countries and institutions, and relying on manual analysis often faces challenges due to the large volume of data and the potential for human error in statistical calculations. Moreover, analyses using a single tool often struggle with the limitation of lacking high-resolution data analysis<sup>203</sup>. To address these challenges, we adopted a multi-tool bibliometric analysis strategy. Specifically, the tools used in our analysis include Citespace (Version 6.2.R6 Advance; Drexel University)<sup>204</sup>, VOSviewer<sup>205</sup> (Version 1.6.19; Leiden University), Bibliometrix<sup>206</sup>, bibliometric<sup>207</sup>, Gephi<sup>208</sup> (Version 0.10.1; Gephi.org), Joinpoint (Version 5.0.2; National Cancer Institute of the United States)<sup>209</sup>, and Cortex (Gustave Eiffel University)<sup>210</sup>. For layout and enhanced visualization, we utilized ScimagoGraphica (Version 1.0.16; Scimago lab)<sup>211</sup> and Pajek64 Portable (Version 5.18; University of Ljubljana)<sup>212</sup>. Data processing, analysis, and visualization were carried out using Origin2021 Pro (Origin Lab) alongside R packages, including ggplot2, reshape2, tidyverse, plyr, scales, and viridis. Detailed analysis strategies for each section are provided in Table 5.

### Data sources and search strategy

Considering the comprehensiveness of the search and the interdisciplinary nature of digital biomarkers, we searched five major databases: Web of Science Core Collection, PubMed, IEEE Xplore Digital Library, Embase, and CINAHL. Before conducting the formal search, all research team members

underwent professional training based on the textbook *Medical Literature Information Retrieval*<sup>213</sup>. With the assistance of librarians, neurologists, and medical informaticians, we developed a search strategy derived from the definition in the BEST glossary<sup>214</sup>. Core keywords included “Alzheimer’s disease,” “digital biomarkers,” and “diagnosis,” combined with terms related to disease behavior or physiological characteristics and associated measurement devices. Boolean operators were used to combine these terms. To minimize inclusion bias due to daily updates of resources in various databases, we conducted a unified search across all platforms on May 1, 2024, and completed the data export process. The scoping review of AI models involved a secondary round of retrieval and selection of all relevant studies published before December 31, 2024. Detailed search strategies and retrieval counts are provided in Supplementary Table 31.

### Inclusion and exclusion criteria

#### Inclusion Criteria:

- Studies must involve digital biomarkers obtained using digital devices or technologies; behavioral or physiological data obtained through other means are excluded as digital biomarkers.
- The literature must be an article published in a peer-reviewed journal.
- Studies must be written in English.
- The research objective must involve the application of digital biomarkers in the screening, diagnosis, or other relevant aspects of AD.

#### Exclusion Criteria:

- Studies where the full text is unavailable or where the content is incomplete.
- Duplicate publications.
- Non-journal literature (e.g., conference papers, books, abstracts).
- Non-research literature (e.g., systematic reviews, scoping reviews, meta-analyses, research protocols).
- Studies that use digital devices or digital biomarkers to assess treatment effects, care, or rehabilitation purposes.

### Screening strategy

Before the formal inclusion and exclusion of literature, two evaluators (WQ and YS) were assigned a randomly selected sample of 50 studies to conduct a

**Table 5 | Combined Analysis Strategy of Various Tools**

Content	Citespace	VOSviewer	Bibliometrix	Gephi	Bibliometric	Context
Publications Number	√					
Author		√	√			
Organization		√				
Country		√			√	
Keyword	√	√	√			
Keyword Cluster	√					
Grant	√					
Interdisciplinary collaboration					√	
Field-specific publication						√
Content	ScimagoGraphica	Pajek	Python	R	Joinpoint	Origin
Map visualization	√					
Social network optimization		√				
Variable coding				√		
Keyword data cleaning					√	
Identification of potential change points						√
Polynomial trendline fitting						√
Identify potential similar fields			√			

preliminary screening test to ensure the reliability of the screening process. The final Cohen's kappa value was 0.88, indicating high consistency, and thus no modifications were made to the inclusion criteria or the evaluators. During the formal independent screening process, any disagreements were resolved by SC, who intervened and participated in the decision-making. The final screening and verification process was completed on June 15, 2024. After identifying specific research hotspots, a further review was conducted to select studies related to AI models that met the required criteria.

### Extraction and classification of digital biomarkers

Based on the FDA's definition of digital biomarkers, and to ensure comprehensive classification, we referenced the digital biomarker research by Lampros C<sup>20</sup> and Antoine<sup>18</sup>, as well as the digital biomarker classification scheme for Parkinson's disease, a similarly neurodegenerative disease<sup>215</sup>. Additionally, we consulted with experts from the Digital Medicine Subcommittee of the Chinese Medical Association. To highlight the unique characteristics of different biomarkers and minimize redundancy in classification, we classified digital biomarkers into 11 categories: limb movement biomarkers, eye movement biomarkers, speech biomarkers, natural driving biomarkers, home activity biomarkers, digital measurement biomarkers based on mobile or dedicated ICT devices, non-dedicated ICT biomarkers (i.e., biomarkers derived from non-specialized information and communication technology devices), physiological signal biomarkers, sleep pattern biomarkers, other biomarkers, and multi-type biomarkers (i.e., multimodal biomarkers). The specific classification scheme is detailed in Supplementary Table 32.

### Data extraction

We designed two comprehensive data extraction tables. One table was used to extract extensive information from 431 studies to provide an overview of the field. This table includes details such as publication year, author information, institutional affiliations, country, funding sources, keywords, subject areas, and citation counts, which illustrate the development trends and evolution of digital biomarkers in AD. Additionally, we applied Louis' method to extract the disciplinary backgrounds of each researcher, enabling us to identify the interdisciplinary nature of the field<sup>216</sup>. The emerging trends of various digital biomarkers, as well as the devices and task paradigms used, were categorized and summarized after thoroughly reviewing the full text of each article. The second table was used to extract key information related to

AI models. This table summarized essential details such as the types of research, digital biomarkers, collection devices, tasks, algorithm types, performance distributions, model validation and calibration, and data processing methods. The extraction of collection paradigms was detailed, specifying the names of devices and types of tasks. Model performance was presented through boxplots, and the mean values were calculated. Other significant data were displayed either visually or in tabular form. The extraction process was conducted by two evaluators (WQ and YS). Discrepancies were resolved through discussion, and if consensus could not be reached, a third author (GX) was consulted. The two data extraction tables are provided in Supplementary Table 33 and Supplementary Table 34, respectively.

### Data cleaning

In the multidimensional landscape analysis, we standardized the representations of the same domain across different databases. For authors with similar names, we conducted a further review to determine if they were the same individual. This verification process included checking the consistency of their ORCID (Open Researcher and Contributor ID), historical publications, affiliation with the same institution, and information on professional platforms such as ResearchGate. For authors affiliated with multiple institutions, we adopted Seojin Nam's institution cleaning model<sup>217</sup>, using the first-listed institution as the primary affiliation. Additionally, we standardized the abbreviations and full names of all institutions. In our analysis of international collaborations, we accepted cases where authors were affiliated with multiple international institutions, as this could indicate potential visiting scholars or other forms of international cooperation. For funding analysis, we reviewed and consolidated various representations of sponsor names (e.g., full names and abbreviations) to ensure consistency. To ensure the uniformity and accuracy of keywords in co-occurrence analysis, we used the bibliometrix package in R to merge synonyms. A full list of merged keywords is provided in Supplementary Table 35.

### Data synthesis

After data extraction, we employed a narrative synthesis approach to summarize the multidimensional insights from the data extraction tables. For multidimensional landscape analysis, the publication volume analysis used the least squares polynomial method to fit the trend line, with  $R^2$  indicating the fit quality<sup>218</sup>. The Compound Annual Growth Rate (CAGR)

was calculated using: Growth rate = ([number of publications in the last year or number of publications in the first year]1/(last year – first year) – 1) × 100<sup>219</sup>. Joinpoint software (version 4.8.0.1) was used to evaluate time trends, identify significant inflection points, and calculate slope changes, with  $p < 0.05$  considered statistically significant<sup>220</sup>. Author productivity was analyzed using Price's Law, identifying core authors<sup>24</sup>. National analysis included global publication and regional density maps from ScimagoGraphica. International collaboration intensity was assessed via co-occurrence matrices and visualized with chord diagrams. Disciplinary publication patterns were analyzed using time-sliced WoSCC-indexed literature and VOSviewer. High-frequency keywords were clustered in VOSviewer, with the minimum frequency determined by Price's Law<sup>24</sup>. The interdisciplinary collaboration was visualized in the matrix by the level of participation from each discipline and the number of contributors to digital biomarker research.

Given the variety of biomarker types, collection paradigms, AI approaches, and evaluation techniques, our analysis of AI models covered several aspects. We began by examining study characteristics, with a focus on demographic information and study design. We then summarized the types and sources of data used across the studies. On the technical aspects, we evaluated the AI modeling methods, including data processing and model validation. Our synthesis also explored the performance metrics reported, highlighting the best-performing models. We also addressed the causes of data loss during the data collection process and explored multimodal AI modeling approaches. Regarding research transparency and reproducibility, we evaluated the availability of code and adherence to established reporting standards in each study.

## Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files, and the original data and detailed analysis methods can be shared upon reasonable request to the corresponding author.

## Code availability

The code used in the analysis of this study can be made available from the corresponding author upon reasonable request.

Received: 15 October 2024; Accepted: 15 April 2025;

Published online: 16 June 2025

## References

- Crowell, V. et al. Disease severity and mortality in Alzheimer's disease: an analysis using the U.S. National Alzheimer's Coordinating Center Uniform Data Set. *BMC neurology* **23**, 302 (2023).
- 2023 Alzheimer's disease facts and figures. *Alzheimer's & dementia: J. Alzheimer's Assoc.* **19**, 1598–1695 (2023).
- Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet (Lond., Engl.)* **396**, 413–446 (2020).
- Scheltens, P. et al. Alzheimer's disease. *Lancet (Lond., Engl.)* **397**, 1577–1590 (2021).
- Li, X. et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019. *Front. aging Neurosci.* **14**, 937486 (2022).
- Jack, C. R. Jr. et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **14**, 535–562 (2018).
- Fagan, A. M. et al. Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch. Neurol.* **64**, 343–349 (2007).
- Klunk, W. E. et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann. Neurol.* **55**, 306–319 (2004).
- Lemmens, S. et al. Combination of snapshot hyperspectral retinal imaging and optical coherence tomography to identify Alzheimer's disease patients. *Alzheimer's Res. Ther.* **12**, 144 (2020).
- Yilmaz, A. et al. Targeted Metabolic Profiling of Urine Highlights a Potential Biomarker Panel for the Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment: A Pilot Study. *Metabolites* **10**, <https://doi.org/10.3390/metabo10090357> (2020).
- Palmqvist, S. et al. Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA* **324**, 772–781 (2020).
- Vasudevan, S., Saha, A., Tarver, M. E. & Patel, B. Digital biomarkers: Convergence of digital health technologies and biomarkers. *NPJ digital Med.* **5**, 36 (2022).
- Dorsey, E. R., Papapetropoulos, S., Xiong, M. & Kieburtz, K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digital Biomark.* **1**, 6–13 (2017).
- Byun, S. et al. Exploring shared neural substrates underlying cognition and gait variability in adults without dementia. *Alzheimer's Res. Ther.* **15**, 206 (2023).
- Celik, Y., Stuart, S., Woo, W. L., Sejdic, E. & Godfrey, A. J. I. F. Multi-modal gait: A wearable, algorithm and data fusion approach for clinical and free-living assessment. *Int. Fusion* **78**, 57–70 (2022).
- Mengoudi, K. et al. Augmenting Dementia Cognitive Assessment With Instruction-Less Eye-Tracking Tests. *IEEE J. Biomed. health Inform.* **24**, 3066–3075 (2020).
- Cay, G. et al. Harnessing Speech-Derived Digital Biomarkers to Detect and Quantify Cognitive Decline Severity in Older Adults. *Gerontology* **70**, 429–438 (2024).
- Piau, A., Wild, K., Mattek, N. & Kaye, J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. *J. Med. Internet Res.* **21**, e12785 (2019).
- Dubois, B., von Arnim, C. A. F., Burnie, N., Bozeat, S. & Cummings, J. Biomarkers in Alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants. *Alzheimer's Res. Ther.* **15**, 175 (2023).
- Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. *NPJ dig. Med.* **2**, <https://doi.org/10.1038/s41746-019-0084-2> (2019).
- Li, R. et al. Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions. *J. Biomed. Info.* **127**, 104030 (2022).
- Bazarbekov, I. et al. A review of artificial intelligence methods for Alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis. *Biomed. Signal Process. Control* **92**, 106023 (2024).
- Tricco, A. C. et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **169**, 467–473 (2018).
- Price, D. J. Networks of Scientific Papers. *Sci. (N.Y., N.Y.)* **149**, 510–515 (1965).
- Fujita, K. et al. Postural Control Characteristics in Alzheimer's Disease, Dementia With Lewy Bodies, and Vascular Dementia. *J. Gerontol. Ser. A, Biol. Sci. Med. Sci.* **79**, <https://doi.org/10.1093/gerona/glae061> (2024).
- Wang, X. et al. Gait Indicators Contribute to Screening Cognitive Impairment: A Single- and Dual-Task Gait Study. *Brain sciences* **13**, <https://doi.org/10.3390/brainsci13010154> (2023).
- Yamada, Y. et al. Characteristics of Drawing Process Differentiate Alzheimer's Disease and Dementia with Lewy Bodies. *J. Alzheimer's Dis.: JAD* **90**, 693–704 (2022).
- Yu, N. Y. & Chang, S. H. Characterization of the fine motor problems in patients with cognitive dysfunction - A computerized handwriting

- analysis. *Hum. Movement Sci.* **65**, <https://doi.org/10.1016/j.humov.2018.06.006> (2019).
29. Miyazaki, A. et al. Association between upper limb movements during drumming and cognition in older adults with cognitive impairment and dementia at a nursing home: a pilot study. *Front. Rehabil. Sci.* **4**, 1079781 (2023).
  30. Imaoka, Y., Flury, A. & de Bruin, E. D. Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Front. psychiatry* **11**, 572938 (2020).
  31. Davis, R. & Sikorskii, A. Eye Tracking Analysis of Visual Cues during Wayfinding in Early Stage Alzheimer's Disease. *Dement. Geriatr. Cogn. Disord.* **49**, 91–97 (2020).
  32. Li, D. et al. Automating the analysis of eye movement for different neurodegenerative disorders. *Computers Biol. Med.* **170**, 107951 (2024).
  33. Wei, J. & Boger, J. Sleep Detection for Younger Adults, Healthy Older Adults, and Older Adults Living With Dementia Using Wrist Temperature and Actigraphy: Prototype Testing and Case Study Analysis. *JMIR mHealth uHealth* **9**, e26462 (2021).
  34. Mielke, M. M. et al. Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **11**, 1367–1376 (2015).
  35. Buegler, M. et al. Digital biomarker-based individualized prognosis for people at risk of dementia. *Alzheimer's & dementia (Amsterdam, Netherlands)* **12**, e12073 (2020).
  36. Lancaster, C. et al. Evaluating the Feasibility of Frequent Cognitive Assessment Using the Mezurio SmartphoneApp: Observational and Interview Study in Adults With Elevated Dementia Risk. *JMIR mHealth uHealth* **8**, e16142 (2020).
  37. Poosri, T. et al. Gait smoothness during high-demand motor walking tasks in older adults with mild cognitive impairment. *PLoS one* **19**, e0296710 (2024).
  38. López-García, S. et al. Sleep Time Estimated by an Actigraphy Watch Correlates With CSF Tau in Cognitively Unimpaired Elders: The Modulatory Role of APOE. *Front. aging Neurosci.* **13**, 663446 (2021).
  39. Yamada, Y. et al. Distinct eye movement patterns to complex scenes in Alzheimer's disease and Lewy body disease. *Front. Neurosci.* **18**, 1333894 (2024).
  40. Gramkow, M. H. et al. Diagnostic performance of light reflex pupillometry in Alzheimer's disease. *Alzheimer's & dementia (Amsterdam, Netherlands)* **16**, e12628 (2024).
  41. Davoudi, A. et al. Classifying Non-Dementia and Alzheimer's Disease/Vascular Dementia Patients Using Kinematic, Time-Based, and Visuospatial Parameters: The Digital Clock Drawing Test. *J. Alzheimer's Dis.: JAD* **82**, 47–57 (2021).
  42. Konig, A. et al. Use of Speech Analyses within a Mobile Application for the Assessment of Cognitive Impairment in Elderly People. *Curr. Alzheimer Res.* **15**, 120–129 (2018).
  43. Jang, H. et al. Classification of Alzheimer's Disease Leveraging Multi-task Machine Learning Analysis of Speech and Eye-Movement Data. *Front. Hum. Neurosci.* **15**, 716670 (2021).
  44. Vizer, L. M. & Sears, A. Classifying Text-Based Computer Interactions for Health Monitoring. *IEEE pervasive Comput.* **14**, 64–71 (2015).
  45. Bayat, S. et al. GPS driving: a digital biomarker for preclinical Alzheimer disease. *Alzheimer's Res. Ther.* **13**, 115 (2021).
  46. Ren, J. et al. Radar-based gait analysis by Transformer-like network for dementia diagnosis. *Biomed. Signal Process. Control* **91**, 105986 (2024).
  47. Cherachapridi, P. et al. Prescreening MCI and Dementia Using Shank-Mounted IMU During TUG Task. *IEEE Sens. J.* **22**, 24550–24558 (2022).
  48. De Cock, A. M. et al. Gait characteristics under different walking conditions: Association with the presence of cognitive impairment in community-dwelling older people. *PLoS one* **12**, e0178566 (2017).
  49. Gietzelt, M., Wolf, K. H., Kohlmann, M., Marschollek, M. & Haux, R. Measurement of accelerometry-based gait parameters in people with and without dementia in the field: a technical feasibility study. *Methods Inf. Med.* **52**, 319–325 (2013).
  50. Kaser, A. N. et al. A novel speech analysis algorithm to detect cognitive impairment in a Spanish population. *Front. Neurol.* **15**, 1342907 (2024).
  51. Igarashi, T., Umeda-Kameyama, Y., Kojima, T., Akishita, M. & Nihei, M. Questionnaires for the Assessment of Cognitive Function Secondary to Intake Interviews in In-Hospital Work and Development and Evaluation of a Classification Model Using Acoustic Features. *Sensors (Basel, Switzerland)* **23**, <https://doi.org/10.3390/s23115346> (2023).
  52. Gu, D. et al. A Stable and Scalable Digital Composite Neurocognitive Test for Early Dementia Screening Based on Machine Learning: Model Development and Validation Study. *J. Med. Internet Res.* **25**, e49147 (2023).
  53. Chan, J. Y. C. et al. Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study. *J. Med. Internet Res.* **22**, e17332 (2020).
  54. Dawadi, P. N., Cook, D. J., Schmitter-Edgecombe, M. & Parsey, C. Automated assessment of cognitive health using smart home technologies. *Technol. health care: Off. J. Eur. Soc. Eng. Med.* **21**, 323–343 (2013).
  55. Minamisawa, A., Okada, S., Inoue, K. & Noguchi, M. Dementia scale score classification based on daily activities using multiple sensors. *IEEE Access* **10**, 38931–38943 (2022).
  56. Kim, J., Cheon, S. & Lim, J. IoT-based unobtrusive physical activity monitoring system for predicting dementia. *IEEE Access* **10**, 26078–26089 (2022).
  57. Zhang, W. et al. Combination of Paper and Electronic Trail Making Tests for Automatic Analysis of Cognitive Impairment: Development and Validation Study. *J. Med. Internet Res.* **25**, e42637 (2023).
  58. Zhang, X. et al. A tablet-based multi-dimensional drawing system can effectively distinguish patients with amnestic MCI from healthy individuals. *Sci. Rep.* **14**, 982 (2024).
  59. Ter Huurne, D. et al. Validation of an Automated Speech Analysis of Cognitive Tasks within a Semiautomated Phone Assessment. *Digital Biomark.* **7**, 115–123 (2023).
  60. Hajjar, I. et al. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. *Alzheimer's & dementia (Amsterdam, Netherlands)* **15**, e12393 (2023).
  61. Ntracha, A. et al. Detection of Mild Cognitive Impairment Through Natural Language and Touchscreen Typing Processing. *Front. digital health* **2**, 567158 (2020).
  62. Khan, T. & Jacobs, P. G. Prediction of Mild Cognitive Impairment Using Movement Complexity. *IEEE J. Biomed. health Inform.* **25**, 227–236 (2021).
  63. Akl, A., Taati, B. & Mihailidis, A. Autonomous unobtrusive detection of mild cognitive impairment in older adults. *IEEE Trans. bio-Med. Eng.* **62**, 1383–1394 (2015).
  64. Tang, F., Uchendu, I., Wang, F., Dodge, H. H. & Zhou, J. Scalable diagnostic screening of mild cognitive impairment using AI dialogue agent. *Sci. Rep.* **10**, 5732 (2020).
  65. Asgari, M., Gale, R., Wild, K., Dodge, H. & Kaye, J. Automatic Assessment of Cognitive Tests for Differentiating Mild Cognitive Impairment: A Proof of Concept Study of the Digit Span Task. *Curr. Alzheimer Res.* **17**, 658–666 (2020).
  66. Jeon, Y. et al. Early Alzheimer's disease diagnosis using wearable sensors and multilevel gait assessment: A machine learning ensemble approach. *IEEE Sens. J.* **23**, 10041–10053 (2023).

67. Shahzad, A., Dadlani, A., Lee, H. & Kim, K. Automated prescreening of mild cognitive impairment using shank-mounted inertial sensors based gait biomarkers. *IEEE Access* **10**, 15835–15844 (2022).
68. Seifallahi, M. et al. Detection of mild cognitive Impairment from gait using Adaptive Neuro-Fuzzy Inference system. *Biomed. Signal Process. Control* **71**, 103195 (2022).
69. Seifallahi, M., Mehraban, A. H., Galvin, J. E. & Ghoraani, B. Alzheimer's Disease Detection Using Comprehensive Analysis of Timed Up and Go Test via Kinect V.2 Camera and Machine Learning. *IEEE Trans. neural Syst. rehabilitation Eng.: a Publ. IEEE Eng. Med. Biol. Soc.* **30**, 1589–1600 (2022).
70. Ghoraani, B. et al. Detection of Mild Cognitive Impairment and Alzheimer's Disease using Dual-task Gait Assessments and Machine Learning. *Biomed. Sig. Process. Control* **64**, <https://doi.org/10.1016/j.bspc.2020.102249> (2021).
71. Seifallahi, M., Soltanizadeh, H., Hassani Mehraban, A. & Khamseh, F. Alzheimeras disease detection using skeleton data recorded with Kinect camera. *Clust. Comput.* **23**, 1469–1481 (2020).
72. Seifallahi, M., Galvin, J. E. & Ghoraani, B. Detection of mild cognitive impairment using various types of gait tests and machine learning. *Front. Neurol.* **15**, 1354092 (2024).
73. Hall, J. B. et al. Feasibility of Using a Novel, Multimodal Motor Function Assessment Platform With Machine Learning to Identify Individuals With Mild Cognitive Impairment. *Alzheimer Dis. associated Disord.* **38**, 344–350 (2024).
74. Kobayashi, M. et al. Automated Early Detection of Alzheimer's Disease by Capturing Impairments in Multiple Cognitive Domains with Multiple Drawing Tasks. *J. Alzheimer's Dis.: JAD* **88**, 1075–1089 (2022).
75. Li, K. et al. A new early warning method for mild cognitive impairment due to Alzheimer's disease based on dynamic evaluation of the "spatial executive process". *Digital health* **9**, 20552076231194938 (2023).
76. Zhang, Y. et al. What can "drag & drop" tell? Detecting mild cognitive impairment by hand motor function assessment under dual-task paradigm. *Int. J. Hum. Comput. Stud.* **145**, 102547 (2021).
77. Ghaderyan, P., Abbasi, A. & Saber, S. A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of Alzheimer's disease. *Expert Syst. Appl.* **114**, 428–440 (2018).
78. Müller, S. et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *Sci. Rep.* **9**, 3543 (2019).
79. Sweidan, J., El-Yacoubi, M. A. & Rigaud, A. S. Explainability of CNN-based Alzheimer's disease detection from online handwriting. *Sci. Rep.* **14**, 22108 (2024).
80. Zuo, F. et al. Deep Learning-Based Eye-Tracking Analysis for Diagnosis of Alzheimer's Disease Using 3D Comprehensive Visual Stimuli. *IEEE J. Biomed. health Inform.* **28**, 2781–2793 (2024).
81. Yin, Y. et al. Internet of things for diagnosis of alzheimeras disease: A multimodal machine learning approach based on eye movement features. *IEEE Internet Things J.* **10**, 11476–11485 (2023).
82. Opwonya, J., Ku, B., Lee, K. H., Kim, J. I. & Kim, J. U. Eye movement changes as an indicator of mild cognitive impairment. *Front. Neurosci.* **17**, 1171417 (2023).
83. Sun, J., Liu, Y., Wu, H., Jing, P. & Ji, Y. A novel deep learning approach for diagnosing Alzheimer's disease based on eye-tracking data. *Front. Hum. Neurosci.* **16**, 972773 (2022).
84. Pereira, M. et al. Visual Search Efficiency in Mild Cognitive Impairment and Alzheimer's Disease: An Eye Movement Study. *J. Alzheimer's Dis.: JAD* **75**, 261–275 (2020).
85. Pavisic, I. M. et al. Eyetracking Metrics in Young Onset Alzheimer's Disease: A Window into Cognitive Visual Functions. *Front. Neurol.* **8**, 377 (2017).
86. Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A. & Agichtein, E. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *J. Neurosci. methods* **201**, 196–203 (2011).
87. Li, Q. et al. Construction of a prediction model for Alzheimer's disease using an AI-driven eye-tracking task on mobile devices. *Aging Clin. Exp. Res.* **37**, 9 (2024).
88. Song, J. et al. Diagnostic Potential of Eye Movements in Alzheimer's Disease via a Multiclass Machine Learning Model. *Cognit. Comput.* **16**, 3364–3378 (2024).
89. Ambrosini, E. et al. Automatic Spontaneous Speech Analysis for the Detection of Cognitive Functional Decline in Older Adults: Multilanguage Cross-Sectional Study. *JMIR aging* **7**, e50537 (2024).
90. García-Gutiérrez, F. et al. Harnessing acoustic speech parameters to decipher amyloid status in individuals with mild cognitive impairment. *Front. Neurosci.* **17**, 1221401 (2023).
91. Metarugcheep, S. et al. Selecting the Most Important Features for Predicting Mild Cognitive Impairment from Thai Verbal Fluency Assessments. *Sensors (Basel, Switzerland)* **22**, <https://doi.org/10.3390/s22155813> (2022).
92. Yamada, Y. et al. Tablet-Based Automatic Assessment for Early Detection of Alzheimer's Disease Using Speech Responses to Daily Life Questions. *Front. digital health* **3**, 653904 (2021).
93. Tanaka, H. et al. Detecting Dementia Through Interactive Computer Avatars. *IEEE J. Transl. Eng. health Med.* **5**, 2200111 (2017).
94. Castegnaro, A. et al. Assessing mild cognitive impairment using object-location memory in immersive virtual environments. *Hippocampus* **32**, 660–678 (2022).
95. Valladares-Rodríguez, S., Fernández-Iglesias, M. J., Anido-Rifón, L. E. & Pacheco-Lorenzo, M. Evaluation of the Predictive Ability and User Acceptance of Panoramix 2.0, an AI-Based E-Health Tool for the Detection of Cognitive Impairment. *Electronics* **11**, 3424 (2022).
96. Tsai, C. F. et al. A Machine-Learning-Based Assessment Method for Early-Stage Neurocognitive Impairment by an Immersive Virtual Supermarket. *IEEE Trans. neural Syst. rehabilitation Eng.: a Publ. IEEE Eng. Med. Biol. Soc.* **29**, 2124–2132 (2021).
97. Zygouris, S. et al. Detection of Mild Cognitive Impairment in an At-Risk Group of Older Adults: Can a Novel Self-Administered Serious Game-Based Screening Test Improve Diagnostic Accuracy? *J. Alzheimer's Dis.: JAD* **78**, 405–412 (2020).
98. Valladares-Rodriguez, S., Fernández-Iglesias, M. J., Anido-Rifón, L., Facal, D. & Pérez-Rodríguez, R. Episodix: a serious game to detect cognitive impairment in senior adults. A psychometric study. *PeerJ* **6**, e5478 (2018).
99. Valladares-Rodriguez, S. et al. Learning to detect cognitive impairment through digital games and machine learning techniques. *Methods Inf. Med.* **57**, 197–207 (2018).
100. Wu, J. et al. An Effective Test (EOmcISS) for Screening Older Adults With Mild Cognitive Impairment in a Community Setting: Development and Validation Study. *J. Med. Internet Res.* **25**, e40858 (2023).
101. Alegret, M. et al. A computerized version of the Short Form of the Face-Name Associative Memory Exam (FACEmemory®) for the early detection of Alzheimer's disease. *Alzheimer's Res. Ther.* **12**, 25 (2020).
102. Greene, B. et al. The Dual Task Ball Balancing Test and Its Association With Cognitive Function: Algorithm Development and Validation. *J. Med. Internet Res.* **26**, e49794 (2024).
103. Wang, Y. et al. An Ensemble Learning Algorithm for Cognitive Evaluation by an Immersive Virtual Reality Supermarket. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 3761–3772 (2024).

104. Alegret, M. et al. FACEmemory<sup>®</sup>, an Innovative Self-Administered Online Memory Assessment Tool. *J. Clin. Med.* **13**, <https://doi.org/10.3390/jcm13237274> (2024).
105. Park, B. et al. Integrating Biomarkers From Virtual Reality and Magnetic Resonance Imaging for the Early Detection of Mild Cognitive Impairment Using a Multimodal Learning Approach: Validation Study. *J. Med. Internet Res.* **26**, e54538 (2024).
106. Chen, S. et al. A Multi-Modal Classification Method for Early Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease Using Three Paradigms With Various Task Difficulties. *IEEE Trans. neural Syst. rehabilitation Eng.: a Publ. IEEE Eng. Med. Biol. Soc.* **32**, 1477–1486 (2024).
107. Wu, R. et al. Screening for Mild Cognitive Impairment with Speech Interaction Based on Virtual Reality and Wearable Devices. *Brain sciences* **13**, <https://doi.org/10.3390/brainsci13081222> (2023).
108. Li, A. et al. Synergy through integration of digital cognitive tests and wearable devices for mild cognitive impairment screening. *Front. Hum. Neurosci.* **17**, 1183457 (2023).
109. Kim, S. Y. et al. Digital Marker for Early Screening of Mild Cognitive Impairment Through Hand and Eye Movement Analysis in Virtual Reality Using Machine Learning: First Validation Study. *J. Med. Internet Res.* **25**, e48093 (2023).
110. Chai, J. et al. Classification of mild cognitive impairment based on handwriting dynamics and qEEG. *Computers Biol. Med.* **152**, 106418 (2023).
111. Yamada, Y. et al. Combining Multimodal Behavioral Data of Gait, Speech, and Drawing for Classification of Alzheimer's Disease and Mild Cognitive Impairment. *J. Alzheimer's Dis.: JAD* **84**, 315–327 (2021).
112. Jiang, J. et al. A Novel Detection Tool for Mild Cognitive Impairment Patients Based on Eye Movement and Electroencephalogram. *J. Alzheimer's Dis.: JAD* **72**, 389–399 (2019).
113. Boudaya, A., Chaabene, S., Bouaziz, B., Hökelmann, A. & Chaari, L. Mild Cognitive Impairment detection based on EEG and HRV data. *Digit. Signal Process.* **147**, 104399 (2024).
114. Qi, H. et al. A Study of Assisted Screening for Alzheimer's Disease Based on Handwriting and Gait Analysis. *J. Alzheimer's Dis.: JAD* **101**, 75–89 (2024).
115. Kwon, L. N. et al. Automated Classification of Normal Control and Early-Stage Dementia Based on Activities of Daily Living (ADL) Data Acquired from Smart Home Environment. *Int. J. Environ. Res. Public Health* **18**, <https://doi.org/10.3390/ijerph182413235> (2021).
116. Kim, S. K., Kim, H., Kim, S. H., Kim, J. B. & Kim, L. Electroencephalography-based classification of Alzheimer's disease spectrum during computer-based cognitive testing. *Sci. Rep.* **14**, 5252 (2024).
117. Perez-Valero, E., Lopez-Gordo, M., Gutierrez, C. M., Carrera-Munoz, I. & Vilchez-Carrillo, R. M. A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearable EEG. *Computer methods Prog. biomedicine* **220**, 106841 (2022).
118. Lee, K., Choi, K. M., Park, S., Lee, S. H. & Im, C. H. Selection of the optimal channel configuration for implementing wearable EEG devices for the diagnosis of mild cognitive impairment. *Alzheimer's Res. Ther.* **14**, 170 (2022).
119. Hou, C.-J., Chen, Y.-T., Capilayan, M. A., Huang, M.-W. & Huang, J.-J. Analysis of heart rate variability and game performance in normal and cognitively impaired elderly subjects using serious games. *Appl. Sci.* **12**, 4164 (2022).
120. Hata, M. et al. Utilizing portable electroencephalography to screen for pathology of Alzheimer's disease: a methodological advancement in diagnosis of neurodegenerative diseases. *Front. psychiatry* **15**, 1392158 (2024).
121. Hanczár, G. et al. Detection of mild cognitive impairment based on mouse movement data of trail making test. *Inform. Med. Unlocked*. **35**, 101120 (2022).
122. Ghosh, A., Puthusserypady, V., Chan, D., Mascolo, C. & Hornberger, M. Machine learning detects altered spatial navigation features in outdoor behaviour of Alzheimer's disease patients. *Sci. Rep.* **12**, 3160 (2022).
123. Kline, A. et al. Multimodal machine learning in precision health: A scoping review. *NPJ digital Med.* **5**, 171 (2022).
124. Noble, D., Mathur, R., Dent, T., Meads, C. & Greenhalgh, T. Risk models and scores for type 2 diabetes: systematic review. *BMJ (Clin. Res. ed.)* **343**, d7163 (2011).
125. Royston, P., Moons, K. G., Altman, D. G. & Vergouwe, Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clin. Res. ed.)* **338**, b604 (2009).
126. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clin. Res. ed.)* **350**, g7594 (2015).
127. Bhatt, P., Liu, J., Gong, Y., Wang, J. & Guo, Y. Emerging Artificial Intelligence-Empowered mHealth: Scoping Review. *JMIR mHealth uHealth* **10**, e35053 (2022).
128. Junaid, S. B. et al. Recent Advancements in Emerging Technologies for Healthcare Management Systems: A Survey. *Healthcare (Basel, Switzerland)* **10**, <https://doi.org/10.3390/healthcare10101940> (2022).
129. Adams, J. L. et al. Digital Technology in Movement Disorders: Updates, Applications, and Challenges. *Curr. Neurol. Neurosci. Rep.* **21**, 16 (2021).
130. Petermann-Rocha, F. et al. Associations between physical frailty and dementia incidence: a prospective study from UK Biobank - Authors' reply. *Lancet Healthy Longev.* **2**, e68 (2021).
131. Montag, C., Elhai, J. D. & Dagum, P. On Blurry Boundaries When Defining Digital Biomarkers: How Much Biology Needs to Be in a Digital Biomarker? *Front. Psychiatry* **12**, 740292 (2021).
132. Macias Alonso, A. K., Hirt, J., Woelfle, T., Janiaud, P. & Hemkens, L. G. Definitions of digital biomarkers: a systematic mapping of the biomedical literature. *BMJ health & care informatics* **31**, <https://doi.org/10.1136/bmjhci-2023-100914> (2024).
133. Daniore, P. et al. From wearable sensor data to digital biomarker development: ten lessons learned and a framework proposal. *NPJ Digital Med.* **7**, 161 (2024).
134. Cummings, J., Reiber, C. & Kumar, P. The price of progress: Funding and financing Alzheimer's disease drug development. *Alzheimer's Dement. (N.Y., N.Y.)* **4**, 330–343 (2018).
135. EyeLink Hardware. <https://www.sr-research.com/zh/hardware/>
136. Proprietary Software for GAITRite Walkways. <https://www.gaitrite.com/gait-analysis-software>.
137. Software as a Medical Device (SaMD). <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd>.
138. Lazzaretti, L. & Capone, F. How proximity matters in innovation networks dynamics along the cluster evolution. *A study high. Technol. Appl. cultural goods*. **69**, 5855–5865 (2016).
139. Sunderaraman, P. et al. Design and Feasibility Analysis of a Smartphone-Based Digital Cognitive Assessment Study in the Framingham Heart Study. *J. Am. Heart Assoc.* **13**, e031348 (2024).
140. Weiner, M. W. et al. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **13**, 561–571 (2017).
141. Rawtaer, I. et al. Early Detection of Mild Cognitive Impairment With In-Home Sensors to Monitor Behavior Patterns in Community-Dwelling Senior Citizens in Singapore: Cross-Sectional Feasibility Study. *J. Med. Internet Res.* **22**, e16854 (2020).
142. Doherty, J. M. et al. Adverse driving behaviors increase over time as a function of preclinical Alzheimer's disease biomarkers. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **19**, 2014–2023 (2023).

143. Chimamiwa, G., Giaretta, A., Alirezaie, M., Pecora, F. & Loutfi, A. Are Smart Homes Adequate for Older Adults with Dementia? *Sensors (Basel, Switzerland)* **22**, <https://doi.org/10.3390/s22114254> (2022).
144. Tuena, C., Pupillo, C., Stramba-Badiale, C., Stramba-Badiale, M. & Riva, G. Predictive power of gait and gait-related cognitive measures in amnestic mild cognitive impairment: a machine learning analysis. *Front. Hum. Neurosci.* **17**, 1328713 (2023).
145. Albers, M. W. et al. At the interface of sensory and motor dysfunctions and Alzheimer's disease. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **11**, 70–98 (2015).
146. Yang, Q., Li, X., Ding, X., Xu, F. & Ling, Z. Deep learning-based speech analysis for Alzheimer's disease detection: a literature review. *Alzheimer's Res. Ther.* **14**, 186 (2022).
147. Molitor, R. J., Ko, P. C. & Ally, B. A. Eye movements in Alzheimer's disease. *J. Alzheimer's Dis.: JAD* **44**, 1–12 (2015).
148. Shim, J., Fleisch, E. & Barata, F. Circadian rhythm analysis using wearable-based accelerometry as a digital biomarker of aging and healthspan. *NPJ digital Med.* **7**, 146 (2024).
149. Brzenczek, C., Klopfenstein, Q., Hähnel, T., Fröhlich, H. & Glaab, E. Integrating digital gait data with metabolomics and clinical data to predict outcomes in Parkinson's disease. *NPJ digital Med.* **7**, 235 (2024).
150. Rios-Urrego, C. D., Rusz, J. & Orozco-Arroyave, J. R. Automatic speech-based assessment to discriminate Parkinson's disease from essential tremor with a cross-language approach. *NPJ digital Med.* **7**, 37 (2024).
151. Kamaruddin, N., Wahab, A. & Quek, C. Cultural dependency analysis for understanding speech emotion. *Expert Syst. Appl.* **39**, 5115–5133 (2012).
152. Anthes, E. Alexa, do I have COVID-19? *Nature* **586**, 22–25 (2020).
153. Wang, X. et al. Estimating presymptomatic episodic memory impairment using simple hand movement tests: A cross-sectional study of a large sample of older adults. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **20**, 173–182 (2024).
154. Stringer, G. et al. Assessment of non-directed computer-use behaviours in the home can indicate early cognitive impairment: A proof of principle longitudinal study. *Aging Ment. health* **27**, 193–202 (2023).
155. Lucey, B. P. et al. Sleep and longitudinal cognitive performance in preclinical and early symptomatic Alzheimer's disease. *Brain: a J. Neurol.* **144**, 2852–2862 (2021).
156. Administration, FDA. Network of Digital Health Experts. <https://www.fda.gov/medical-devices/digital-health-center-excellence/network-digital-health-experts>.
157. Németh, A. H. et al. Using Smartphone Sensors for Ataxia Trials: Consensus Guidance by the Ataxia Global Initiative Working Group on Digital-Motor Biomarkers. *Cerebellum (Lond., Engl.)* **23**, 912–923 (2024).
158. Liu, D. S., Abu-Shaban, K., Halabi, S. S. & Cook, T. S. Changes in Radiology Due to Artificial Intelligence That Can Attract Medical Students to the Specialty. *JMIR Med. Educ.* **9**, e43415 (2023).
159. Kusters, R. et al. Interdisciplinary research in artificial intelligence: challenges and opportunities. *Front. Big Data* **3**, 577974 (2020).
160. Godfrey, A., Stuart, S. & Tenaerts, P. Tech world and medicine come together to harness digital medicine. *Maturitas* **127**, 95–96 (2019).
161. Dinov, I. D. et al. Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations. *PloS one* **11**, e0157077 (2016).
162. Mohsen, F., Al-Absi, H. R. H., Yousri, N. A., El Hajj, N. & Shah, Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digital Med.* **6**, 197 (2023).
163. Qi, W. et al. Mapping Knowledge Landscapes and Emerging Trends in AI for Dementia Biomarkers: Bibliometric and Visualization Analysis. *J. Med. Internet Res.* **26**, e57830 (2024).
164. Abd-Alrazaq, A. et al. The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. *NPJ digital Med.* **5**, 87 (2022).
165. Debray, T. P. et al. Meta-analysis and aggregation of multiple published prediction models. *Stat. Med.* **33**, 2341–2362 (2014).
166. Salazar de Pablo, G. et al. Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. *Schizophrenia Bull.* **47**, 284–297 (2021).
167. Adamowicz, L., Christakis, Y., Czech, M. D. & Adamusiaik, T. SciKit Digital Health: Python Package for Streamlined Wearable Inertial Sensor Data Processing. *JMIR mHealth uHealth* **10**, e36762 (2022).
168. Brem, A. K. et al. Digital endpoints in clinical trials of Alzheimer's disease and other neurodegenerative diseases: challenges and opportunities. *Front. Neurol.* **14**, 1210974, <https://doi.org/10.3389/neur.2023.1210974> (2023).
169. Ouyang, X. et al. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 404–419.
170. Ford, E., Milne, R. & Curlewis, K. Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **13**, e1492 (2023).
171. Martinez-Martin, N., Insel, T. R., Dagum, P., Greely, H. T. & Cho, M. K. Data mining for health: staking out the ethical territory of digital phenotyping. *NPJ Digit. Med.* **1**, <https://doi.org/10.1038/s41746-018-0075-8> (2018).
172. Bak-Coleman, J. et al. Create an IPCC-like body to harness benefits and combat harms of digital tech. *Nature* **617**, 462–464 (2023).
173. Godfrey, A. et al. Inertial wearables as pragmatic tools in dementia. *Maturitas* **127**, 12–17 (2019).
174. Del Din, S. et al. Measuring gait with an accelerometer-based wearable: influence of device location, testing protocol and age. *Physiological Meas.* **37**, 1785–1797 (2016).
175. Borges do Nascimento, I. J. et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. *NPJ digital Med.* **6**, 161 (2023).
176. Wall, C. et al. Considering and understanding developmental and deployment barriers for wearable technologies in neurosciences. *Front. Neurosci.* **18**, 1379619 (2024).
177. Shandhi, M. M. H. et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. *NPJ digital Med.* **7**, 44 (2024).
178. Wilbanks, J. T. & Topol, E. J. Stop the privatization of health data. *Nature* **535**, 345–348 (2016).
179. Zhai, B., Elder, G. J. & Godfrey, A. Challenges and opportunities of deep learning for wearable-based objective sleep assessment. *NPJ digital Med.* **7**, 85 (2024).
180. Lyall, D. M. et al. Artificial intelligence for dementia—Applied models and digital health. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **19**, 5872–5884 (2023).
181. Blackman, J. et al. Remote evaluation of sleep to enhance understanding of early dementia due to Alzheimer's Disease (RESTED-AD): an observational cohort study protocol. *BMC geriatrics* **23**, 590 (2023).
182. Hillel, I. et al. Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. *Eur. Rev. aging Phys. Act.: Off. J. Eur. Group Res. into Elder. Phys. Act.* **16**, 6 (2019).
183. Costa, A. & Milne, R. Detecting value(s): Digital biomarkers for Alzheimer's disease and the valuation of new diagnostic technologies. *Sociol. health Illn.* **46**, 261–278 (2024).
184. Meier, I. B. et al. Using a Digital Neuro Signature to measure longitudinal individual-level change in Alzheimer's disease: the Altoida large cohort study. *NPJ digital Med.* **4**, 101 (2021).

185. Wall, C., Hetherington, V. & Godfrey, A. Beyond the clinic: the rise of wearables and smartphones in decentralising healthcare. *NPJ digital Med.* **6**, 219 (2023).
186. *Digital Biomarkers Market*. <https://www.rootsanalysis.com/reports/digital-biomarkers-market.html>.
187. Sadowski, J., Viljoen, S. & Whittaker, M. Everyone should decide how their digital data are used - not just tech companies. *Nature* **595**, 169–171 (2021).
188. Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P. & Riva, G. Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Front. Hum. Neurosci.* **14**, 245 (2020).
189. Grammatikopoulou, M. et al. Assessing the cognitive decline of people in the spectrum of AD by monitoring their activities of daily living in an IoT-enabled smart home environment: a cross-sectional pilot study. *Front. aging Neurosci.* **16**, 1375131 (2024).
190. Weiner, M. W. et al. Increasing participant diversity in AD research: Plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's Disease Neuroimaging Initiative 4. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **19**, 307–317 (2023).
191. Koychev, I., Young, S., Holve, H., Ben Yehuda, M. & Gallacher, J. Dementias Platform UK Clinical Studies and Great Minds Register: protocol of a targeted brain health studies recontact database. *BMJ open* **10**, e040766 (2020).
192. Reid, G. et al. The usability and reliability of a smartphone application for monitoring future dementia risk in ageing UK adults. *Br. J. Psychiatry.: J. Ment. Sci.* **224**, 245–251 (2024).
193. Fu, H., Chen, H. & Xing, G. Demo Abstract: AD-CLIP: privacy-preserving, low-cost synthetic human action dataset for alzheimer's patients via CLIP-based models. *2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 257–258 (2024).
194. Xiangmao, C., Cheng, P., Guoliang, X., Tian, H. & Gang, Z. iSleep: A Smartphone System for Unobtrusive Sleep Quality Monitoring. *ACM Trans. Sens. Netw.* **16**, 27 (2020). (32 pp.)–27 (32 pp.).
195. Li, Y., Yu, D. S. F., Chen, S., Xing, G. & Chen, H. Demo: EmoMarker: A Privacy-Preserving, Multi-Modal Sensing System for Dyadic Digital Biomarkers of Expressed Emotions for Patients with Dementia. The 22nd Annual International Conference on Mobile Systems, Applications and Services (MobiSys), 614–615 (2024).
196. Chang, X., Li, G., Xing, G., Zhu, K. & Tu, L. DeepHeart: A Deep Learning Approach for Accurate Heart Rate Estimation from PPG Signals. *Acm Transactions on Sensor Networks* **17**, <https://doi.org/10.1145/3441626> (2021).
197. Fu, H., Chen, H., Lin, S. & Xing, G. SHADE-AD: An LLM-Based Framework for Synthesizing Activity Data of Alzheimer's Patients. The 23rd ACM Conference on Embedded Networked Sensor Systems (SenSys), (2025).
198. Wu, H., Liu, K., Jiang, S., Zhao, Z., Yan, Z., Xing, G. Demo abstract: Caringfm: An interactive in-home healthcare system empowered by large foundation models. *23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2024
199. Liu, Y. et al. Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response. *Sci. Transl. Med.* **14**, eadc9669 (2022).
200. He, H. et al. What radio waves tell us about sleep! *Sleep* **48**, <https://doi.org/10.1093/sleep/zsae187> (2025).
201. Yang, Y. et al. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat. Med.* **28**, 2207–2215 (2022).
202. Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E. & Herrera, F. Science mapping software tools: Review, analysis, and cooperative study among tools. *J. Am. Soc. Inform. Sci. Tech.* **62**, 1382–1402 (2011).
203. Osinska, V. & Klimas, R. Mapping science: Tools for bibliometric and altmetric studies. *Inf. Res Int. Electron. J.* <https://doi.org/10.47989/irpaper909> (2021).
204. Chen, C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Tech.* **57**, 359–377 (2006).
205. van Eck, N. J. & Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **84**, 523–538 (2010).
206. Aria, M. & Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **11**, 959–975 (2017).
207. *bibliometric*. <https://bibliometric.com/>.
208. *Gephi*. <https://gephi.org/>.
209. *Joinpoint trend analysis software*. <https://surveillance.cancer.gov/joinpoint/>.
210. *Cortex Manager Documentation*. <https://docs.context.net/>.
211. He, D. et al. Virtual Reality Technology in Cognitive Rehabilitation Application: Bibliometric Analysis. *JMIR serious games* **10**, e38315 (2022).
212. De Nooy, W., Mrvar, A. & Batagelj, V. *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software*. Vol. 46 (Cambridge university press, 2018).
213. Aijing L, S. Y. & Lu M. *Medical Literature Information Retrieval. Third Edition.*, (People's Medical Publishing House, 2005).
214. Group, F.-N. B. W. In *BEST (Biomarkers, EndpointS, and other Tools) Resource*, (Food and Drug Administration (US) National Institutes of Health (US), 2016).
215. Sun, Y. M., Wang, Z. Y., Liang, Y. Y., Hao, C. W. & Shi, C. H. Digital biomarkers for precision diagnosis and monitoring in Parkinson's disease. *NPJ digital Med.* **7**, 218 (2024).
216. Agha-Mir-Salim, L. et al. Interdisciplinary collaboration in critical care alarm research: A bibliometric analysis. *Int. J. Med. Inform.* **181**, 105285 (2024).
217. Nam, S., Kim, D., Jung, W. & Zhu, Y. Understanding the Research Landscape of Deep Learning in Biomedical Science: Scientometric Analysis. *J. Med. Internet Res.* **24**, e28114 (2022).
218. Golub, G. H. & Van Loan, C. F. An analysis of the total least squares problem. *SIAM J. Numer. Anal.* **17**, 883–893 (1980).
219. Garrett, S. J. *An Introduction to the Mathematics of Finance: A Deterministic Approach*. (Elsevier Science, 2013).
220. Qiu, H., Cao, S. & Xu, R. Cancer incidence, mortality, and burden in China: a time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. *Cancer Commun. (Lond., Engl.)* **41**, 1037–1048 (2021).
221. Winchester, L. M. et al. Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. *Alzheimer's Dement.: J. Alzheimer's Assoc.* **19**, 5860–5871 (2023).
222. Teh, S. K., Rawtaer, I. & Tan, H. P. Predictive Accuracy of Digital Biomarker Technologies for Detection of Mild Cognitive Impairment and Pre-Frailty Amongst Older Adults: A Systematic Review and Meta-Analysis. *IEEE J. Biomed. health Inform.* **26**, 3638–3648 (2022).

## Acknowledgements

Zhejiang Province Traditional Chinese Medicine, Science, and Technology Project (2023ZF134) supported by SC, Zhejiang Provincial Medical and Health Science and Technology Program Project (2022KY1052) supported by SC, First-Class Course of Zhejiang Province (2022-1133) supported by SC, Basic Public Welfare Research Project/ Joint Fund Project of Zhejiang Province: Research on accurate localization of gait disorder brain region in Parkinson's disease based on pfMRI data set and hierarchical Bayesian model(LBY23H200002), supported by XL.

## Author contributions

S.C. and G.X. contributed to the conceptualization of the study. W.Q. conceptualized the research, established the methodology, performed the full-text data visualization, and drafted the first part of the manuscript. X.Z. and B.W. contributed to the research design and respectively wrote the second and third sections of the manuscript. Y.S. and C.D. provided software and other resources, and assisted in drafting and reviewing the original version of the manuscript. S.S., J.L., and K.Z. organized the data, established the methodology, and contributed to the drafting of the initial version. Y.H., M.Z., S.Y., Y.D., and H.S. handled data cleaning and management, and conducted formal analyses. J.K., X.L., G.J., L.M.M., G.X., S.C., and Z.Y. participated in the classification of digital biomarkers, data extraction, and contributed to the methodology. HF and LP handled data visualization for the entire manuscript and participated in drafting the initial version. X.L., H.C., L.M.M.B., Z.Y., H.C., G.X., and S.C. supervised the research. All authors reviewed the manuscript, approved its submission for publication, and agreed to be accountable for all aspects of the work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at

<https://doi.org/10.1038/s41746-025-01640-z>.

**Correspondence** and requests for materials should be addressed to Guoliang Xing or Shihua Cao.

**Reprints and permissions information** is available at  
<http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025