Review

Artificial Intelligence for Caregivers of Persons With Alzheimer’s Disease and Related Dementias: Systematic Literature Review

Bo Xie¹,², BSc, MSc, PhD; Cui Tao³, PhD; Juan Li⁴, PhD; Robin C Hilsabeck⁵, PhD; Alyssa Aguirre⁵, LCSW

¹School of Nursing, The University of Texas at Austin, Austin, TX, United States
²School of Information, The University of Texas at Austin, Austin, TX, United States
³School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
⁴Department of Computer Science, North Dakota State University, Fargo, ND, United States
⁵Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Bo Xie, BSc, MSc, PhD
School of Nursing
The University of Texas at Austin
1710 Red River
Austin, TX, 78712
United States
Phone: 1 512 232 5788
Email: boxie@utexas.edu

Abstract

Background: Artificial intelligence (AI) has great potential for improving the care of persons with Alzheimer’s disease and related dementias (ADRD) and the quality of life of their family caregivers. To date, however, systematic review of the literature on the impact of AI on ADRD management has been lacking.

Objective: This paper aims to (1) identify and examine literature on AI that provides information to facilitate ADRD management by caregivers of individuals diagnosed with ADRD and (2) identify gaps in the literature that suggest future directions for research.

Methods: Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for conducting systematic literature reviews, during August and September 2019, we performed 3 rounds of selection. First, we searched predetermined keywords in PubMed, Cumulative Index to Nursing and Allied Health Literature Plus with Full Text, PsycINFO, IEEE Xplore Digital Library, and the ACM Digital Library. This step generated 113 nonduplicate results. Next, we screened the titles and abstracts of the 113 papers according to inclusion and exclusion criteria, after which 52 papers were excluded and 61 remained. Finally, we screened the full text of the remaining papers to ensure that they met the inclusion or exclusion criteria; 31 papers were excluded, leaving a final sample of 30 papers for analysis.

Results: Of the 30 papers, 20 reported studies that focused on using AI to assist in activities of daily living. A limited number of specific daily activities were targeted. The studies’ aims suggested three major purposes: (1) to test the feasibility, usability, or perceptions of prototype AI technology; (2) to generate preliminary data on the technology’s performance (primarily accuracy in detecting target events, such as falls); and (3) to understand user needs and preferences for the design and functionality of to-be-developed technology. The majority of the studies were qualitative, with interviews, focus groups, and observation being their most common methods. Cross-sectional surveys were also common, but with small convenience samples. Sample sizes ranged from 6 to 106, with the vast majority on the low end. The majority of the studies were descriptive, exploratory, and lacking theoretical guidance. Many studies reported positive outcomes in favor of their AI technology’s feasibility and satisfaction; some studies reported mixed results on these measures. Performance of the technology varied widely across tasks.

Conclusions: These findings call for more systematic designs and evaluations of the feasibility and efficacy of AI-based interventions for caregivers of persons with ADRD. These gaps in the research would be best addressed through interdisciplinary collaboration, incorporating complementary expertise from the health sciences and computer science/engineering–related fields.

(JMIR Med Inform 2020;8(8):e18189) doi: 10.2196/18189

KEYWORDS
Alzheimer disease; dementia; caregiving; technology; artificial intelligence
Introduction

Alzheimer’s disease and related dementias (ADRD) have become a major public health concern in the United States. An estimated 5.6 million Americans aged 65 and older (10% of the US population) were living with ADRD in 2019, and this number is expected to grow dramatically as the population continues to age. By 2025, the number of Americans aged 65 or older with ADRD is expected to reach 7.1 million, nearly a 27% increase from 2019, and by 2050, this population is projected to be 13.8 million, with the highest growth among those in ADRD’s advanced stage [1].

Persons with ADRD require progressively extensive assistance in their daily lives, the majority of which is provided by family members, friends, and other unpaid caregivers [1]. It is estimated that in 2018, American caregivers of persons with ADRD provided 18.5 billion hours of informal unpaid assistance, valued at $233.9 billion [1]. Family caregivers (hereafter “caregivers”) of persons with ADRD are expected to make important care decisions for their family members with ADRD on a daily basis. However, these caregivers report being unprepared for their roles and responsibilities, uninformed about care options, and unsupported by professionals in their decision making [2-5]. Caregiving for persons with ADRD is stressful [6-10], and it can severely affect the caregiver’s own health and well-being [7]. There is an urgent need to better prepare caregivers to manage their daily lives and those of their family members with ADRD, yet there are critical knowledge gaps regarding the types and amounts of information that caregivers may want to have in order to better manage ADRD. To provide patient-centered care for people with ADRD and enhance caregivers’ quality of life, we must address these gaps.

Artificial intelligence (AI) is showing great promise in areas of health care—in precision treatments, patient education, virtual assistance, and cost reduction [11]. Some attempts have been made to apply AI for persons with ADRD and their caregivers in order to improve patients’ daily functioning, quality of life, and well-being, as well as reduce caregiver burden (eg, social robots to facilitate social interaction and engagement, assistive robots to facilitate daily activities such as handwashing, tea making, or dressing) [12-16]. To date, however, there has been little systematic review to identify research on AI for ADRD management by caregivers and gaps that remain in our understanding of AI for ADRD management. We have conducted this systematic review to identify and examine literature on AI that provides information to facilitate ADRD management by caregivers of individuals diagnosed with ADRD and to identify gaps in the literature that suggest future directions for research.

Methods

Overview

Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for conducting systematic literature reviews and following procedures used in previous systematic literature reviews [17-19], we performed 3 rounds of search in selected databases. Because this review focuses on AI and ADRD management, we searched databases commonly used for research not only in the health sciences but also in computer science and engineering: PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL) Plus with Full Text, PsycINFO, IEEE Xplore Digital Library, and ACM Digital Library. First, we searched titles and abstracts using keywords. Next, we screened the titles and abstracts using inclusion and exclusion criteria. Finally, we screened the papers’ full texts to ensure that they met the inclusion or exclusion criteria.

Round 1: Keyword Search

On August 23, 2019, we searched titles and abstracts in PubMed using the following 3 sets of keywords: (“dementia” OR “Alzheimer”) AND (“caregiver*” OR “proxy” OR “proxies” OR “surrogate*”) AND (“artificial intelligence” OR “intelligent”). These sets of keywords were inclusive but in line with our study’s aims. For the same reason, we did not use built-in limiters in PubMed. This yielded 16 papers. Next, we performed the same search of medical subject heading (MeSH) terms in PubMed, excluding “proxies” and “intelligent,” which are not MeSH terms: (“dementia” OR “Alzheimer”) AND (“caregiver*” OR “proxy” OR “surrogate*”) AND (“artificial intelligence”). This yielded 10 papers. Of the 26 papers from these two searches, 1 was a duplicate, yielding a combined total of 25 nonduplicate papers (4 were reviews; the other 21 reported original data). In addition, we searched both CINAHL Plus with Full Text and PsycINFO, using the same 3 sets of keywords that we used for titles and abstracts in PubMed. CINAHL yielded 10 papers, including 7 duplicates. PsycINFO yielded 10 papers, including 6 duplicates. Excluding duplicates, 7 papers remained, for a total of 32 papers across the 3 health sciences databases.

Again, on September 9, 2019, using the same sets of keywords, a search of all metadata (titles, abstracts, and indexing terms) for all available years in the IEEE Xplore Digital Library yielded 47 papers. We also searched the ACM Digital Library (ACM Full-Text Collection) for abstracts or titles that matched any of the following words or phrases: “Alzheimer’s,” “dementia,” “caregiver,” “proxy,” “proxies,” “surrogate,” “artificial intelligence,” “intelligent.” These results were sorted by relevance, and the first 200 records were manually inspected; this generated 36 papers. No duplicates were found between the ACM and IEEE databases. However, when merged with the first 32 papers from PubMed, CINAHL, and PsycINFO, 2 duplicates were found, yielding a total of 113 nonduplicate papers.

Round 2: Screening of Titles and Abstracts

Next, 3 of the authors (BX, CT, JL) each screened approximately one-third of the titles and abstracts of the 113 papers. The results were cross-examined by the other 2 authors to ensure accuracy and consistency. Differences were resolved through several rounds of discussion. This round of screening was based on the rationale that the focus of our systematic literature review was AI tools that could provide information and service to facilitate ADRD management by caregivers of persons diagnosed with ADRD. Other topics were outside of the scope of our review. Specifically, we removed any paper that met at least one of the
following exclusion criteria: (1) primary focus on using artificial intelligence to automatically collect information from users (eg, via sensors), not to provide information to users (n=32); (2) paper did not report empirical data from human participants (eg, literature review, book review, column/commentary, system architecture; n=9); (3) primary focus on screening, identification, or diagnosis of dementia or detecting or modeling anxiety or burnout in caregivers instead of providing services to persons already diagnosed with dementia or their caregivers (n=5); (4) study participants were paid or volunteer caregivers and did not include any family caregivers (n=3); (5) full text not in English (n=3).

This round of screening resulted in the removal of 52 papers, with 61 papers remaining.

**Round 3: Screening of Full Text**

In the next round of screening, we eliminated 31 more papers because they met at least one of the aforementioned exclusion criteria: (1) did not report empirical data from human participants (n=18); (2) study participants did not include family caregivers (n=4); (3) primary focus on using artificial intelligence to automatically collect information from users (eg, via sensors), not to provide information to users (n=4); (4) technology under investigation was not artificial intelligence (eg, videogames; n=3); (5) primary focus on screening, identification, or diagnosis of dementia or detecting or modeling anxiety or burnout in caregivers (n=1); (6) report of essentially the same content as in another paper (n=1).

A total of 30 papers remained in the final sample [20-49]. The selection process is summarized in Figure 1 according to the PRISMA guidelines [50].
Round 4: Coding of Full Text

The 30 papers in our final sample were coded using a framework consistent with our prior work [17-19], summarizing key information from each paper. The coding included each study’s publication year, study aim, research method, participant characteristics, sample size, country/area where data collection took place, dosage of AI technology (ie, amount and frequency of time exposed to the AI technology), outcome measures, and key findings. The results of the coding are presented in Multimedia Appendix 1. In addition, we assessed levels of evidence reported in the 30 papers [51].

Results

Our initial searches yielded 113 papers. Through multiple rounds of screening, we removed 83 of them to arrive at our final sample. The reasons for excluding these 83 papers are summarized in Table 1.

Table 1. Summary of the reasons for excluded papers.

<table>
<thead>
<tr>
<th>Reason for exclusion</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary focus was using AI&lt;sup&gt;b&lt;/sup&gt; to automatically collect information from users (eg, sensors), not to provide information to users</td>
<td>36</td>
</tr>
<tr>
<td>Did not report empirical data from human participants (eg, literature review, book review, column/commentary, system architecture)</td>
<td>27</td>
</tr>
<tr>
<td>Study participants did not include any family caregivers</td>
<td>7</td>
</tr>
<tr>
<td>Primary focus was on screening/identification/diagnosis of dementia or detecting/modeling anxiety or burnout in caregivers (instead of providing services to persons already diagnosed with dementia or their caregivers)</td>
<td>6</td>
</tr>
<tr>
<td>The technology under investigation was not AI</td>
<td>3</td>
</tr>
<tr>
<td>Full text not in English</td>
<td>3</td>
</tr>
<tr>
<td>Reporting essentially the same content as another paper (that was already included in the final sample)</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>83</td>
</tr>
</tbody>
</table>

<sup>a</sup>Number of excluded papers.
<sup>b</sup>AI: artificial intelligence.

Key characteristics of the final 30 papers are summarized in Multimedia Appendix 1. The papers were published from 2001 to 2019, averaging 2 per year. The number of publications was consistently low (1 per year) until rising in 2008. The year 2018 had the most papers (4), suggesting an increasing interest in our topic. AI technologies included in the 30 papers varied. We categorized these technologies according to their intended use. As Table 2 shows, the majority (20/30, 67%) focused on using AI to assist in activities of daily living. A limited number of specific daily activities were targeted in these studies, particularly handwashing, tea making, and dressing.

Table 2. Summary of artificial intelligence technology’s intended use.

<table>
<thead>
<tr>
<th>AI&lt;sup&gt;a&lt;/sup&gt; technology use</th>
<th>n&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assist in activities of daily living (eg, assistive robots to aid handwashing, tea making, or dressing)</td>
<td>20</td>
</tr>
<tr>
<td>Facilitate social interaction (eg, social robots)</td>
<td>2</td>
</tr>
<tr>
<td>Provide cognitive stimulation (eg, computerized activities to stimulate cognition)</td>
<td>2</td>
</tr>
<tr>
<td>Ensure safe home environments (eg, smart homes)</td>
<td>2</td>
</tr>
<tr>
<td>Educate (eg, through a teleconferencing program or virtual reality platform)</td>
<td>2</td>
</tr>
<tr>
<td>Assist in reminiscence therapy</td>
<td>2</td>
</tr>
<tr>
<td>Total, N</td>
<td>30</td>
</tr>
</tbody>
</table>

<sup>a</sup>AI: artificial intelligence.
<sup>b</sup>Number of papers.

Aims of the 30 studies fell into one of three major categories: (1) to test the feasibility, usability, or perceptions of a prototype AI technology; (2) to generate preliminary data on the technology’s performance (primarily accuracy in detecting target events, such as falls); and (3) to understand user needs and preferences for the design and functionality of to-be-developed technologies. The majority of these studies used qualitative research methods, with interviews, focus groups, and observation being the 3 most common methods. Cross-sectional surveys were also common, but with small convenience samples. The majority of the studies were descriptive, exploratory, and lacking in theoretical guidance; they were not intended to test theory-informed hypotheses.
The sample sizes of all 30 studies were small, ranging from 6 to 106, with the vast majority on the low end. A total of 7 studies reported data from healthy volunteers or health care professionals but did not include actual patients or caregivers as research participants. We included these studies in our analysis because the technologies under consideration were intended for use by patients and caregivers. One study did not report any participant characteristics.

Nearly half of the studies were conducted in Canada (14/30, 47%), 8 of the 30 (27%) in Europe, and 4 of the 30 (13%) in the United States, Israel, Japan, Mexico, and Taiwan each had one study (1/30, 3%). At least 7 of the 30 studies (23%) were conducted in a research lab; 6 others did not report the setting for data collection. The remaining studies took place in a facility (eg, senior living facility, hospital) or private home.

We also analyzed the AI technology's dosage (ie, the amount of time and frequency that users were exposed to the AI technology in each study). A total of 9 of the 30 studies used interviews or surveys, so the dosage criterion was not applicable to them. Among the 21 studies that involved exposure to AI technology, 5 did not report dosage. The dosages reported in the remaining 16 studies varied widely in terms of both total time and frequency of exposure, ranging from as much as 24/7 access for 4 to 6 weeks or 2 hours per week over 12 weeks to as little as 15 to 20 minutes in a single session.

Outcome measures varied widely as well. Overall, they included both objective and subjective measures. Outcome measures included (1) feasibility, satisfaction, and stress, which were subjective measures; (2) performance, such as the accuracy of AI technology in completing its intended task, measured objectively; (3) usability (self-reported ease of use and perceptions of usefulness); (4) usage patterns (eg, which AI features were used, frequency/duration of usage), also measured objectively; and (5) user needs and requirements for the technology, another set of subjective measures.

Many studies reported positive outcomes in favor of the AI technology being studied (or to be developed) in terms of the technology’s feasibility (with acceptability used as the most common measure of feasibility) and satisfaction (positive perceptions of the technology). One of those studies reported a high dropout rate (65%) [22], making it difficult to interpret the study’s reported positive outcomes. One study reported preliminary evidence supporting limited efficacy of a social robot in reducing patients’ stress [29]. Some studies reported mixed results for feasibility and satisfaction, with some participants reporting that they liked the AI technology but others reporting that they did not [23,26,47,49]. Notably, in 2 separate studies, caregivers reported more positive attitudes than did patients toward the use of AI technology in home care [23,49].

Performance of the technology, measured primarily by accuracy in detecting target events, varied widely across different tasks, ranging from as low as 23% in detecting incorrect dressing events [28] to as high as 98% in detection of falls [46]. In assisting with daily activities, assistive AI devices helped reduce patients’ dependence on caregivers [42,43]. Usage patterns also varied widely, ranging from continuous active use to inactive use. Mixed results were reported for the AI technology’s features, with some features easier to use and more popular than others [33].

A range of user needs was identified, including needs for assistance in home care, getting information (about time, schedule, care options, etc), and communication and social interactions. There is a great need for AI technology to provide tailored assistance to meet these user needs [37]. However, several factors make it challenging to design tailored technology. These include variation in patients’ needs and abilities from day to day and even during the day [39], patients’ varying and evolving identities and preferences for a technology’s styles and features [40], users’ diverse technology literacy levels [41], and challenges associated with ethical issues [48], particularly conflicting needs between caregivers and patients [36] and privacy concerns in assisting in private tasks [31,32].

Regardless of the findings, the levels of evidence [51] of all studies in our final sample were low due to their small convenience samples and exploratory research methods.

**Discussion**

AI has great potential for improving the care for persons with ADRD and the quality of life of family caregivers. To date, however, there has been little effort to systematically review literature on AI for ADRD management by caregivers and to determine what still needs to be done to understand the impact of AI on ADRD management. In this study, we have addressed those gaps. We have identified work on AI that provides information to facilitate ADRD management by family caregivers of patients diagnosed with ADRD, and we have identified gaps in existing work, which suggest future directions for research. The majority of the AI studies included in our final sample (20/30, 67%) focused on using AI to assist in activities of daily living. A limited number of specific daily activities were targeted. The aims of the 30 studies suggested three major purposes: (1) to test the feasibility, usability, or perceptions of a prototype AI technology; (2) to generate preliminary data on the technology’s performance (primarily accuracy in detecting target events, such as falls); and (3) to understand user needs and preferences for the design and functionality of to-be-developed technologies. The majority of these studies used qualitative research methods, with interviews, focus groups, and observation being the 3 most common methods. Cross-sectional surveys were also common, but with small convenience samples. The sample sizes of the 30 studies were small, ranging from 6 to 106, with the vast majority on the low end. The majority of the studies were descriptive, exploratory, and lacking in theoretical guidance. Many studies reported positive outcomes in favor of AI technology’s feasibility and user satisfaction; some reported mixed results for these measures. Performance of technology varied widely across different tasks.

Our findings illustrate important characteristics of research to date on the use of AI that provides information to aid ADRD management by family caregivers. First, only a few studies (N=30) have focused on this topic. Given the topic’s interdisciplinary nature, we intentionally searched databases...
Conflicts of Interest

None declared.
Multimedia Appendix 1

Summary of the 30 studies in the final sample.
[DOCX File, 27 KB-Multimedia Appendix 1]

References


21. Amiribesheli M, Bouchachia A. Smart homes design for people with dementia. 2015 Presented at: International Conference on Intelligent Environments; July 15-17, 2015; Prague, Czech Republic URL: https://doi.org/10.1109/IE.2015.33 [doi: 10.1109/ie.2015.33]


