Title

Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review

Authors

Graham, Sarah A

Lee, Ellen E

Jeste, Dilip V

et al.

Publication Date

2020-02-01

DOI

10.1016/j.psychres.2019.112732

Peer reviewed

HHS Public Access

Author manuscript

Psychiatry Res. Author manuscript; available in PMC 2021 February 01.

Published in final edited form as:

Psychiatry Res. 2020 February ; 284: 112732. doi:10.1016/j.psychres.2019.112732.

Artificial Intelligence Approaches to Predicting and Detecting Cognitive Decline in Older Adults: A Conceptual Review

Sarah A. Graham, PhD[1,2], **Ellen E. Lee, MD**[1,2,3], **Dilip V. Jeste, M.D.**[1,2,4], **Ryan Van Patten, PhD**[1,2], **Elizabeth W. Twamley, PhD**[1,2,3], **Camille Nebeker, EdD, MS**[2,5], **Yasunori Yamada, PhD**[6], **Ho-Cheol Kim, PhD**[7], **Colin A. Depp, PhD**[1,2,3]

[1]Department of Psychiatry, University of California San Diego, La Jolla, CA, US

[2]Sam and Rose Stein Institute for Research on Aging, University of California San Diego, La Jolla, CA, US

[3]VA San Diego Healthcare System, San Diego, CA, US

[4]Department of Neurosciences, University of California San Diego, La Jolla CA, US

[5]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, US

[6]Accessibility and Aging (YY), IBM Research-Tokyo, Tokyo, Japan

[7]Scalable Knowledge Intelligence, IBM Research-Almaden, San Jose, CA, US

Abstract

Preserving cognition and mental capacity is critical to aging with autonomy. Early detection of pathological cognitive decline facilitates the greatest impact of restorative or preventative treatments. Artificial Intelligence (AI) in healthcare is the use of computational algorithms that mimic human cognitive functions to analyze complex medical data. AI technologies like machine learning (ML) support the integration of biological, psychological, and social factors when approaching diagnosis, prognosis, and treatment of disease. This paper serves to acquaint clinicians and other stakeholders with the use, benefits, and limitations of AI for predicting, diagnosing, and classifying mild and major neurocognitive impairments, by providing a conceptual overview of this topic with emphasis on the features explored and AI techniques employed. We present studies that fell into six categories of features used for these purposes: 1) sociodemographics; 2) clinical and psychometric assessments; 3) neuroimaging and neurophysiology; 4) electronic health records and claims; 5) novel assessments (e.g., sensors for

Corresponding Author: Dilip V. Jeste, M.D., Senior Associate Dean for Healthy Aging and Senior Care, Estelle and Edgar Levi Memorial Chair in Aging, Distinguished Professor of Psychiatry and Neurosciences, Director, Sam and Rose Stein Institute for Research on Aging, Co-Director, IBM-UCSD Artificial Intelligence for Healthy Living Center, University of California San Diego, 9500 Gilman Drive, Mail Code #0664, La Jolla, CA, USA, 92093-0664, djeste@ucsd.edu, Telephone: (858) 534-4020.

Author Agreement

This statement is to certify that all authors have seen and approved the final version of the manuscript being submitted. All authors warrant that the article is their original work, has not received prior publication, and is not under consideration for publication elsewhere.

digital data); and 6) genomics/other omics. For each category we provide examples of AI approaches, including supervised and unsupervised ML, deep learning, and natural language processing. AI technology, still nascent in healthcare, has great potential to transform the way we diagnose and treat patients with neurocognitive disorders.

Keywords

Dementia; Mild cognitive impairment; Machine learning; Sensors; Natural language processing

I. Introduction

The World Health Organization (WHO) defines healthy aging as *the process of developing and maintaining the functional ability that enables well-being in older age* (World Health Organization, 2019). Cognitive health is one of the most important determinants of functional ability of older adults (Beaton et al., 2015; Dodge et al., 2005; Gross et al., 2011), and is critical to aging with autonomy (Depp and Jeste, Dilip, 2006; Willis et al., 2006). Healthy aging is associated with some cognitive decline in select abilities (e.g., processing speed, fluid reasoning, episodic memory (Der et al., 2010; Eckert, 2011)). A proportion of older adults develop mild cognitive impairment (MCI; labeled mild neurocognitive disorder in the DSM-5 (Association, 2013)), and 5-15% progress to dementia (major neurocognitive disorder) annually (Association, 2013; Mitchell and Shiri-Feshki, 2009, 2008; Petersen, 2011). Worldwide, 50 million people have dementia (World Health Organization, 2019). As there is no known cure for dementia, tools for the earliest possible detection of cognitive decline are necessary to achieve the greatest impact of current and novel treatment approaches to delay pathological cognitive aging (Graham and Depp, 2019).

Unfortunately, early detection of cognitive impairment is a challenging psychometric endeavor due to the insidious progression of symptoms, which, in the early stages, may be mistaken for normal age-related cognitive impairment (Deary et al., 2009; Petersen et al., 2001). MCI can be difficult to clearly identify, due to multiple sets of diagnostic criteria and need for longitudinal follow-up (Brodaty et al., 2017). Furthermore, MCI may precede varying types of dementia and does not lead to dementia in a sizable proportion of patients. Knowing which patients warrant a comprehensive cognitive screening can be challenging for clinicians, and neuropsychological test batteries are time-consuming and require trained administration. An ideal diagnostic tool must be sensitive to the earliest signs of cognitive decline, non-invasive, practical, and scalable for use in clinics worldwide. Similar efforts are already underway (Balota et al., 2010; Patten et al., 2018; Silverberg et al., 2011) with incremental progress, but there remains much room for improvement.

The purpose of this conceptual review is to provide a primer for clinicians on the understanding and use of an exciting new approach to supporting clinical decision-making such as diagnosis, prediction, and differentiation between the various types of MCI and dementias – i.e., artificial intelligence (AI). AI refers to the scientific field within the discipline of computer sciences concerned with building systems or machines (computers) to accomplish tasks that typically require human intelligence, such as making decisions.

Machine learning (ML), deep learning (DL), and natural language processing (NLP) are techniques of AI. For a machine to act intelligently, it needs to learn from data (trained with data). In ML, algorithms are used to enable the machine to learn through structured data input and past experience to detect patterns in the data and use uncovered patterns to predict future human data. ML can be supervised (i.e., tested against dependent variable data that are known or labeled) or unsupervised (i.e., with data that are unknown or unlabeled). DL is a subset of ML that is useful when there is a large amount of complex and unstructured data. DL involves multiple layers of algorithms called artificial neural networks (ANN), each providing a hierarchically different interpretation to the data. NLP is family of techniques that focuses on analysis of natural human language (usually written) and can be integrated with any of the ML approaches. AI applications specifically for drug discovery, causal disease modeling, clinical trials recruitment, and neuropsychiatric symptoms are outside the scope of this review and have been previously examined in the literature (Jiang et al., 2017; Zhavoronkov et al., 2019).

II. Artificial Intelligence Primer for Predicting and Detecting Cognitive Decline

AI in healthcare is the use of computational algorithms and software that mimic human cognitive functions to analyze complex structured and unstructured medical data like images or clinical notes (Jiang et al., 2017; Yu et al., 2018). AI tools use these high-dimensional (i.e., multi-feature) data to determine potential predictors of normal versus pathological changes in cognitive functioning. AI analytic techniques are ideally suited to handle large volumes and complexity of datasets and can do this more efficiently than humans (Raghupathi and Raghupathi, 2014; Wang et al., 2016). Machine learning (ML) is a subset of AI that involves various methods of enabling an algorithm to learn from datasets, or update itself based on new data (Chen et al., 2017; Nevin, 2018). Standard statistics emphasize fitting a specific model and hypothesis testing to understand underlying mechanisms. In contrast, ML algorithms do not require *a priori* hypotheses about relationships among variables, and instead, emphasize prediction accuracy and can often detect unforeseen relationships and complicated nonlinear interactions within data (Graham et al., 2019). The results or “performance” of an AI algorithm depend on the model selected, available data, and the input features the researchers selected to predict an outcome. Below we narrate the most common classes of ML used for healthcare purposes: supervised and unsupervised machine learning (SL and UL) (Bzdok et al., 2018; Fabris et al., 2017; Miotto et al., 2016), and deep learning (DL) (Esteva et al., 2019; Miotto et al., 2017) (Figure 1a), which may or may not involve natural language processing (NLP) (Demner-Fushman et al., 2009; Hirschberg and Manning, 2015) (Figure 1b).

Supervised Learning (SL) approaches require pre-labeled data (e.g., diagnosis of cognitive impairment vs. unimpaired) that serve as known outcomes for training an algorithm along with features derived from additional datastreams (e.g., clinical notes, neuroimaging) (Bzdok et al., 2018; Fabris et al., 2017). The algorithm then determines which features are most predictive of the pre-labeled outcome. The diagnosis of cognitive impairment could be based on either categorical classification (yes or no) or continuous regression (e.g., score on

a neurocognitive assessment) (Fig. 1a). The validity of SL algorithms rely heavily on the "ground truth" behind the labeled outcomes, which may require longitudinal follow-up or other information to bolster the determination of outcomes such as cognitive impairment.

Unsupervised Learning (UL) algorithms are provided with unlabeled data. While the data may contain, for example, individuals with cognitive impairment and those without, the algorithm is not privy to this information (Miotto et al., 2017). Instead, the algorithm searches unstructured data (e.g., clinical notes) for relationships or clusters with the goal of segmenting the data by some shared characteristics, or detecting anomalies that do not belong to a particular group. Identified clusters generally require clinical expertise to derive their meaning (Fig. 1a).

Deep Learning (DL) functions using both SL and UL but is capable of exploiting the unknown structure from data using artificial neural networks (ANNs) that automatically derive features from raw data (i.e., feature engineering) when they learn, instead of requiring human input for obtaining features from raw data (Esteva et al., 2019; Miotto et al., 2017)23,24]. This type of learning requires very large datasets in comparison to other forms of ML that can work with smaller data size and extensive computation power. Complex, high-dimensional data like neuroimaging and speech are well suited to DL (Fig. 1a).

Natural Language Processing (NLP) refers to how computers understand natural language (e.g., speech, text) in terms of language translation, semantic understanding, and summarization (Demner-Fushman et al., 2009; Hirschberg and Manning, 2015). The process of NLP is to transfer text from an unstructured into a structured format to enable analyses. Studies that use NLP generally follow with one of the aforementioned learning techniques (SL, UL, DL) to determine the accuracy of using speech/written/text data to model cognitive function (Fig. 1b).

Performance metrics of AI results

AI studies most commonly report results of algorithm performance as percent accuracy and receiver operating characteristic area under the curve (ROC AUC). Accuracy is the proportion of correct predictions: true positives + true negatives divided by all observations (true positives and negatives + false positives and negatives) (Hossin and Sulaiman, 2015; Huang and Ling, 2005). In comparison, AUC provides information about the tradeoff between sensitivity (true positive rate) and specificity (true negative rate) at various threshold settings. The benefit of using AUC instead of, or in addition to, percent accuracy, is that unlike accuracy this metric is not affected by class imbalance (e.g., a smaller number of subjects in the sample with dementia compared to healthy controls) (Hossin and Sulaiman, 2015).

When evaluating the efficacy or quality of the results of AI studies, we should pay close attention to the validation methods used to arrive at the performance metrics. A study has been internally validated if methods like cross validation (CV) were used. CV is considered "internal" validation because all of the data are used at some point in the training phase (e.g., leave one out CV; 5-fold CV) (Blagus and Lusa, 2015). The performance is reported as the average across the testing folds. CV enables the researcher to double-check the accuracy of a

model on different subsets of data, though the algorithm has not been vetted on a population external to the one used for training. In contrast, external validation involves testing the algorithm performance on a completely different dataset than the training set (Park and Han, 2018). This step is crucial before an algorithm's clinical usefulness can be determined.

III. Overview of Select Studies Focused on AI for Cognitive Decline

We did not perform a meta-analysis of all studies related to neurocognitive disorders and AI. Instead, our goal was to provide a guide to aid clinicians in understanding the heterogeneity and potential value and limitations of a variety of neurocognitive features for AI applications. Using a broad MEDLINE inquiry with several search terms [("artificial intelligence" or "machine learning" or "NLP") AND ("cognition" or "cognitive testing)]," we then selected studies to illustrate the diversity of data sources and research questions addressed using AI, preferring those with larger sample sizes and clear explanations of the ML approach.

We selected studies that showcased common classes of features used for detecting, classifying, or predicting cognitive status and that employed the most common AI techniques emerging in healthcare: SL, UL, DL, and NLP (Jiang et al., 2017). Six feature categories (i.e., types of datasets) emerged from the studies selected: sociodemographic data, clinical and psychometric assessments; neuroimaging and neurophysiological data; EHR and claims data; novel assessments (e.g., handwriting and speech analyses); and genomic and other omic data. Table 1 showcases different AI techniques used with each feature category, with its strengths and limitations.

Sociodemographic Data (Table 1 section A)

Sociodemographic and other forms of population data offer rich information from large datasets (e.g., the US Health and Retirement Study (Michigan, 2019); Ageing and Retirement in Europe (SHARE-ERIC, 2019); Korean Longitudinal Study of Aging (Service., 2015)). Langavant and colleagues (De Langavant et al., 2018) developed an UL-based algorithm for identifying participants with high likelihood of dementia from population-based surveys, without clinical diagnosis, using both American and European subjects, with the potential to flag individuals within a large population-based sample for cognitive screening. The Na study (Na, 2019) used variables commonly collected in community health care institutions (sociodemographics, health, subjective well-being) and found that age and education were particularly important in predicting cognitive decline in a community sample of Korean adults.

A benefit of such data is that they are often stratified geographically and cover various demographic groups. They are also easy to collect for reasonable cost and can be widely disseminated. Such data can potentially help with early risk stratification and subsequent identification of high-risk individuals in need of more detailed assessments (De Langavant et al., 2018). These data may also contain social determinants of health (e.g., education), often overlooked in other clinical data. Because many countries collect population data regarding health, socioeconomic status, and social and family networks of older adults, such information may also provide an opportunity to compare outcomes across different countries

and infer global health estimates of dementia burden. However, simply identifying putative risk and protective factors for cognitive decline from sociodemographic data may be of limited use for predicting future cognitive impairment for an individual. Furthermore, the findings from one country/setting may not relate directly to participants in other nations, e.g., extrapolating from a Korean sample to a US sample or *vice versa*. However, when combined with clinical measures like the Mini-Mental State Examination (MMSE), as shown by Na (Na, 2019), sociodemographics could be a useful addition to a ML algorithm. Multi-modal variables are most meaningful when their complex interactions are analyzed comprehensively, and longitudinally (e.g., Na (Na, 2019)), using ML models.

Clinical and Psychometric Assessments (Table 1 section B)

Clinical assessment data offer readily available, inexpensive, and rich sources of information. The three studies highlighted in this category show how AI techniques can be used to streamline a cognitive assessment battery for dementia (Lins et al., 2017), incorporate information from clinical notes to improve diagnostic accuracy of MCI and dementia (Moreira and Namen, 2018), and best distinguish between normal cognition and MCI using neuropsychological measures (Senanayake et al., 2017). The three studies differ widely in sample size, input data, and algorithms, demonstrating the varied applications of such data. Given that every major healthcare provider collects clinical variables, these data promote generalizability of ML algorithms and can potentially involve large samples if every individual in an area or healthcare system is included. Similar to population-based sociodemographics, clinical data may be best for identifying high-risk individuals who need additional assessments and clinical interventions to help focus resources most efficiently. However, clinical assessments are not streamlined or standardized (primary versus subspecialty settings), and different clinicians may use different measures (e.g., the Montreal Cognitive Assessment (MoCA) versus the MMSE). AI may be able to address the limitations of heterogeneous data by using a heterogeneous training set, or by testing models in different populations. ML techniques may also help to rank the factors that are critical for assessing cognitive impairment and thus help to focus on these factors. The quality and accuracy of clinical data can be variable and require detailed record-keeping and access to the data to be useful for AI.

Neuroimaging and Neurophysiological Data (Table 1 section C)

Neuroimaging and neurophysiological techniques have grown considerably in the past decade. Research continues to demonstrate their use for providing important information about the brain’s structure and function (Khandai and Aizenstein, 2013). Brain imaging is often used to detect neurological causes (e.g., tumors, stroke), but not psychopathology (Vernooij et al., 2019). In the interpretation of radiological images, AI techniques can outperform specialists in detecting early or “preclinical” degradation of neuroanatomy because AI is particularly well suited to detecting abnormalities within image and signal data through training (i.e., pattern recognition) (Ahmed et al., 2019; Hosny et al., 2018). AI offers the potential to improve interpretability and clinical utility of neuroimaging and neurophysiological data that are commonly obtained but incompletely understood. We may learn from AI about new aspects of brain function and connectivity and generate new hypotheses regarding brain-based mechanisms of neuropsychiatric diseases. The four

examples of AI used with brain imaging data show how different EEG (Fan et al., 2018) and brain imaging profiles (Gamberger et al., 2017; Grassi et al., 2018; Iizuka et al., 2019) can be used to identify cognitive impairment (Fan et al., 2018; Iizuka et al., 2019) and predict prognostic trajectories (Gamberger et al., 2017; Grassi et al., 2018) in different populations.

Neuroimaging and neurophysiological data are considered high dimensional data—data where the number of features often greatly exceeds the number of observations. Because most statistical analyses are better suited for lower dimensional data, ML is an ideal alternative for traditional neuroimaging/neurophysiology analyses. Given recent initiatives to grow open source datasets like the Alzheimer's Disease Neuroimaging Initiative (ADNI) (e.g., (Gamberger et al., 2017)), the "big data" required for optimal AI techniques are also available. Ultimately, combining AI techniques with rich biological information contained in neuroimaging will enable faster, safer, cheaper, and more accurate imaging results, usable for informing diagnoses, prognoses, and treatment decisions. However, neuroimaging and neurophysiological assessments are not commonly offered in all medical settings because of high costs and safety issues like radiation exposure. There is also considerable heterogeneity among datasets regarding imaging modalities (MRI versus PET), machines (different strengths of MRI machines), and processing approaches which continue to evolve.

Electronic Health Record (EHR) and Claims Data (Table 1 section D)

The EHR includes huge amounts of patient-specific information containing both structured (coded) and unstructured (free text) entries (Hayrinen et al., 2008). EHR may also contain some sociodemographic and clinical data mentioned above, depending on the vendor and/or health organization. The four highlighted studies in this section used nationwide administrative claims data (Id et al., 2019), EHR from a regional Veterans Affairs healthcare system (a publicly administered program) (Shao et al., 2019), EHR from a regional not-for-profit academic healthcare system (Wang et al., 2019), and EHR from two hospital-based samples (Wang et al., 2018). These datasets record and help to manage patient care and offer a relatively inexpensive source of information collected over long time periods on large numbers of patients. The large size of these databases (i.e., thousands of individuals) enables studies of rare conditions, and the longitudinal aspect of the data enables researchers to investigate effects of treatment(s) over time.

Nori and colleagues (Id et al., 2019) and Wang and colleagues (Wang et al., 2019) utilized the longitudinal nature of EHR data to find features related to increased incidence of near-term (4-5 years) dementia (Id et al., 2019) and mortality (Wang et al., 2019). ML algorithms can deal with very large numbers of potential input features (e.g., (Id et al., 2019) used over 10,000 clinical, pharmaceutical, and demographic variables) and rapidly develop predictive models without specific selection of variables, enabling automated selection of high value predictors. However, EHR data alone have relatively limited predictive power when analyzed in the absence of other social determinants of health (e.g., population-based sociodemographic data)(Freij et al., n.d.).

EHR systems are primarily designed for streamlining billing purposes; thus, the data for deciphering and supporting clinical decision-making may not always be available. The quality and quantity of EHR data are also dependent on external factors (e.g., severity of

illness, insurance rules, regional practices, availability of resources) and are heterogeneous in organization and level of detail. For example, the findings from a regional VA health system (as in (Shao et al., 2019)) may be more representative of care at other VA health systems, whereas there may be considerable regional differences within other nationwide insurers (e.g., Blue Cross Blue Shield versus Kaiser Permanente) due to different patient populations and plan structures. AI will be particularly useful with these data if it can "learn" the different styles of documentation from different providers and different healthcare systems - a excellent area for NLP applications. Finally, AI could help healthcare providers to better and more efficiently understand their patient's clinical history and guide their decision-making process.

Claims data, like those used by (Id et al., 2019), are generated primarily for the administration of payment for health services delivered. These data offer structured information on patient interactions with a healthcare system (e.g., billed services, prescriptions) and has the ability to link records with other large registries (e.g., death records, cancer databases). Unlike EHRs, claims only offer limited information on clinical severity and patients' health status, without laboratory, imaging, and other diagnostic test results. Furthermore, claims records do not reflect treatments and assessments that were suggested by clinicians and refused by patients. Claims data have the advantage, however, of collecting data from various sites that may not be included in a single EHR and result in a nationally representative sample. They may help to identify and reduce common biases in healthcare, e.g., when combined with other clinical data, they can help determine which conditions were undiagnosed in some patients, and at what point in time, so that future ML algorithms can detect early markers and indicators of future disease. Potential disadvantages to claims data include differences in values between billed and paid claims, confidentiality issues, and negative consequences like premiums based on personal traits potentially affecting insurability.

Novel Assessments (Sensors, Handwriting, Speech) (Table 1 section E)

Novel features like sensor (digital) data, handwriting (text), and speech (audio), offer unique opportunities to identify new indicators of cognitive decline (Kourtis et al., 2019). The five exemplar studies in Table 1 include home-based motion sensors (Akl et al., 2015), computerized handwriting analyses (Angelillo et al., 2019), videotaped handwashing tasks (Ashraf and Taati, 2016), multi-modal wearable activity monitors (Gwak et al., 2018), and audio-recorded speech data (Vincze, 2018) for detecting cognitive impairment. These data (particularly environmental and wearable sensors) have the potential for continuous, longitudinal tracking of cognitive changes. For example, Akl and colleagues (Akl et al., 2015) installed passive infrared motion sensors in participants' homes to assess movements and general activity by location that may be indicative of MCI over a 3-year period. They found that novel features like the trajectories of weekly walking speed were among the most important for detecting MCI in older adults. However, current relationships of these novel data with cognitive status are not yet well characterized. Furthermore, the sensor data contain artifacts (e.g., visitors, noise) and have considerable heterogeneity across individuals and environments (e.g., one- versus two-faucet handles, microphone position). Nonetheless, sensors offer an opportunity for tracking real-world behaviors in more ecologically valid

environments than traditional laboratory or clinic settings. Longitudinal sensor data are particularly difficult to visualize, understand, and manage without specialized algorithms provided by ML.

Genomic and Other Omics Data (Table 1 section F)

Genomic data are probably the best example of the big data ideally suited to ML analytic techniques. DL, in particular, is most useful when large amounts of data are available, and the human genome comprises more than 3 billion base pairs with a multitude of complex processes governing the expression of different genes (Libbrecht and Noble, 2017). Despite major advances in genomics, we still do not fully understand the various genes' functions and how they impact our physiology and health. Gene expression data can be used to learn to distinguish among different disease phenotypes and identify potentially valuable disease biomarkers. Alzheimer's Disease (AD), for example, is partially heritable and genetically complex. Large genome databases can offer enough training data to build accurate prediction models relating to gene expression, genomic regulation, or variant interpretation associated with AD and other cognitive impairments. The three studies highlighted in this section include a study of specific genes from a large NIH database (Jamal et al., 2016), gut microbiome analyses (Haran et al., 2019), and single-nucleotide polymorphism (SNP) data integrated with brain imaging (Zhou et al., 2018). The field of genomics is central to the precision medicine movement, as the illnesses an individual may experience are determined to a variable extent by their genes. ML has also enabled direct-to-consumer applications of genomic analyses like "23andMe" and "Ancestry.com." ML approaches have been leveraged to annotate a variety of genomic sequence elements (e.g., splice sites, promotors, enhancers), differentiate among different disease phenotypes, identify disease biomarkers, and investigate mechanisms underlying gene expression. However, genome-wide association studies (GWAS) for polygenic diseases like AD require extremely large sample sizes, which may limit the depth of phenotypic data and thus reduce the accuracy of these algorithms (e.g., 80% for Jamal and colleagues (Jamal et al., 2016); 65% for Zhou and colleagues (Zhou et al., 2018)).

IV. Discussion

High-dimensional data for AI

Different feature types for helping to detect, classify, and predict early pathological cognitive decline in older adults have varied strengths and limitations. The best-performing AI algorithms will require multi-feature data (Jiang et al., 2017) to personalize the findings to the level of the individual patient with their unique bio-psycho-social makeup (Havelka et al., 2009). For example, models based on only EHR data are likely to be biased due to the lack of important information about everyday functioning (e.g., physical function, social connections) that is also critical for health aging (Jeste et al., 2019). Based on this small subsample of studies, a wide variety of features (sociodemographic and clinical factors, specific cognitive tests, functional impairments, mobility problems, speech patterns, EEG measures, MRI-derived brain structures, PET and SPECT scan findings, and genes) were found to be associated with or predictive of cognitive impairment. To improve diagnosis and prognosis for adults with cognitive decline, AI research will require large, comprehensive,

multi-feature datasets that are collected longitudinally to better predict cognitive trajectories over time (Chi et al., 2017).

Developing such datasets entails several inherent challenges. Ongoing efforts to continually curate large-scale datasets like the ADNI and the UK Biobank databases will be key to the clinical success of AI, though they are costly and labor-intensive. Some claims and EHR companies are currently in search of feasible and legal ways to link these data with health risk assessments, sociodemographic data, and vital signs on a broad basis to create a more holistic picture of patients' health (Freij et al., n.d.). Furthermore, large-scale availability of novel features may be limited by proven clinical utility. For example, while neuroimaging or biosensor data can provide rich, multi-feature input for an AI algorithm, such data would not be available without broad insurance coverage and access to laboratory facilities (Crown, 2015).

Future Directions for AI and Neurocognitive Research

AI's strength lies in its ability to accommodate large quantities of multimodal data. Thus, AI can aid better understanding of unique factors and behaviors associated with cognitive decline that have been previously difficult to quantify, e.g., loneliness or social isolation (Biddle et al., 2019; Linggonegoro and Torous, 2019), resilience and wisdom (Meeks and Jeste, 2009), and behavioral symptoms like agitation and psychosis (Cheng, 2017; Feast et al., 2016). Capturing these factors and behaviors will require leveraging technology and novel inputs like mobile devices and sensor signals that are continually increasing in popularity and place low burden on the healthcare system (Kourtis et al., 2019).

The temptation may be to include the "kitchen sink" when developing a ML model because these algorithms enable a much larger set of predictor variables than commonly used in clinical research. However, features should still be evaluated for their validity in terms of potential relationships to the outcome of interest. It is also possible to create increasingly precise algorithms with additional features or continually fine-tuning the ML algorithm – though this may raise the likelihood of overfitting the model such that the algorithm is too customized for the particular training data and would not transfer well to another sample (Park and Han, 2018).

ML methods are subject to the same challenges and sources of bias encountered in observational data analyses using traditional statistical approaches. While small and labeled datasets for specific tasks are easier to collect, the resultant algorithms may not transfer to other datasets. In contrast, large and unlabeled datasets are also fairly easy to collect, but require a shift toward semi-supervised or unsupervised learning techniques that are harder to train (Esteva et al., 2019). Implementation of standards for AI/ML studies will be key to ensuring study quality. The US Food and Drug Administration (FDA) recently released a white paper (Administration, 2019) soliciting advice (by June 3rd, 2019) from stakeholders to help developers bring AI devices to market. The considerations discussed therein pertain to transparency, interpretability, and replication as components of "good ML practices". The World Economic Forum has also recommended a governance structure, safety and efficacy regulations, and responsible practices in the development of technological tools (World Economic Forum, 2019). Governmental regulation may be essential to establish regulatory

guidelines for AI applications in research like those endorsed by the EQUATOR network (EQUATOR Network, n.d.). The Computing Community Consortium also recently published a 20-year community roadmap for AI research (Gil and Selman, 2019), citing integrated intelligence (e.g., creating open-shared repositories of machine-understandable world knowledge); meaningful interaction (e.g., techniques for productive collaboration in mixed teams of humans and machines); and self-aware learning (e.g., developing causal and steerable models from numerical data and observations) as research priorities to realize societal benefits.

All of the studies presented in this overview focused on diagnosis or prediction of a neurocognitive disease. Algorithms to detect neurocognitive impairments may be able to support the decision-making capabilities of an experienced clinician, but they will not replace clinical expertise. No studies to date have directly compared clinical diagnostic accuracy of a neurocognitive disorder head-to-head with an AI approach, so the efficacy of these algorithms remains to be determined, with a few exceptions (Brinker et al., 2019; Lindsey et al., 2018; Nam et al., 2019). An accurate prediction of a patient diagnosis also does not provide clinicians with direction to change that outcome. However, AI could potentially expedite patient diagnoses if it can flag patients that are in need of immediate care or follow-up (Savage, 2019). If AI could further supplement clinical knowledge with less common datastreams, it may lend considerable support to individualizing prognoses and treatment decisions. Clinicians will require background knowledge regarding AI to decipher results and gauge the utility of such information (for an excellent guide applied to radiology, see (Park and Han, 2018)). Collaboration between clinicians and AI experts will be key to continual development of AI models, as clinicians can share their deep understanding of clinical populations – and the heterogeneity among individuals and over time – that will aid AI researchers in refining AI algorithms and transferring them to other populations.

Ethics of using AI for Neurocognitive Disorders

The ethical and social implications of using AI for detection and prediction of neurocognitive disorders include the need to weigh benefits against potential risks to patients. The benefits could be better healthcare; however, it is important to consider bias and accountability (Challen et al., 2019). For example, a risk may stem from whether the algorithm was built upon data that are not representative of the patient in question (e.g., older adults from underrepresented minorities), and subsequently presents a diagnosis that is questionable. Moving forward, there will need to be procedures to account for, and take action to mitigate, potential bias to avoid exacerbating inequities. AI models must be deployed in diverse samples to ensure generalizability. Moreover, how a decision is derived by the algorithm needs to be transparent to the clinician (Samek et al., 2017) so that a questionable recommendation can be examined before action is taken.

Within the context of diagnosing and predicting the trajectory of dementia, there are many disease-specific concerns. Once an individual is diagnosed with dementia, there can be serious legal and financial consequences, including the ability to make decisions, live independently, and even drive motor vehicles (Cornett and Hall, 2008). Algorithms can increasingly be applied to smartphones and other products that are widely distributed, based

on inputs such as keyboard typing patterns (White et al., 2018). While highly scalable, data ownership and privacy issues are a concern especially since regulations to protect user privacy are lacking, which may expose more people to surreptitious cognitive health surveillance. For example, passive surveillance tools applied to smartphone usage or social media posts could negatively impact ones job security, driving license, and insurance premiums (Rosenfeld and Torous, 2017). With such high stakes, the medical community must follow evidence-based practices to diagnose and treat their patients and pharmacotherapies must undergo rigorous clinical trials prior to approval by the FDA. Similarly, AI-derived algorithms must meet clinical standards. However, the threshold of proof and utility of AI models is not yet established.

Adopting AI algorithms in clinical practice carries the additional challenge of establishing trust in the model. The “black box” of ML presents a unique problem in how we reconcile the AI model’s results with our clinical experience and the scientific literature. The movement to develop Explainable AI (XAI) may aid the ability of clinicians to communicate these findings with other clinicians as well as with patients and their families to guide clinical decision-making (Gunning, 2017). XAI involves efforts to address a machine’s ability to explain its decisions and actions to users. The goal is explainable models that still have a high level of performance. Ultimately, healthcare liability remains with the clinician; thus, AI tools need to best support clinicians.

Limitations of this Review

Caution is necessary when generalizing the results of the studies presented in this paper, as they are not exhaustive, and therefore, not representative of the entire body of literature on AI and neurocognitive disorders. Due to the use of multiple definitions of MCI, the *a priori* labeling of MCI versus dementia groups may not reflect the longitudinal outcomes. There are potentially more recent exemplar studies within these feature categories that we did not capture. We also have not summarized these studies in any quantitative manner, as our goal was to highlight the breadth and range of studies that use AI methods to examine features of datasets relevant to neurocognitive disorders. This research is in too early a stage and consists of too much heterogeneity in methods to enable meaningful systematic analysis.

Conclusions

AI technology holds remarkable promise for transforming the way we diagnose and treat patients with neurocognitive disorders. There exist a large variety of potential features that in combination can comprehensively characterize the bio-psycho-social determinants of a unique individual and thus enable more personalized understanding of cognitive decline. The performance and potential clinical utility of ML algorithms for detecting, diagnosing, and predicting cognitive decline using these features will continue to improve as we leverage multi-feature and large datasets. Establishing guidelines for research involving AI applications in healthcare will be necessary to ensure the quality of results, as will engagement of clinicians (as well as patients and their caregivers) so that they may contribute their expertise in the refinement of AI algorithms. With the assistance of AI, early detection of cognitive decline may not be as difficult as it is today.

Acknowledgments

This study was supported, in part, by the National Institute of Mental Health T32 Geriatric Mental Health Program (grant MH019934 to DVJ [PI]), the IBM Research AI through the AI Horizons Network IBM-UCSD AI for Healthy Living (AIHL) Center, by the by the Stein Institute for Research on Aging at the University of California San Diego, and by the National Institutes of Health, Grant UL1TR001442 of CTSA funding. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

Administration, U.F. and D., 2019 Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based Software as a Medical Device (SaMD)-discussion paper and request for feedback.

Ahmed R, Zhang Y, Member S, Feng Z, Lo B, Member S, Inan OT, Member S, 2019 Neuroimaging and Machine Learning for Dementia Diagnosis : Recent Advancements and Future Prospects. IEEE Rev. Biomed. Eng 12, 19–33. 10.1109/RBME.2018.2886237 [PubMed: 30561351]

Akl A, Taati B, Mihailidis A, 2015 Autonomous Unobtrusive Detection of Mild Cognitive Impairment in Older Adults. IEEE Trans Biomed Eng. 62, 1383–1394. 10.1109/TBME.2015.2389149 [PubMed: 25585407]

Angelillo MT, Balducci F, Impedovo D, Pirlo G, Vessio G, 2019 Attentional Pattern Classification for Automatic Dementia Detection. IEEE Access 7, 57706–57716. 10.1109/ACCESS.2019.2913685

Ashraf A, Taati B, 2016 Automated Video Analysis of Handwashing Behavior as a Potential Marker of Cognitive Health in Older Adults. IEEE J. Biomed. Heal. Informatics 20, 682–690. 10.1109/JBHI.2015.2413358

Association, A.P., 2013 Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5, 5th ed. American Psychiatric Association, Arlington.

Balota DA, Tse C-S, Hutchison KA, Spieler DH, Duchek JM, Morris JC, 2010 Predicting Conversion to Dementia of the Alzheimer Type in a Healthy Control Sample: The Power of Errors in Stroop Color Naming. Psychol Aging 25, 208–218. 10.1038/jid.2014.371 [PubMed: 20230140]

Beaton K, Mcevoy C, Grimmer K, 2015 Identifying indicators of early functional decline in community-dwelling older people: A review. Geriatr. Gerontol. Int 15, 133–140. 10.1111/ggi.12379 [PubMed: 25303103]

Biddle KD, Uquillas O, Jacobs HIL, Ph D, Zide B, Kirn DR, Rentz DM, Psy D, Johnson KA, Sperling RA, Donovan NJ, 2019 Social Engagement and Amyloid- b -Related Cognitive Decline in Cognitively Normal Older Adults. Am. J. Geriatr. Psychiatry 27, 1247–1256. 10.1016/j.jagp.2019.05.005 [PubMed: 31248770]

Blagus R, Lusa L, 2015 Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models. BMC Bioinformatics 16, 1–10. 10.1186/s12859-015-0784-9 [PubMed: 25591917]

Brinker TJ, Hekler A, Hauschild A, Berking C, Schilling B, Enk AH, Haferkamp S, Karoglan A, von Kalle C, Weichenthal M, Sattler E, Schadendorf D, Gaiser MR, Klode J, Utikal JS, 2019 Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. Eur. J. Cancer 111, 30–37. 10.1016/j.ejca.2018.12.016 [PubMed: 30802784]

Brodaty H, Aerts L, Crawford JD, Heffernan M, Kochan NA, Reppermund S, Kang K, Maston K, Draper B, Trollor JN, Sachdev PS, 2017 Operationalizing the Diagnostic Criteria for Mild Cognitive Impairment: The Salience of Objective Measures in Predicting Incident Dementia. Am. J. Geriatr. Psychiatry 25, 485–497. 10.1016/J.JAGP.2016.12.012 [PubMed: 28110876]

Bzdok D, Krzywinski M, Altman N, 2018 Machine learning: supervised methods. Nat. Methods 15, 5–6. 10.1007/s10741-014-9462-7.Natural [PubMed: 30100821]

Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K, 2019 Artificial intelligence, bias and clinical safety. BMJ Qual. Saf 28, 231–237. 10.1136/bmjqs-2018-008370

Chen M, Hao Y, Hwang K, Wang L, Access, L.W.-I., 2017, U., 2017 Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5, 8869–8879.

Cheng S, 2017 Dementia Caregiver Burden : a Research Update and Critical Analysis. 10.1007/s11920-017-0818-2

Chi CL, Zeng W, Oh W, Borson S, Lenskaia T, Shen X, Tonellato PJ, 2017 Personalized long-term prediction of cognitive function: Using sequential assessments to improve model performance. J. Biomed. Inform 76, 78–86. 10.1016/j.jbi.2017.11.002 [PubMed: 29129622]

Cornett PF, Hall JR, 2008 Issues in disclosing a diagnosis of dementia. Arch. Clin. Neuropsychol 23, 251–256. 10.1016/j.acn.2008.01.001 [PubMed: 18299184]

Crown WH, 2015 Potential Application of Machine Learning in Health Outcomes Research and Some Statistical Cautions. Value Heal. 18, 137–140. 10.1016/j.jval.2014.12.005

De Langavant LC, Bayen E, Yaffe K, 2018 Unsupervised machine learning to identify high likelihood of dementia in population-based surveys: Development and validation study. J. Med. Internet Res 20, e10493 10.2196/10493 [PubMed: 29986849]

Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE, Penke L, Rafnsson SB, Starr JM, 2009 Age-associated cognitive decline. Br. Med. Bull. 2009; 92, 135–152. 10.1093/bmb/ldp033

Demner-Fushman D, Chapman WW, McDonald CJ, 2009 What can natural language processing do for clinical decision support? J. Biomed. Inform 42, 760–772. 10.1016/j.jbi.2009.08.007 [PubMed: 19683066]

Depp CA, Jeste Dilip, V., 2006 Definitions and Predictors of Successful Aging: A Comprehensive Review of Larger Quantitative Studies. Am J Geriatr Psychiatry 14, 6–21. [PubMed: 16407577]

Der G, Allerhand M, Starr JM, Hofer SM, Deary IJ, 2010 Age-related changes in memory and fluid reasoning in a sample of healthy old people. Aging, Neuropsychol. Cogn 17, 55–70. 10.1080/13825580903009071

Dodge HH, Kadowaki T, Hayakawa T, Yamakawa M, Sekikawa A, Ueshima H, 2005 Cognitive impairment as a strong predictor of incident disability in specific ADL-IADL tasks among community-dwelling elders: The Azuchi study. Gerontologist 45, 222–230. 10.1093/geront/45.2.222 [PubMed: 15799987]

Eckert MA, 2011 Slowing down: Age-related neurobiological predictors of processing speed. Front. Neurosci 5, 1–13. 10.3389/fnins.2011.00025 [PubMed: 21390287]

Esteva A, Robicquet A, Ramsundar B, Kuleshov V, Depristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J, 2019 A guide to deep learning in healthcare. Nat. Med 25 10.1038/s41591-018-0316-z

Fabris F, de Magalhães JP, Freitas AA, 2017 A review of supervised machine learning applied to ageing research. Biogerontology 18, 171–188. 10.1007/s10522-017-9683-y [PubMed: 28265788]

Fan M, Yang AC, Fuh J, Chou C, 2018 Topological Pattern Recognition of Severe Alzheimer ' s Disease via Regularized Supervised Learning of EEG Complexity. Front. Neurosci 12, 12 10.3389/fnins.2018.00685 [PubMed: 29410610]

Feast A, Moniz-Cook E, Stoner C, Charleworth G, Orrell M, 2016 A systematic review of the relationship between Behavioural and Psychological Symptoms (BPSD) and caregiver wellbeing. Int. Psychogeriatrics 28, 1761–1774.

Forum WE, 2019 Empowering 8 Billion Minds Enabling Better Mental Health for All via the Ethical Adoption of Technologies.

Freij M, Dullabh P, Lewis S, Smith SR, Dhopeshwarkar R, n.d. Incorporating Social Determinants of Health in Electronic Health Records : Qualitative Study of Current Practices Among Top Vendors Corresponding Author: 7, 1–12. 10.2196/13849

Gamberger D, Lavrač N, Srivatsa S, Tanzi RE, Doraiswamy PM, 2017 Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. Sci. Rep 7, 6763 10.1038/s41598-017-06624-y [PubMed: 28755001]

Gil Y, Selman B, 2019 A 20-Year Community Roadmap for Artificial Intelligence Research in the US.

Graham S, Depp C, Lee E, Nebeker C, Tu X, Kim H, Jeste D, 2019 Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. Curr. Psychiatry Rep.

Graham SA, Depp CA, 2019 Artificial intelligence and risk prediction in geriatric mental health: what happens next? Int. Psychogeriatrics 31, 921–923. 10.1017/S1041610219000954

Grassi M, Loewenstein DA, Caldirola D, Schruers K, Duara R, Perna G, 2018 A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion: Further evidence

of its accuracy via a transfer learning approach. Int. Psychogeriatrics 10.1017/S1041610218001618

Gross AL, Rebok GW, Unverzagt FW, Willis SL, Brandt J, 2011 Cognitive Predictors of Everyday Functioning in Older Adults: Results From the ACTIVE Cognitive Intervention Trial 66, 557–566. 10.1093/geronb/gbr033.

Gunning D, 2017 Explainable Artificial Intelligence (XAI), Defense Advanced Research Projects Agency (DARPA). 10.1111/fct.12208

Gwak M, Sarrafzadeh M, Woo E, 2018 Support for a Clinical Diagnosis of Mild Cognitive Impairment Using Photoplethysmography and Gait Sensors, in: Proceedings, APSIPA Annual Summit and Conference 2018 pp. 671–678.

Haran JP, Bhattarai SK, Foley SE, Dutta P, Ward DV, Bucci V, McCormik BA, 2019 Alzheimer's Disease Microbiome Is Associated with Dysregulation of the Anti-Inflammatory P-Glycoprotein Pathway. MBio 10, e00632–19. 10.1128/mBio.00632-19 [PubMed: 31064831]

Havelka M, Despot J, Lu D, 2009 Biopsychosocial Model – The Integrated Approach to Health and Disease. Coll. Antropol 33, 303–310. [PubMed: 19408642]

Hayrinen K, Saranto K, Nykanen P, 2008 Definition, structure, content, use and impacts of electronic health records: A review of the research literature. Int. J. Med. Inform 77, 291–304. 10.1016/j.ijmedinf.2007.09.001 [PubMed: 17951106]

Hirschberg J, Manning CD, 2015 Advances in natural language processing. Sci. Mag 349, 261–266.

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL, 2018 Artificial intelligence in radiology. Nat. Rev. Cancer 18, 500–510. 10.1038/S41568-018-0016-5 [PubMed: 29777175]

Hossin M, Sulaiman MN, 2015 A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process 5, 1–11.

Huang J, Ling CX, 2005 Using AUC and Accuracy in Evaluating Learning Algorithms. IEEE Trans. Knowl. Data Eng 17, 299–310.

Id VSN, Hane CA, Martin DC, Kravetz AD, Sanghavi M, 2019 Identifying incident dementia by applying machine learning to a very large administrative claims dataset 1–15. 10.1371/journal.pone.0203246

Iizuka T, Fukasawa M, Kameyama M, 2019 Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. Sci. Rep 9, 8944 10.1038/s41598-019-45415-5 [PubMed: 31222138]

Jamal S, Goyal S, Shanker A, Grover A, 2016 Integrating network, sequence and functional features using machine learning approaches towards identification of novel Alzheimer genes. BMC Genomics 17, 807 10.1186/s12864-016-3108-1 [PubMed: 27756223]

Jeste DV, Glorioso D, Lee EE, Daly R, Graham S, Liu J, Paredes AM, Nebeker C, Tu X, Twamley EW, Van Patten R, Yamada Y, Depp C, Kim H-C, 2019 Study of Independent Living Residents of a Continuing Care Senior Housing Community: Sociodemographic and Clinical Associations of Cognitive, Physical, and Mental Health. Am. J. Geriatr. Psychiatry 10.1016/j.jagp.2019.04.002

Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Yilong, Dong Q, Shen H, Wang Yongjun, 2017 Artificial intelligence in healthcare: Past, present and future. Stroke Vasc. Neurol 2, 230–243. 10.1136/svn-2017-000101 [PubMed: 29507784]

Khandai AC, Aizenstein HJ, 2013 Recent advances in neuroimaging biomarkers in geriatric psychiatry. Curr. Psychiatry Rep 15 10.1007/s11920-013-0360-9

Kourtis LC, Regele OB, Wright JM, Jones GB, 2019 Digital biomarkers for Alzheimer ' s disease: the mobile / wearable devices opportunity. npj Digit. Med 1–9. 10.1038/S41746-019-0084-2 [PubMed: 31304351]

Libbrecht MW, Noble WS, 2017 Machine learning in genetics and genomics Maxwell. Nat Rev Genet. 16, 321–332. 10.1038/nrg3920.Machine

Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, 2018 Deep neural network improves fracture detection by clinicians. PNAS 115, 11591–11596. 10.1073/pnas.1806905115 [PubMed: 30348771]

Linggonegoro DW, Torous J, 2019 Expanding Technology for Engagement in Dementia While Ensuring Equity, Interoperability, and Privacy. Int. Psychogeriatrics

Lins AJCC, Muniz MTC, Garcia ANM, Gomes AV, Cabral RM, Bastos-Filho CJA, 2017 Using artificial neural networks to select the parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. Comput. Methods Programs Biomed 152, 93–104. 10.1016/j.cmpb.2017.09.013 [PubMed: 29054264]

Meeks TW, Jeste DV, 2009 Neurobiology of Wisdom. Arch Gen Psychiatry 66, 355–365. [PubMed: 19349305]

Michigan TR of the U. of, 2019 Health and Retirement Study. [WWW Document].

Miotto R, Li L, Kidd BA, Dudley JT, 2016 Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci. Rep 6, 1–10. 10.1038/srep26094 [PubMed: 28442746]

Miotto R, Wang F, Wang S, Jiang X, Dudley JT, 2017 Deep learning for healthcare: Review, opportunities and challenges. Brief. Bioinform 19, 1236–1246. 10.1093/bib/bbx044

Mitchell A, Shiri-Feshki M, 2009 Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009 119, 252–265. 10.1111/j.1600-0447.2008.01326.x

Mitchell AJ, Shiri-Feshki M, 2008 Temporal trends in the long term risk of progression of mild cognitive impairment: A pooled analysis. J. Neurol. Neurosurg. Psychiatry 79, 1386–1391. 10.1136/jnnp.2007.142679 [PubMed: 19010949]

Moreira LB, Namen AA, 2018 A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. Comput. Methods Programs Biomed 165, 139–149. 10.1016/j.cmpb.2018.08.016 [PubMed: 30337069]

Na KS, 2019 Prediction of future cognitive impairment among the community elderly: A machine-learning based approach. Sci. Rep 9, 1–9. 10.1038/s41598-019-39478-7 [PubMed: 30626917]

Nam JG, Park S, Hwang EJ, Lee JH, 2019 Development and Validation of Deep Learning – based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology 290, 218–228. [PubMed: 30251934]

Network, E., n.d. Enhancing the QUAlity and Transparency Of health Research [WWW Document].

Nevin L, 2018 Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. PLoS Med. 15, 4–7. 10.1371/journal.pmed.1002708

Organization, W.H., 2019 What is healthy ageing? [WWW Document].

Park SH, Han K, 2018 Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis. Radiology 286, 800–809. [PubMed: 29309734]

Van Patten R, Fagan AM, Kaufman DAS, 2018 Differential Cued-Stroop Performance in Cognitively Asymptomatic Older Adults with Biomarker-Identified Risk for Alzheimer's Disease: A Pilot Study. Curr. Alzheimer Res 15, 820–827. 10.2174/1567205015666180404170359 [PubMed: 29623843]

Petersen R, 2011 Mild Cognitive Impairment. N. Engl. J. Med 364, 2227–2234. [PubMed: 21651394]

Petersen RC, Stevens JC, Ganguli M, Tangalos EG, 2001 Practice parameter : Early detection of dementia : Mild cognitive impairment (an evidence-based review) Report of the Quality Standards Subcommittee of the 1133–1142.

Raghupathi W, Raghupathi V, 2014 Big data analytics in healthcare : promise and potential 1–10.

Rosenfeld L, Torous J, 2017 Data Security and Privacy in Appsfor Dementia: An Analysis of Existing Privacy Policies. Am. J. Geriatr. Psychiatry 25, 873–877. 10.1016/J.JAGP.2017.04.009 [PubMed: 28645535]

Samek W, Wiegand T, Müller K-R, 2017 Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. arXiv Prepr. arXiv 1708.08296.

Savage N, 2019 Artificial-intelligence technology could help radiologists and pathologists to diagnose disease. Nature 573, S98–S99. [PubMed: 31554994]

Senanayake U, Sowmya A, Dawes L, Kochan NA, Wen W, 2017 Deep Learning Approach for Classification of Mild Cognitive Impairment Subtypes 655–662. 10.5220/0006246306550662

Service., K.E.I., 2015 Korean Longitudinal Study of Ageing (KLoSA) [WWW Document].

Shao Y, Zeng QT, Chen KK, Shutes-david A, Thielke SM, Tsuang DW, 2019 Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records 4, 1–11.

SHARE-ERIC, 2019 SHARE: Survey of Health, Ageing and Retirement in Europe. [WWW Document].

Silverberg NB, Ryan LM, Carrillo MC, Sperling R, Petersen Ronald, C., Posner HB, Snyder PJ, Hilsabeck R, Gallagher M, Raber J, Rizzo A, Possin K, King J, Kaye J, Ott BR, Albert MS, Wagster MV, Schinka JA, Cullum CM, Farias ST, Balota D, Rao S, Loewenstein D, Budson AE, Brandt J, Manly JJ, Barnes L, Strutt A, Gollan TH, Ganguli M, Babcock D, Litvan I, Kramer JH, Ferman TJ, 2011 Assessment of cognition in early dementia. Alzheimers Dement. 7, e60–e76. 10.1016/j.jalz.2011.05.001.Assessment [PubMed: 23559893]

Vernooij MW, Pizzini FB, Schmidt R, Smits M, Yousry TA, Bargallo N, Frisoni GB, Haller S, Barkhof F, 2019 Dementia imaging in clinical practice: a European-wide survey of 193 centres and conclusions by the ESNR working group. Neuroradiology 633–642. 10.1007/S00234-019-02188-y [PubMed: 30852630]

Vincze V, 2018 A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech 2050, 130–138. 10.2174/1567205014666171121114930

Wang L, Lakin J, Riley C, Korach Z, Frain LN, Zhou L, 2018 Disease Trajectories and End-of-Life Care for Dementias : Latent Topic Modeling and Trend Analysis Using Clinical Notes, in: AMIA Annual Symposium Proceedings, pp. 1056–1065.

Wang L, Sha L, Lakin JR, Bynum J, Bates DW, Hong P, Zhou L, 2019 Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions 2, 1–14. 10.1001/jamanetworkopen.2019.6972

Wang Y, Kung LA, Byrd TA, 2016 Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technol. Forecast. Soc. Change 126, 3–13. 10.1016/j.techfore.2015.12.019

White RW, Doraiswamy PM, Horvitz E, 2018 Detecting neurodegenerative disorders from web search signals. npj Digit. Med 1, 18–21. 10.1038/S41746-018-0016-6 [PubMed: 31304302]

Willis SL, Tennstedt SL, Marsiske M, Ball K, Elias J, Koepke KM, Morris JN, Rebok GW, Unverzagt FW, Stoddard AM, Wright E, 2006 Long-term effects of cognitive training on everyday functional outcomes in older adults. J. Am. Med. Assoc 296, 2805–2814. 10.1001/jama.296.23.2805

World Health Organization, 2019 Dementia [WWW Document].

Yu K, Beam AL, Kohane IS, 2018 Artificial intelligence in healthcare. Nat. Biomed. Eng 2, 719–731. 10.1038/s41551-018-0305-z [PubMed: 31015651]

Zhavoronkov A, Mamoshina P, Vanhaelen Q, Scheibye-Knudsen M, Moskalev A, Aliper A, 2019 Artificial intelligence for aging and longevity research: Recent advances and perspectives. Ageing Res. Rev 49, 49–66. 10.1016/J.ARR.2018.11.003 [PubMed: 30472217]

Zhou T, Thung K, Zhu X, Shen D, 2018 Feature Learning and Fusion of Multimodality Neuroimaging and Genetic Data for Multi-status Dementia Diagnosis 132–140. 10.1007/978-3-319-67389-9

Highlights

- AI holds remarkable promise for transforming the way we diagnose and treat patients with neurocognitive disorders.
- Machine learning can leverage large volumes of longitudinal, high-dimensional data like imaging and genomics that may best predict cognitive trajectories.
- Regulations pertaining to interpretability, replication, ethics, and safety are needed to establish best practices for AI research.

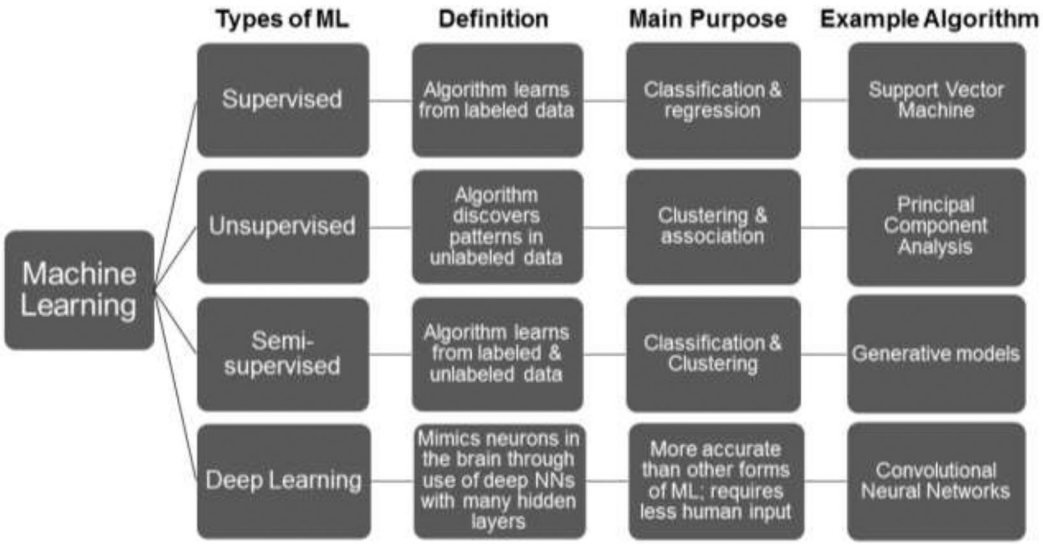

Figure 1a.

The most common subcategories of machine learning (ML) used for healthcare purposes. NN=neural network

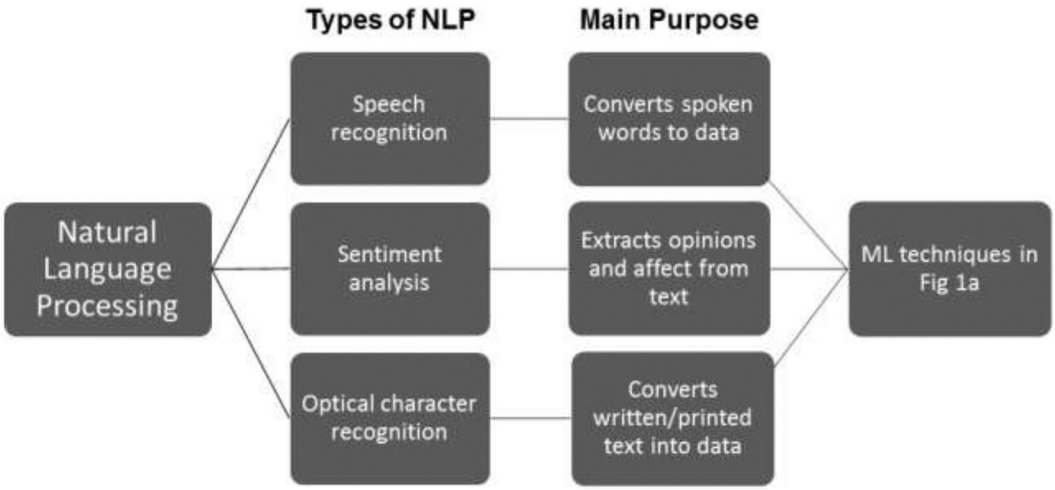

Figure 1b.

The most common subcategories of natural language processing (NLP) used for healthcare purposes.

Table 1.

Summary of Characteristics of Selected Studies of AI for Cognitive Impairment

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
Sociodemographic Data (Section A)									
Langavant et al., 2018 [41] *Journal of medical internet research* University of Paris, Créteil, France	Identify participants with high likelihood of dementia in population-based surveys without clinical diagnosis	n=18,165 training (59% F) n=58,202 test (57% F) Training: US Adults >50 years from Health & Retirement Study (HRS; 2002-2003) (n=856 received in-home assessment of dementia using clinical criteria) Test: European adults >50 years from SHARE; 2010-2012	Survey-based data including demographics, health, health care utilization, & cognition	UL		X	X	**Algorithm:** Hierarchical clustering: identified 3 clusters based on functional & motor (walking, climbing) limitations **Performance:** Cluster 3 (high risk for dementia) accuracy=93.1% AUC=0.91 **Main findings:** UL identified high likelihood of dementia in population-based surveys, even without cognitive & behavioral measures & without clinical diagnosis of dementia	**Strengths:** -More generalizable to other samples due to being commonly collected data -Larger sample sizes due to public registries -Inclusive of all demographic groups -Contains social determinants of health -Beneficial for resource poor areas with limited primary care access & limited cognitive testing capacities. **Limitations:** -Lack clinical/ biological information that may allow for more precise diagnoses
Na 2019 [42] *Scientific reports* Gachon University College of Medicine, Incheon, Republic of Korea	Predict cognitive impairment using variables commonly collected in community health care institutions	N=3,424 community-dwelling older adults 70.4 ± 7.0 without cognitive impairment based on MMSE (53.7% F) Data from KLoSA 2014 to 2016	Sociodemographic, health, functional, & subjective well being	SL	X			**Algorithm:** GBM **Performance:** AUC = 0.921 **Main findings:** Cognitive decline best predicted by: age, MMSE, & education.	
Clinical and Psychometric Assessments (Section B)									
Lins et al., 2017 [43] *Computer Methods and Programs in Biomedicine* Federal Rural University of Pernambuco, Northeast Brazil	Predict MCI & dementia from demographic & standard neurocognitive test features	N=151 (25% held out as test set); n=70 adults with clinical diagnosis of MCI 71.3 ±7.5 yrs; n=56 adults with dementia 76.9 ± 7.5 yrs; 25 HCs 69.1 ± 5.1 yrs Database from Molecular Markers in Degenerative Diseases	Gender, age, level of education, study time, & scores from cognitive tests (MMSE, Semantic Verbal Fluency Test, CDR, & Ascertaining Dementia).	SL	X	X		**Algorithm:** RF: only cognitive tests **Performance:** Accuracy=96.8%, sensitivity=0.98, specificity=0.96 **Main findings:** Using only cognitive testing (MMSE, CDR, AD8) was best for predicting cognitive status.	**Strengths:** -Direct assessment of cognitive functioning -Standard scoring metrics and comparison to validated norms -High relevance to clinicians **Limitations:** -Certain assessments require clinical suspicion and more resources to obtain (e.g., neuropsychological testing)

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
Moreira & Namen 2018 [44] *Computer Methods and Programs in Biomedicine* North Fluminense State University, Rio de Janeiro, Brazil.	Determine whether unstructured mining of medical texts improves diagnosis of MCI & AD	N=605; characterized in model as ≥65 or <65 years (gender not specified but also included in model) patients attending the Alzheimer & Parkinson Center in the city of Campos dos Goytacazes	Demographic, clinical, neuropsych screening tests, and clinical notes	NLP UL SL		X		**Algorithm:** Best model for AD: J48 with AdaBoost ensemble method including UL x-means clustering **Performance:** accuracy = 0.80 & AUC = 0.85 **Algorithm:** Best model for MCI: NB with Bagging including UL k-means or x-means clustering **Performance:** accuracy = 0.85; AUC = 0.87 **Main findings:** greater effectiveness of a hybrid (UL/SL) model for diagnosing AD and MCI; clinician notes contain important information that should not be ignored.	-Cognitive data are derived from contrived testing situations
Senanayake et al., 2017 [45] *ICPRAM* UNSW, Sydney, Australia	Distinguish between CN & MCI using neuropsychological test scores	N=1,037; 70-90 years (57% F) community-dwelling, non-demented adults (MMSE score ≥24) MCI (CDR=0.5 criteria) further divided into aMCI & naMCI Data from the Sydney Memory & Aging Study (MAS)	Between 28-35 neuropsychological test scores (depending on enrollment wave)	DL	X			**Algorithm:** SAE: CN vs. MCI **Performance:** Accuracy=83%; AUC=88% MCI subtypes Accuracy=76%; AUC=80% **Main findings:** Neuropsychological measures can differentiate between MCI & its subtypes. DL SAE has significant advantages over conventional classifiers; SAE can be used as an unsupervised feature extractor; model will further improve with higher dimensional data	
Neuroimaging and Neurophysiologic Data (Section C)									
Fan et al., 2018 [50] *Frontiers in Neuroscience* Northeastern University, Boston, MA	Discover the altered spatiotemporal patterns of EEG complexity associated with AD pathology in different severity levels	N=123 adults from the Dementia Clinic at the Neurological Institute, Taipei Veterans General	Multiscale Entropy of EEG, a complexity measure of time series signals	SL	X	X		**Algorithm:** LASSO: HC vs. AD3 **Performance:** accuracy=79.5% AD1 vs. AD3 accuracy = 82% AD2 vs. AD3 accuracy = 72.4%	**Strengths:** -Systematic approach to complex, multi-layered imaging data. (In other fields, AI techniques can detect imaging

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
		Hospital in Taiwan (AD diagnosed with NINCDS-ADRDA Criteria) HC N=15; AD1 (CDR=0.5) N=15; AD2 (CDR=1) N=69; & AD3 (CDR=2) N=24						**Main findings:** Temporal & occipitoparietal brain regions were more discriminative for classifying severe AD vs.NC, but more diverse & distributed patterns of EEG complexity in the brain were exhibited across individuals in early stages of AD	abnormalities on par with trained physicians) -Potential to improve interpretability and clinical utility of certain commonly-obtained but incompletely-understood imaging data -Hypothesis-generating potential for brain-based mechanisms -Can guide development of targeted therapies using neurostimulation approaches. **Limitations:** -Expensive (thus, smaller sample sizes, less generalizability) -Heterogeneity of datasets (imaging modalities, machines, processing approaches) that make it challenging to harmonize data -Less tightly correlated with real-world functional outcomes than clinical and neuropsychological data
Gamberger et al., 2017 [51] *Scientific Reports* Duke University Medical Center	Identify different prognostic cognitive trajectories of MCI patients through discovering homogenous clusters based on neuroimaging, clinical data, & cognitive tests	N=562; 74.0 ± 7.5 years Data from ADNI database: ADNI-1 & ADNI-2 late MCI subjects with at least one postbaseline visit (criteria available in ADNI procedures manual [http://www.adni-info.org/]) (39% F)	Clinical, cognitive, & biomarker (volumetric brain MRI, amyloid PET, FDG-PET, spinal fluid) tests	UL		X		**Algorithm:** Multi-layer clustering; two clusters identified—rapid vs. slow decliners. **Performance:** Best predictor: baseline ADAS13 > 19.50 yielded 92% sensitivity & 93.7% specificity in ADNI1 & 98.4% sensitivity & 90% specificity in ADNI2 **Main findings:** Pathological differences between rapid vs. slow decliners included an almost 5-fold greater rate of converting to dementia in rapid vs. slow cluster, & a lower rate of reverting back to cognitively normal (0% vs. 13%)	
Grassi et al., 2018 [52] *International Psycho geriatrics* Mount Sinai Medical Center, Miami Beach, Florida, & the Community & Memory Disorders Center at the University of South Florida	Prediction of 3-year conversion to AD in subjects with MCI & Pre-MCI from clinical & MRI data	n=75 older adults with DSM-criteria diagnosis of AD; age NR n=197 HC; age NR n=61 older adults meeting CDR criteria for MCI (out of sample) 70+ years (60% F)	clinical & neuro psychological testing, cardiovascular risk, rating of MRI data	SL	X		X	**Algorithm:** SVM **Performance:** AUC=0.996 for AD vs. HC **Results out of sample:** SVM: AUC = 0.821 MCI **Main findings:** Clinically available data can be used to predict 3-year conversion from MCI to AD	
Iizuka et al., 2019 [53] *Scientific Reports*	Diagnose DLB & AD from SPET scans	n=240 (80 each for DBL 77.9 ± 5.3 years, AD	SPECT images with emphasis on CIS	DL	X	X		**Algorithm:** CNN **Performance:** accuracy for	

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
Fukujuji Hospital, Japan		77.8 ± 5.42 years, & NL 77.7 ± 5.0 years) training; n=60 (20 each) for training DLB, AD, & NL (McKeith criteria & NINCDS-ADRDA) (52% F)						differentiating DLB-NL=93%; DBL-AD=89%; AD-NL=92% AUCs for differentiating DLB–NL=0.95; DLB–AD=0.94; AD-NL=0.94 **Main findings:** DL was useful for differentiating DLB from AD, & for predicting clinical features of DLB. CIS was more involved in discrimination of DLB–AD rather than DLB–NL	
EHR and Claims Data (Section D)									
Nori et al., 2019 [55] *Plos One* Optum Labs, Cambridge, MA,	Predict ADRD 4–5 years in advance from administrative claims data	N=44,945 with ADRD N=2,901,044 NC 77.2 ±7.0 yrs training data; (62% F training) (27% training; 73% test) ADRD diagnosis (medical claim codes) Data from 2001-2015 from the Optum Labs Data Warehouse (OLDW); all 50 states represented	over 10,000 clinical, pharmaceutical, and demographic variables	SL	X	X		**Algorithm:** LASSO & regularized logistic regression **Performance:** AUC 0.69 test data **Main findings:** Patients identified by the model 6.4 times more likely to be diagnosed with dementia in the near-term	**Strengths:** -Potential to detect at-risk patients seeking healthcare for reasons other than cognitive decline -Large and longitudinal datasets **Limitations:** -Quality and quantity of EHR data for individuals are dependent on external factors (severity of illness, insurance rules, psychosocial resources, regional practices and resources). For example, sicker patients will likely have more contact with the healthcare system and more documentation within the EHR. - EHR data may not reflect assessments or work-up that were recommended by providers but declined by the patient. -EHR data is heterogeneous in organization and level of detail on the provider, clinic, and system-levels, e.g., a geriatric specialty clinic
Shao et al., 2019 [56] *BMC Medical Informatics & Decision Making* VA Puget Sound	Identify cases of undiagnosed dementia from both structured & unstructured EHR data	n=1,861 Veterans with (79.8 yrs) & n=9,305 without (79.5 yrs) ICD-9 dementia codes (3.3% F) Data from the clinical data warehouse (CDW) within the Veterans Affairs Informatics & Computing Infrastructure (VINCI)	Structured data (diagnosis [ICD codes], procedures [CPT codes], medications, & clinical document types); unstructured data (clinical document text)	UL/SL		X		**Algorithm:** LDA **Performance:** 853 features identified (290 topics, 174 non-dementia ICD codes, 159 CPT codes, 59 medications, & 171 note types) **Main findings:** imperfect data (e.g., ICD codes in combination with other EHR features) can be used to detect Veterans with undiagnosed dementia	
Wang et al., 2019 [57]	Predict mortality from demographic	Patients with dementia	959,628 clinical notes,	DL NLP	X	X		**Algorithm:** LSTM 6-month mortality	

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
JAMA network open Harvard Medical School, Boston, Massachusetts	& clinical notes, highlight topics that best predict mortality to detect patients that may benefit from palliative services	Training: n=24,229 [60% F, 74.8 ± 13.2 years] Test: n=2,692 [61% F, 75.0 ±12.6 years] Data from Partners HealthCare System patients who visited from1/1/11 through 12/31/17	demographics, death status					**Performance:** AUC 0.978 test data 1-year mortality AUC 0.956 test data 2-year mortality AUC 0.943 test data **Main findings:** Top-ranked latent topics associated with 6-month, 1- & 2-year mortality included palliative & end-of-life care, cognitive function, delirium, testing of cholesterol levels, cancer, pain, use of health care services, arthritis, nutritional status, skin care, family meeting, shock, respiratory failure, & swallowing function	may order a different panel of tests and assessments compared to a primary care clinic.
Wang et al., 2018 [58] *AMIA Annual Symposium Proceedings* Harvard Medical School, Boston, MA	Evaluated topic models for important themes mentioned in care provider notes about dementia patients; explored patterns & trends of topics over the final 2 years of life	n=7,875; 84.3 ± 9.5 years at death with dementia (54.5% F) (432,007 clinical notes) n=133,394 HC 71.9 ± 16.5 years at death Patients with dementia from two PHS hospitals: Brigham & Women's Hospital & Faulkner Hospital	All types of inpatient & ambulatory notes--office visit notes, progress notes, discharge summaries, emergency department notes, consultations, nutrition notes, social work notes, phone calls	UL NLP				**Algorithm:** Topic modeling (LDA) **Performance:** generated 224 stable topics classified into 72 unique categories **Main findings:** Patterns & trends of identified topics provided unique findings & insights not documented in EHR; e.g., functional status, mental status, & palliative care.	
Novel Assessments (Speech, Handwriting, Sensors) (Section E)									
Akl et al., 2015 [61] *IEEE Trans Biomed Eng.* University of Toronto, Canada	Detect MCI using home-based sensing technology	N=97 older adults 80+ years from Portland, Oregon, metropolitan area living alone either CIN or MCI (CDR criteria) (90% F)	Motion detected with passive infrared motion sensors & walking speed	SL	X			**Algorithm:** SVM **Performance:** AUC=0.97 **Main findings:** Trajectories of weekly walking speed, CoV of walking speed, CoV of morning & evening walking speeds, age, & gender were most important for detecting MCI in older adults	**Strengths:** -Potential to discover new biomarkers and biological mechanisms of cognitive decline -Potential for monitoring in real-world settings -Continuous, longitudinal monitoring enables pattern identification **Limitations:**
Angelillo et al., 2019 [62]	Detect dementia automatically from	N=65 total n=29 HC age	Handwriting information: x & y	SL	X			**Algorithm:** Ensemble classifier	

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
IEEE Access University of Bari, Italy	results of a digital version of the attentional matrices test (AMT)	65±13 years n=36 with diagnosis of dementia 75±9 years (%F NR)	coordinates of pen position; pen inclination; pen pressure; pen airtime vs. contact time; horizontal & vertical Shannon entropy					**Performance:** accuracy=84%; AUC=87% **Main findings:** Digitalization of the AMT enables capturing a larger set of performance measures than can be obtained by the paper-based test; the best variable for screening for cognitive impairment was prolonged in-air time	-Less is known about relationships between novel measures and cognitive decline -Particularly if used in isolation from other measurements, may have lower accuracies due to exploratory nature of these data -Current research is exploratory and has smaller sample sizes -High heterogeneity across individuals and environments
Ashraf & Taati 2016 [63] *IEEE journal of biomedical & health informatics* University of Toronto, Canada	Predict cognitive status by monitoring hand-washing behaviors	N=27 participants; 82.4 ± 9.5 years; (81.4% F) with MMSE scores ranging from no to severe impairment	Video-tapes of hand-washing in one bathroom at a long-term care facility	SL	X			**Algorithm:** RF 4-class classifier (aware, mild, moderate, severe) **Performance:** all features accuracy=52.1%; collapsed features=70.4% **Main findings:** Computer-rated aspects of handwashing (occupancy of sink areas & hand motions) can predict MMSE scores & classifications	
Gwak et al., 2018 [64] *In Proceedings APSIPA Annual Summit & Conference* University of California Los Angeles	Classify MCI vs. CH using PPG & gait sensor data	N=69 older adults 72.5±10.6 years recruited for the longitudinal aging study from the Department of Neurology, Psychiatry, & Computer Science (51% F)	PPG & gait accelerometer & gyroscope sensor data	SL	X			**Algorithm:** RF & logistic regression **Performance:** RF accuracy=82% PPG data only; logistic regression accuracy = 86% **Main findings:** Classification accuracy using the optimal feature subset was higher than when only using a neuropsychological test score (CVLT) (76% & 79%)	
Tóth et al., 2018 [65] *Current Alzheimer Research* Memory ambulance of the University of Szeged, Hungary	Detect MCI based on acoustic features from spontaneous speech	n=48 adults with clinical diagnosis of MCI 73 years (55-93) & n=38 HC 64 years (57-84) (65.5% F)	Acoustic parameters from spontaneous speech recall of 2 short black & white films	NLP SL	X			**Algorithm:** SVM with manual feature selection **Performance:** accuracy = 71%, AUC=71% **Algorithm:** RF with automatic feature selection **Performance:** accuracy = 71%, AUC=70%	

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
								Main findings: Most significant differences between groups in speech tempo from delayed recall task, & number of pauses for question-answering task	
Genomic and other -Omic Data (Secttion F)									
Jamal et al., 2016 [67] *BMC Genomics* Jawaharlal Nehru University, New Delhi, India	Predict probable AD- associated genes from a large pool of genes & identify therapeutic targets	Entrez gene database at the National Centre for Biotechnology Information (NCBI) 458 genes which may cause AD; 55947 non-AD genes	56405 genes belonging to Homo sapiens species	SL	X			**Algorithm:** NB **Performance:** Accuracy = 80% **Main findings:** Identified 13 novel candidate genes that could have a potential role in AD pathology; demonstrated that AL-108, an investigational AD-specific drug, had strong binding affinity for all novel drug targets	**Strengths:** -Existence of large databases -Discover new roles of genes in the pathology of cognitive decline -Genes are purported to play a large role in neurodegenerative pathogenesis -Discover new drug targets for neurodegenerative diseases like AD **Limitations:** -Lack of access to biological samples -Not routinely collected -Often used in the absence of other clinical information (lowers accuracy) -Limited phenotypic data in many large genetic databases
Haran et al., 2019 [68] *mBio* University of Massachusetts Medical School, Worcester, Massachusetts	Identify numerous microbial taxa & functional genes that act as predictors of AD in comparison to elders without dementia or with other dementia types	N=108 nursing home residents (47.2% no dementia 83.0 ± 10.2 years, 22.2% AD 84.7±8.1, & 30.6% other dementia types 87.9±7.9 (CDR scores) (49% F)	Longitudinal stool samples for intestinal microbiota (P-glycoprotein expression)	UL				**Algorithm:** t-distributed stochastic neighbor embedding (tSNE) **Performance:** identified lower proportions of key butyrate-producing species in AD; Jaccard distances between AD samples more similar than those from individuals with no dementia or other dementia types **Main findings:** Microbiome of AD shows a lower proportion & prevalence of bacteria with the potential to synthesize butyrate, & higher abundances of taxa known to cause proinflammatory states	
Zhou et al., 2017 [69] *Mach Learn Med Imaging.* University of North Carolina, Chapel Hill	Predict AD & its prodromal status from multimodal imaging & genetic data	ADNI dataset 190 AD 75.2±7.5 years, 389 MCI 74.9±7.3 years, 26 HC 75.8±5.0 years (43% F)	MRI, PET (93 ROIs) & SNP (3023 features)	DL	X			**Algorithm:** DNN **Performance:** MRI +PET+SNP highest accuracy = 65% **Main findings:** The combination of brain imaging & genetic features	

Authors / Journal / Location	Primary Aim	Subjects / Dataset	Predictors (features) used by AI algorithm	AI method	Validation			Best algorithm and performance / Main finding(s)	Strengths and limitations of using these features with AI analytic approaches
				SL UL DL NLP	CV	In sample test	Out of sample test		
								produced the highest accuracy in classifying AD vs. MCI vs. HC	

AD=Alzheimer's Disease; ADNI= Alzheimer's Disease Neuroimaging Initiative; AI=artificial intelligence; AMT= attentional matrices test; AUC=area under the curve; CDR=clinical dementia rating; CN=control; CNN=convolutional neural network; CoV=coefficient of variation; CV=cross validation; DL=deep learning; DNN=deep neural network; F=female; GBM=gradient boosting model; HC=healthy control; KLoSA=Korean Longitudinal Study of Aging; LDA=latent Dirichlet allocation; LSTM=long short-term memory; MCI=mild cognitive impairment; MMSE=mini mental state examination; NB=naïve bayes; NLP=natural language processing; PPG=photoplethysmography; RF=random forest; ROI=region of interest; SAE=stacked auto-encoder; SHARE= Survey of Health, Ageing & Retirement; SL=supervised learning; SNP= Single Nucleotide Polymorphism; SVM=support vector machine; UL=unsupervised learning