Health and Technology (2024) 14:201–237
https://doi.org/10.1007/s12553-024-00823-0

REVIEW PAPER

Early detection of dementia using artificial intelligence and multimodal features with a focus on neuroimaging: A systematic literature review

Ovidijus Grigas[1] · **Rytis Maskeliunas**[1] · **Robertas Damaševičius**[2]

Received: 6 April 2023 / Accepted: 16 January 2024 / Published online: 10 February 2024

Abstract

Purpose This paper is a systematic literature review of the use of artificial intelligence techniques to detect early dementia. It focuses on multi-modal feature analysis in combination with neuroimaging. The paper examines what past research suggests about issues in the field, what dementia types researchers focus on, what are state-of-the-art methods in the different dementia detection groups, what combinations of modalities (images, text, speech, etc.) are frequently used, how models are evaluated and validated, what datasets researchers use, what are common pre-processing and feature extraction from neuroimages techniques, what are key issues in this research area, and what are potential future research areas.

Materials and methods The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method was used to collect and summarize research in the scope of the defined problem. This study investigated early dementia detection problem from a multi-modal perspective, with neuroimaging being used as one of the modalities.

Results Five databases were queried and 2881 sources were identified and processed in the literature review. 59 sources were selected after eligibility assessment. The study identified all points defined in the purpose of the research.

Conclusions The main findings of the study were that Alzheimer's disease and Mild Cognitive Impairment (MCI) are the most researched dementia types in the field; typical choice for dementia detection is Machine Learning (ML) methods; the most popular modalities combination is T1w + Fluorodeoxyglucose - Positron Emission Tomography (FDG-PET); accuracy, sensitivity and specificity are the main evaluation metrics used by the researchers; k-fold validation is being used the most; Alzheimer's disease neuroimaging initiative (ADNI) is the most used dataset by researchers; intensity and spacial normalization, skull stripping and segmentation are the most common pre-processing techniques for neuroimages; voxel average intensities are being used the most as features in classification extracted from neuroimages; explainability still persists as one of the main issues in adoption of developed methods in clinical practise; there is a lack of studies on Vascular dementia, Frontotemporal dementia, Parkinson's disease and Huntington's disease.

Keywords Multi-modal data · Early dementia · Artificial intelligence · Machine learning · Deep learning · Neuroimaging · Systematic review

1 Introduction

Dementia is frequently described as a condition which progressively degrades a person's cognitive abilities and should not be mistaken by being a composing part of aging. World Health Organization [1] describes dementia as "Dementia is a syndrome in which there is deterioration in cognitive function beyond what might be expected from the usual consequences of biological ageing." and National Institute of Aging [2] describes it as "Dementia is the loss of cognitive functioning - thinking, remembering, and reasoning - to such an extent that it interferes with a person's daily life and activities", however US Department of Health and Human Services [3] describes it as "Dementia is an umbrella term used to describe a range of neurological conditions affecting the brain that gets worse over time". Following these descriptions, it is easy to derive the meaning of dementia - a group of neurological diseases, which is associated

✉ Robertas Damaševičius
robertas.damasevicius@vdu.lt

[1] Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania

[2] Department of Applied Informatics, Vytautas Magnus University, Kaunas, Lithuania

with degradation of cognitive abilities and brain function. This disorder, particularly in later stages, affects a person's abilities to function as an individual and forces him/her to count on family or friend assistance. In later stages dementia is usually categorized as one of the diseases, where the most frequent is Alzheimer's disease followed by vascular dementia and dementia with Lewy bodies as well as Parkinson's disease.

Dementia has three stages (early, middle, late) and the first stage is the hardest to detect due to very mild effects on the cognitive abilities that can be misidentified as lack of rest or sleep. Therefore, detecting early dementia is still a challenging task for researchers. Early diagnosis of dementia is necessary for people to prepare for a (presently) inevitable future as there is no cure, however with existing symptomatic treatment patient can lengthen the period of time in which he can be independent or in rare cases reverse cognitive decline effects.

Typical ways to detect dementia are cognitive tests such as Clinical Dementia Rating (CDR) or Mini-Mental State Examination (MMSE), genetic test, laboratory tests or neuroimaging technologies - Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computed Tomography (CT), Diffusion Tensor Imaging (DTI). With the raging artificial intelligence applications in medical sciences and knowledge that Artificial Intelligence (AI) are excellent pattern recognizers it is crucial to try and employ Machine Learning (ML) and Deep Learning (DL) to solve the need for accurate diagnosis of early dementia utilizing multi-modal data, because most often dementia is detected in the last stage, where medication can no longer help the patient.

After reviewing existing literature we identified the main research gap - a lack of review, which would focus on multi-modal data usage in detection of early dementia. One of the most frequent conclusions from reviewed articles was the lack of multi-modal methods. Therefore, this systematic literature review article will focus on reviewing existing research works which focus on multi-modal data utilization in detection algorithms.

List of scope elements (aims) to be identified in this research:

- Different dementia detection types;
- State-of-the-art methods to detect early dementia through multi-modal feature analysis;
- Most frequently used combinations of modalities with neuroimaging - neuroimaging proved itself throughout the years being one of the time-tested default choices when it comes to diagnosing dementia;
- Evaluation of performance and validation of such models;
- Datasets used in the studies;
- Common pre-processing and feature extraction techniques from neuroimages;
- Existing issues in the field;
- Future research areas.

Most of the found literature reviews were not looking at the early dementia detection from multi-modal perspective. Therefore, our main contribution and novelty of this research is a systematic literature review which only focuses on multi-modality as being the key driver in improving AI models' performance. That means we only review research works, which utilize multi-modal data.

The paper is structured as follows: Section 2 summarizes existing literature reviews. Section 3 briefly describes research area background knowledge. Section 4 defines methodology used in this systematic literature review and gives a brief overview of the found articles. Section 5 presents found scientific articles during research, answers defined research questions and discuss identified issues, future research areas. Section 6 concludes the paper. Section 5.10 briefly describes limitations of this study.

2 Related works

In this section all articles will be listed that were found during the first stage of literature review. Stages will be discussed in the research methodology Section 4. In the introduction chapter we briefly mentioned all found and reviewed related literature reviews. In Table 1, we show the summary of all identified papers: paper first author and reference, when the paper is published, what dementia type was the main focus of the literature review, what issues in the research area and future prospects we could identify from the paper, and some additional conclusions.

Findings are summarized and discussed in results subsection 5.1. Concluding findings of the related literature reviews is one of the research aims.

The main difference between existing literature reviews and this study is the focus on multi-modal data, where one of the modalities used is neuroimaging. After reviewing all found literature reviews, we identified that there is a lack of such study and therefore decided to investigate this research area, which can be defined as a novelty of this study.

3 Problem domain

This section aims to provide background knowledge on early dementia detection categorized into several topics: risk factors, stages, types, traditional diagnosis, treatment, statistical influence of dementia on economics and demographics. After reading this section hopefully the reader will have a better understanding of the severity of this problem.

Table 1 Found related literature reviews

Source	Year	Focus	Issues / Additional extracted conclusions	Future research areas
Tăuţan et al. [4]	2021	Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD)	**A**vailable annotated data is not large enough for DL requirements. **N**o standard for evaluating models. **D**ata leakage. **B**rain-imaging is powerful, but expensive and cannot be easily accessed by public.	**G**enerative Adversarial Networks (GAN) could be used to generate synthetic data for small datasets problems. **T**ransfer learning for better accuracy. **U**se of electroencephalography (EEG) signals could be better explored in consumer environment. **M**ulti-modality.
Li et al. [5]	2022	Dementia	**M**ost diagnosis tools require trained personnel to administer results. **L**ack of tools which recognizes early dementia. **P**erformance of tools are not great. **M**odalities such as gait, eye, hand movements are not yet fully utilized, because of the need for additional equipment. **D**eep learning methods require a lot of data. **D**eep learning methods lack of explainability which slows down acceptance. **C**ombining speech and text yields better performance than running tests separately.	**S**mart environments, which utilize sensors, cameras, and computer vision to monitor and diagnose dementia. **M**ulti-modality usage because most diagnosis tools currently created to work separately.
Battineni et al. [6]	2022	Dementia	**L**ack of interpretability. **A**I is quick to analyze large population of data in comparison to trained personnel.	**E**xplainability of AI may improve clinicians' adoption. **S**elf-learning AI to improve generalizability. **S**mart environments. **M**ulti-modality.
Saravanan et al. [7]	2022	PD	**I**mbalanced datasets. **E**EG signals have limitations due to low spatial resolution and contaminate various artifacts. **S**peech signals and the quality of speech is affected due to ambient environment and motion artifacts.	**F**ine-tuning can improve deep learning model performance. **O**ther metrics than specificity and sensitivity should be experimented with to evaluate model performance. **M**ulti-modality may improve prediction accuracy.
Goyal et al. [8]	2022	AD, Mild Cognitive Impairment (MCI)	**M**ost of tools use different modalities separately and can generate large and redundant data that provides false results. **T**ransform-based features provide good results compared to other feature extraction methods.	**M**ulti-modality could increase accuracy and improve evaluation of the classification process.
Billeci et al. [9]	2020	AD, MCI	**T**here is a lack of study concerning MCI. **C**lassification between AD and HC (Healthy Control or Cognitive Normal) performs better than between AD and MCI or MCI and HC.	**M**ulti-modality could be utilized to improve performance of classifications.

Table 1 (continued)

Source	Year	Focus	Issues / Additional extracted conclusions	Future research areas
Loh et al. [10]	2021	PD	**A**doption of existing deep learning tools is nonexistent. **B**lackbox models, impossible to interpret. **T**rust issues of doctors. **M**ethods do not utilize multi-modality enough. **I**mprecise nature of clinical diagnoses, and a lack of neurology experts in PD worldwide usually leads to delayed diagnosis and suboptimal management of PD. **S**uccess of gene therapy may be influenced by early diagnosis tools.	**V**isual cues - segmentation of abnormal areas in brain as a normal output of model (to improve interpretability and explainability). **S**elf-assessment (may encourage individuals to seek help). **M**ore smart-home solutions, which can monitor and diagnose.
Fathi et al. [11]	2022	AD, MCI	**D**ata leakage - using parts of tests set in the training phase. **S**mall datasets might yield overfitting models. **N**o benchmark platform to validate all models in the same manner. **S**election of hyper-parameters is usually done via heuristic search, but that could cause non-optimal models. **P**ET based models yields excellent results but are not reliable due to data leakage.	**B**enchmark platform to evaluate all models in the same way. **M**ulti-modality. **T**ransfer-learning alternative for training from scratch. **S**ome recent Convolutional Neural Network (CNN) architectures could be investigated in early AD detection (HRNet, CapsNet, MobileNet-V2, ConvNeXt).
Pereira et al. [12]	2019	PD	**I**mage processing techniques are strongly dependent to the quality of the input and are more prone to errors. **M**ost tools try to utilize signal analysis, which are gathered from on-body sensors.	**S**mart home environments for detection and monitoring. **A**ugmented-reality or virtual-reality research area could be more explored. **I**mage-based data is not yet fully explored in PD detection.
Khan et al. [13]	2021	AD, PD	**E**xplainability of diagnosis. **T**he quality of available annotated datasets is poor because it is time consuming to prepare annotations. **H**aving many different diagnostic tools, it is possible that regulations and laws might be valid for some, but not for others. This introduces interoperability problems. **S**ome DL methods require a lot of computational power. **D**L/ML algorithms should be retrained separately with single disease-oriented datasets, because models trained for couple diseases might not work well with other if dataset is imbalanced or poor quality.	**E**xplainable AI might make use of created models in clinical practice. **E**xpanding datasets with synthetic data. **M**ulti-modality. **T**here is a need to investigate whether existing techniques are resource efficient when detecting brain disorders. **G**AN's can be used to reduce data scarcity problem.
Ahmed et al. [14]	2019	Dementia	**R**adiotracers used in image collection may have some side-effects. **L**arge neuroimaging datasets require a lot of labeling time to train models. **N**oise in the images reduces accuracy. **D**ata heterogeneity. **C**linical features in medical imaging may be subjective and thus ground truth images may not be perfect. **O**ther diseases than Alzheimer's lack of focus.	**T**here would be a routine clinical practice of neuroimaging to diagnose dementia. **M**ulti-modality. **E**EG could be utilized better as a modality and data may potentially be collected in patient's home.

Table 1 (continued)

Source	Year	Focus	Issues / Additional extracted conclusions	Future research areas
Agarwal et al. [15]	2021	AD	**O**verfitting due to small datasets. **H**ard to reproduce solutions due to randomness being involved in training. **D**ifficult to determine what features are better in terms of accurate decision. **D**eep learning models with Three-dimensional (3D) data tend to train for long time. **M**inimum pre-processing is recommended for neuroimaging modalities - intensity normalization and registration.	**t**au-PET and amyloid-PET could be explored as additional biomarkers in diagnosis. **M**ulti-modality for better feature extraction. **T**ransfer learning for better accuracy. **U**se different datasets for training and validation to develop reproducible results.
Blanco et al. [16]	2023	MCI, AD	**M**odels trained on small datasets usually overfit. **A**bsence of studies in Latin America, Carribean and Africa regions. **L**ongtitudal studies usually use multi-modal data. **A**rea Under Curve (AUC) is a more robust performance metric, that considers both sensitivity and specificity.	**M**ulti-modality. **T**ools in longtitudal studies should handle sparse and imcomplete data. Increase explainability and interpretability with white-box ML models.
Zhao et al. [17]	2023	AD	**S**mall datasets in AD classification compared to computer vision due to medical data privacy. **C**lass imbalance. **D**ata leakage. **H**igh computation cost for DL methods.	**M**odel should be chosen based on hardware, specific application requirements, balancing performance and complexity.
Viswan et al. [18]	2023	AD	**E**xplainable Artificial Intelligence (XAI) researchers resort to their own intuition on what is a good explanation without advising medical professionals. **X**AI based AD diagnosis usually lacks ground truth data. **C**onfidence of neural networks should be crucial when prediction may be life-threatening. **N**o comprehensive benchmark for real-world validation.	**D**ecision explanations have to be evaluated for intuitiveness and also being the optimal ones. **M**edical experts involved in the interpretation for every data modality is a way forward. **M**ore data modalities is better.
Aggarwal et al. [19]	2023	PD	**D**ata imbalance. **S**mall datasets tend to overfit. **D**L is computationaly expensive. **E**EG and Magnetoencephalography (MEG) do not directly visualize brain structure, but provide a way to have insights on the functional aspects of the brain. They should be used together with neuroimaging modalities.	**M**ulti-modality. **T**here is a need for standard Parkinson's disease dataset, which would include CT, functional MRI (fMRI), DTI and MRI modalities.
Arya et al. [20]	2023	AD	**E**nsemble learning usually improve performance. **T**o overcome small dataset problem, multiple databases could be merged or transfer learning technique used.	**M**ulti-modality. **E**nsemble learning. **E**xplainability.

Table 1 (continued)

Source	Year	Focus	Issues / Additional extracted conclusions	Future research areas
Huang et al. [21]	2023	AD, PD	**I**ncomplete and noisy data. **C**ollecting neuroimaging data is time consuming and expensive. **E**xisting multi-modal solutions only focus on neuroimaging modalities, but do not complement with other types of modalities.	**M**ulti-modality.
Muhammed Niyas and Thiyagarajan [22]	2023	MCI	**I**dentifying relevant bio-markers and features from neuroimaging data is a challenging task for researchers. **M**ulti-modal solutions show better results.	**M**ore experiments with different regions around the world. **F**inding quick MCI to AD converters (within 6 months) using less number of follow up data.

Table 2 The Clinical Dementia Rating

Score	Impairment severity
0	Cognitive normal
0.5	Questionable or very mild
1	Mild
2	Moderate
3	Severe

3.1 Risk factors

There are several identified risk factors of dementia [23]: excessive use of alcohol, traumatic brain injury, air pollution, lack of education, high blood pressure (hypertension), loss of hearing, heavy smoking, overweight, depression, physical inactivity, diabetes, social contact. Some of these risk factors can be managed, modified or avoided. Vascular risk factors (factors which can cause cardiovascular disease development) [24]: hypertension, diabetes, obesity, smoking, can be managed to reduce the possibility of developing dementia. Employing a healthy, active lifestyle, quitting smoking, treating hypertension and diabetes on time, can greatly reduce the risk of dementia [25–27]. Modifiable risk factors are physical inactivity, social contact, education. Avoidable risk factors are use of alcohol and smoking. However, we cannot easily control air pollution, traumatic brain injuries or a loss of hearing.

3.2 Stages

Dementia has a different effect on every person, which depends on the type of it, whether he/she has multiple dementias, and at which stage the disease was noticed. There are three stages of dementia: early, middle and late [1]. Each stage has a set of common symptoms. In diagnosis practice, psychiatrists use scales, which allow to quantify cognitive abilities of a patient into severity categories logically equivalent to different stages of dementia. Most frequently used scales are Clinical Dementia Rating (CDR) [28] and Mini-Mental State Examination (MMSE) [29], which are used as cognitive abilities tests [30]. Both quantify cognitive impairment in continuous scale, CDR (0 - 3) and MMSE (0 - 30). See the CDR scale in Table 2 below:

See the MMSE scale in Table 3 below:

From the scales, we can see that CDR is more granular than MMSE, although they have similar categories of

Table 3 Mini-Mental State Examination scores

Score	Impairment severity
0-17	Cognitive normal
18-23	Mild
24-30	Severe

Table 4 Common symptoms of dementia during different stages

Stage of dementia	Common symptoms
No dementia or cognitive normal	No memory loss, difficulties with orientation, problem solving; Can live independently and does not require assistance for self-care.
Early (Very mild)	Occasional memory loss, but frequently being misidentified as lack of rest or sleep; Does not experience difficulties with orientation, however, sometimes one can catch himself thinking of the solutions for common problems longer than usual; Does not require any assistance.
Middle (Mild)	Constant moderate memory loss; Difficulty with orientation of time, location, problem solving; Usually cannot live independently and requires slight assistance for self-care.
Late (Severe)	Severe memory loss; No orientation except for people; Unable to solve problems; Usually in nursing home due to challenging nature of taking care of such people; Requires significant assistance with self-care.

severity: Cognitive Normal (CN), Mild Cognitive Impairment (MCI) and severe cognitive impairment. Common symptoms between different stages of dementia are briefly described in Table 4 below:

There are also other type of cognitive impairment tests such as: Alzheimer's disease assessment scale (ADAS) [31], Rey Auditory Verbal Learning Test (RAVLT) [32], Geriatric Depression Scale (GDS) [33], or questionnaires such as: Functional Activities Questionnaire (FAQ) [34], Neuropsychiatric Inventory Questionnaire (NPIQ) [35].

3.3 Types

This syndrome can be categorized into multiple different sub-categories (diseases or disorders), which have their own symptoms. See Table 5 for symptoms categorized for each common dementia type, which were picked out from a set [36–44] of articles.

Each type of dementia has its symptoms. However, common symptoms are shared across all of them: some degree of memory loss, difficulties of doing usual activities, confusion, mood or behavior changes. Symptoms of dementia tend to get worse over time, therefore it is important to diagnose it as early as possible to prepare for the future.

3.4 Traditional diagnosis

Traditional ways of diagnosing dementia are cognitive abilities tests, such as CDR [28], MMSE, Abbreviated Mental Test Score (AMTS) and Modified Mini-Mental State Examination (3MS) [30], which detects cognitive impairments. Some recent research [45] finds new tests such as Cognetivity's Integrated Cognitive Assessment (CognICA) being a sufficient detector of early-stage Dementia. Laboratory tests are another way of diagnosing dementia, one of more recent research [46] finds reference between specific proteins, particularly higher pTau181 and lower beta-amyloid, found in patients' blood can provide enough evidence about having Alzheimer's. There are also other types of evaluations, for example, a visit to psychiatric may help to identify any mental health issues that may cause dementia symptoms. Also, genetic tests exist [47], which detects single-gene changes, that can influence the development of the dementia. Other most frequently used diagnostic tools are brain scans.

Table 5 Symptoms specific to each common dementia type

Disease/Dementia	Symptoms
Alzheimer's disease	Memory loss issues, confusion, difficulties while talking and thinking, muscle coordination issues. Patients find it hard to recognize objects.
Vascular dementia	Similar to Alzheimer's, but with less obvious memory loss issues in the early stage. Most likely have mood issues, even depression, finds difficulties while thinking rationally, keeping attention. Difficulties walking, stroke-like symptoms: weak muscles, paralysis.
Frontotemporal dementia	A group of neurodegenerative diseases. Change of personality, aphasia (linguistic problems), obsessiveness, behavior changes, compulsiveness.
Dementia with Lewy bodies (Parkinson's disease is a sub-type)	Has similar to Alzheimer's symptoms with combination of confusion, hallucinations, fainting and sleep disorders, parkinsonism (tremors, stiffness, postural instability).
Huntington's disease	Aggressiveness, mood problems, delusions, depression, feeling dissatisfied, lack of initiative, poor self-care.
Human Immunodeficiency Viruse (HIV) dementia	This type of dementia is classified as a complication of HIV infection. Frequent symptoms are cognitive impairment, behavior change, movement and/or muscle problems. Often leads to death.
Mixed dementia	Symptoms mixed and depends on the combination of dementias patient has.

Typically, a combination of different brain scans can show multiple problems, which can cause dementia symptoms. CT, MRI, PET scanning are common choices to detect a possible cause of dementia. These scans can show tumors, strokes or identify transformations in the brain. Although usually dementia is diagnosed by eliminating everything else, which could cause any symptom by known diseases.

3.5 Treatment

Presently, there is no cure for dementia [1, 48–50]. However, early diagnosis and symptomatic treatment [51–56] can lengthen the period where one can live independently and actively until the decline of cognitive abilities becomes unbearable. One of the treatment options is medication. Majority of medications are focused on Alzheimer's disease, because it is the most common type of Dementia [57]. These are the frequently prescribed medications (acetylcholinesterase inhibitors): donepezil, rivastigmine, galantamine. Studies finds that these medications help to treat Lewy body and vascular dementia, Parkinson's and Alzheimer's diseases, mixed dementia forms [57]. Another type of medication is memantines (medications which treat memory loss), these can also treat the mentioned dementia types.

Other forms of treatments are related to being active, stimulating brain activity by training memory, learning new languages, and improving problem-solving skills. These can be grouped into Cognitive Stimulation Therapy (CST) which might help people who have mild to moderate dementia. However, there is a lack of evidence, that these activities can be effective as medication [58].

3.6 Statistics

The World Health Organization (WHO) estimates that presently 55 M people live with dementia. This number could reach 78 M by 2030 and 139 M by 2050 [1]. In Europe alone there are around 11 M people with dementia [59].

If we look how the disease prevalence is categorized by sex, we can see that women are more affected by dementias in all age ranges [60].

Alzheimer's disease is number 1 diagnosed dementia with around 50-60% contribution [60, 61]. Where vascular dementia falls to second place with 25% and dementia with Lewy bodies - third with 15%.

Ten to fifteen percent of patients with MCI develop dementia each year [62].

Alzheimer's Association states in their facts and figures that [63] "1 in 3 seniors dies with Alzheimer's or another dementia. It kills more than breast and prostate cancer combined". Dementia and Alzheimer's disease is listed with number 7 in top 10 causes of deaths worldwide [64] reaching nearly 2 M deaths. Dementia being in the top 10 of causes of deaths emphasizes importance of help for people suffering from this disease. On average people with Alzheimer's and other dementias over age 65 live 4 - 8 years after diagnosis, sometimes even 20 years [63]. This explains how sometimes dementia progression can be slow, tiring, and unpredictable.

The cost of care for patients who have dementia is increasing. Annually one patient care costs around US $19k. The calculated cost worldwide in 2015 reached $167B and predicted to cost $500B in 2030 and $1.9T in 2050 [65]. This reflects the need to find a cure for dementia and related diseases.

4 Methodology

Systematic literature review allows us to identify, examine, and evaluate all existing relevant research to the purpose of this study. This systematic literature review is based on Preferred Report Items for Systematic Reviews and Meta-Analysis statements (PRISMA) methodology, which serves purpose, that allows to perform easier replication and quicker evaluation of search results as well as finding all research area gaps, that can be utilized for new research options.

4.1 Design

This systematic literature review will be carried out by raising questions relevant to the research area and then answering with discussion of results conducted through database screenings.

4.1.1 Research questions

The research will address these questions:

1. Is there any existing work in this area of research (literature reviews)? What conclusions about future research areas and existing problems other researchers identified?
2. What types of dementia detection researchers focus their research on?
3. What are state-of-the-art methods in detection of early dementia in the field of AI/ML/DL?
4. What modalities are being used in combination with neuroimaging to detect early dementia?
5. How is model performance evaluated and validated?
6. What datasets researchers used in their studies?
7. What are the commonly used pre-processing and feature extraction from neuroimages techniques?
8. What are the key issues in detection of early dementia?
9. What are the potential future research areas?

4.1.2 Process

The systematic literature review will involve two stages:

1. Collecting existing literature reviews relevant to the research purpose and the discussion of the results (answering first research question);
2. Collecting existing research works relevant to research purpose and discussion of the results (answering the rest of the research questions).

Each stage involves database screening with relevant to research area keywords. This research focuses on finding journal articles from accredited online databases. Search engines were used to find relevant studies.

4.1.3 Database screening

As previously mentioned, online databases were used to collect papers relevant to this research. Selected databases can be found in Table 6. These databases were selected, because they can provide good enough search results for highest quality journal article papers relevant to the detection of dementia using AI and ML. Institutional access of Kaunas University of Technology was used in retrieval of the research papers. Only full text papers were included in the search results. To accommodate the necessity to find the state-of-the-art methods in the research area, all search results are filtered to be published from 2018 until present.

These keywords were identified as search terms: early, dementia, Alzheimer's, Parkinson's, Huntington's, mild cognitive impairment, MCI, artificial intelligence, machine learning, deep learning, neuro, medical, diagnosis, recognition, identification, detection.

For the first stage of systematic literature review, additional search terms were included: review, systematic review. Search terms are aggregated into a search query with logical operations (AND) and (OR). Search queries are listed below:

- IEEE Xplore

  ("Review" OR "Systematic review") AND "Early" AND ("Dementia" OR "Alzheimer's" OR "Parkinson's" OR "Huntington's" OR "Mild cognitive impairment" OR "MCI") AND ("Artificial intelligence" OR "Deep learning" OR "Machine learning" OR "Neuro" OR "Medical") AND ("Diagnosis" OR "Recognition" OR "Identification" OR "Detection"). Part of the query: "("Review" OR "Systematic review") AND", is removed on Stage 2.
- Scopus
  - Stage 1:

    TITLE-ABS-KEY("Review" OR "Systematic review") AND TITLE-ABS-KEY("Early") AND TITLE-ABS-KEY("Dementia" OR "Alzheimer's" OR "Parkinson's" OR "Huntington's" OR "Mild cognitive impairment" OR "MCI") AND TITLE-ABS-KEY("Artificial intelligence" OR "Deep learning" OR "Machine learning" OR "Neuro" OR "Medical") AND TITLE-ABS-KEY("Diagnosis" OR "Recognition" OR "Identification" OR "Detection") AND PUBYEAR>2017 AND (LIMIT-TO(PUBSTAGE, "final")) AND (LIMIT-TO(DOCTYPE, "re")) AND (LIMIT-TO(SUBJAREA, "COMP")) AND (LIMIT-TO(LANGUAGE, "English")).
  - Stage 2:

    TITLE-ABS-KEY("Early") AND TITLE-ABS-KEY("Dementia" OR "Alzheimer's" OR "Parkinson's" OR "Huntington's" OR "Mild cognitive impairment" OR "MCI") AND TITLE-ABS-KEY("Artificial intelligence" OR "Deep learning" OR "Machine learning" OR "Neuro" OR "Medical") AND TITLE-ABS-KEY("Diagnosis" OR "Recognition" OR "Identification" OR "Detection") AND PUBYEAR>2017 AND PUBYEAR<2023 AND (LIMIT-TO(PUBSTAGE, "final")) AND (LIMIT-TO(SUBJAREA, "COMP")) AND (LIMIT-TO(DOCTYPE, "cp") OR LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(SRCTYPE, "j") OR LIMIT-TO(SRCTYPE, "p")).
- Web of Science
  - Stage 1:

    (AB= "Review" OR AB="Systematic review") AND AB="Early" AND (AB="Dementia" OR

Table 6 Selected databases

Database	URL
IEEE Xplore	https://ieeexplore.ieee.org/search/advanced, (last access on 14 December 2023)
Scopus	https://www.scopus.com/search/form.uri?display=basic#basic, (last access on 14 December 2023)
Web of Science	https://www.webofscience.com/wos/woscc/advanced-search, (last access on 14 December 2023)
Science Direct	https://www.sciencedirect.com/search, (last access on 14 December 2023)
Springer link	https://link.springer.com/advanced-search, (last access on 14 December 2023)

AB="Alzheimer's" OR AB="Parkinson's" OR AB="Huntington's" OR AB="Mild cognitive impairment" OR AB="MCI") AND (AB="Artificial intelligence" OR AB="Deep learning" OR AB="Machine learning" OR AB="Neuro" OR AB="Medical") AND (AB="Diagnosis" OR AB="Recognition" OR AB="Identification" OR AB="Detection") OR (TI="Review" OR TI="Systematic review") AND TI="Early" AND (TI="Dementia" OR TI="Alzheimer's" OR TI="Parkinson's" OR TI="Huntington's" OR TI="Mild cognitive impairment" OR TI="MCI") AND (TI="Artificial intelligence" OR TI="Deep learning" OR TI="Machine learning" OR TI="Neuro" OR TI="Medical") AND (TI="Diagnosis" OR TI="Recognition" OR TI="Identification" OR TI="Detection") AND Review Article (Document Types) AND 2018 OR 2019 OR 2020 OR 2021 OR 2022 (Publication Years) AND Review Article (Document Types) AND Book Chapters (Exclude - Document Types) AND Engineering OR Computer Science (Research Areas).

  - Stage 2:
    TI="Early" AND (TI="Dementia" OR TI="Alzheimer's" OR TI="Parkinson's" OR TI="Huntington's" OR TI="Mild cognitive impairment" OR TI="MCI") AND (TI="Artificial intelligence" OR TI="Deep learning" OR TI="Machine learning" OR TI="Neuro" OR TI="Medical") AND (TI="Diagnosis" OR TI="Recognition" OR TI="Identification" OR TI="Detection") OR AB="Early" AND (AB="Dementia" OR AB="Alzheimer's" OR AB="Parkinson's" OR AB="Huntington's" OR AB="Mild cognitive impairment" OR AB="MCI") AND (AB="Artificial intelligence" OR AB="Deep learning" OR AB="Machine learning" OR AB="Neuro" OR AB="Medical") AND (AB="Diagnosis" OR AB="Recognition" OR AB="Identification" OR AB="Detection") AND 2022 OR 2021 OR 2020 OR 2019 OR 2018 (Publication Years) AND Article (Document Types) AND Early Access OR Book Chapters OR Proceeding Paper (Exclude - Document Types) AND English (Languages) AND Computer Science (Research Areas).

- Science Direct

  - Stage 1:
    Find articles with these terms: ("Review" OR "Systematic review") AND "Early" AND ("Dementia" OR "Alzheimer's" OR "Parkinson's" OR "Huntington's" OR "Mild cognitive impairment" OR "MCI"). Title, abstract or author-specified keywords: ("Artificial intelligence" OR "Deep learning" OR "Machine learning" OR "Neuro" OR "Medical") AND ("Diagnosis" OR "Recognition" OR "Identification" OR "Detection"). Query is filtered to only "Review articles" in a field of "Computer Science".
  - Stage 2:
    Same as Stage 1, but part of the query: "("Review" OR "Systematic review")", is removed. Query is filtered to only "Research articles" in a field of "Computer Science".

- Springer link
  ("Review" OR "Systematic review") AND "Early" AND ("Dementia" OR "Alzheimer's" OR "Parkinson's" OR "Huntington's" OR "Mild cognitive impairment" OR "MCI") AND ("Artificial intelligence" OR "Deep learning" OR "Machine learning" OR "Neuro" OR "Medical") AND ("Diagnosis" OR "Recognition" OR "Identification" OR "Detection"). Part of the query: "("Review" OR "Systematic review") AND", is removed on Stage 2. Query was filtered to only "Computer Science" field. All finding were further filtered to only include "Articles".

4.2 Reviewing procedure

This section will describe the process which was used to carry out this systematic literature review.

4.2.1 Selection of papers

As described in the previous sections, this systematic literature review will have two stages. In the first one, database screening is done to find all already existing reviews, which correlates to the purpose of this research. The same process will be done for the second stage in which we collect all research papers instead of review articles.

4.2.2 Inclusion and exclusion criteria

To carry out systematic literature review inclusion and exclusion criteria were defined to help differentiate studies relevant to selected research area. All papers which did not meet inclusion criteria were removed from the eligibility study.

Inclusion Criteria

- Stage 1:
  1. Articles reviewing different AI and ML methods for early detection of dementia.
- Stage 2:
  1. Articles proposing a method for the detection of early dementia;

2. Articles which compare different methods for the detection of early dementia.

Exclusion Criteria

- Stage 1:
  1. Not a systematic review article.
- Stage 1 and 2:
  1. Articles which do not mention anything related to dementia in title or abstract;
  2. Non-English articles;
  3. Articles which do not focus on early detection of dementia;
  4. Articles which do not use any AI/ML/DL methods/techniques;
  5. Duplicate articles;
  6. Articles which do not clearly report results and conclusions;
  7. Conference proceedings or papers;
- Stage 2:
  1. Review article.
  2. Does not use multi-modal approach;
  3. Does not use neuroimaging as one of modalities;
  4. Does not provide details about feature extraction;
  5. Does not provide details about validation scheme;
  6. Does not provide any details whether data was pre-processed.
  7. Articles which do not provide details about evaluation techniques;

Short summary of proportion of initial and final selected papers can be seen in Fig. 1. Figure represents initial count of selected sources for each database which was screened as well as final number after eligibility study.

4.2.3 Collection of data

Data collected during the database screening is described below:

1. Authors;
2. Title;
3. Year of publication;
4. Publication medium – journal title;
5. Number of references;
6. Country of origin – researchers' country of origin;
7. Contribution of the article – main contribution of the article as described by the authors;
8. Subjective summary – summary of the article from this research author's perspective;
9. What dementia type was investigated;
10. Types of modalities – what types of modalities were used in the detection of dementia (images, text, speech etc.);
11. AI/ML/DL methods used – what type of AI/ML/DL methods were used in the research;
12. Evaluation techniques – what type of evaluation was performed for the proposed or compared method(s).
13. Validation techniques - how models were validated;
14. Pre-processing techniques - what kind of pre-processing methods were used;
15. Feature extraction techniques - how features were extracted from data samples;

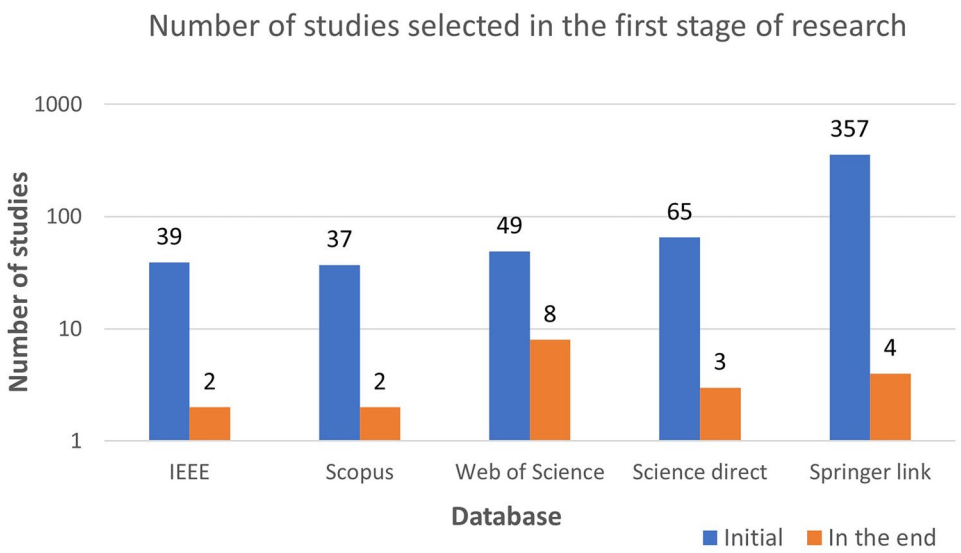

Fig. 1 Proportion of papers selected in the first stage of the study

16. Issues - what issues were identified by the authors;
17. Future research areas – what future research areas were identified by the authors.

4.3 Overview of collected data

4.3.1 Stage 1 overview

During the first stage, 547 review articles were identified through database screening. After removing duplicates, and title with abstract screening step, 476 sources were eliminated leaving with 71 review articles on which full text eligibility inspection was carried out. After full text inspection an additional 52 articles were eliminated. Only 19 review articles met inclusion criteria. High level overview of performed steps is illustrated in Fig. 2.

All review articles used in this study are from 2018 to 2023. Figure 3 represents year distribution of articles after assessing eligibility. Most of the articles are published in 2023 and 2022 - 12 articles out of 19 selected. This means that the research area is relatively new and has become more popular in recent years.

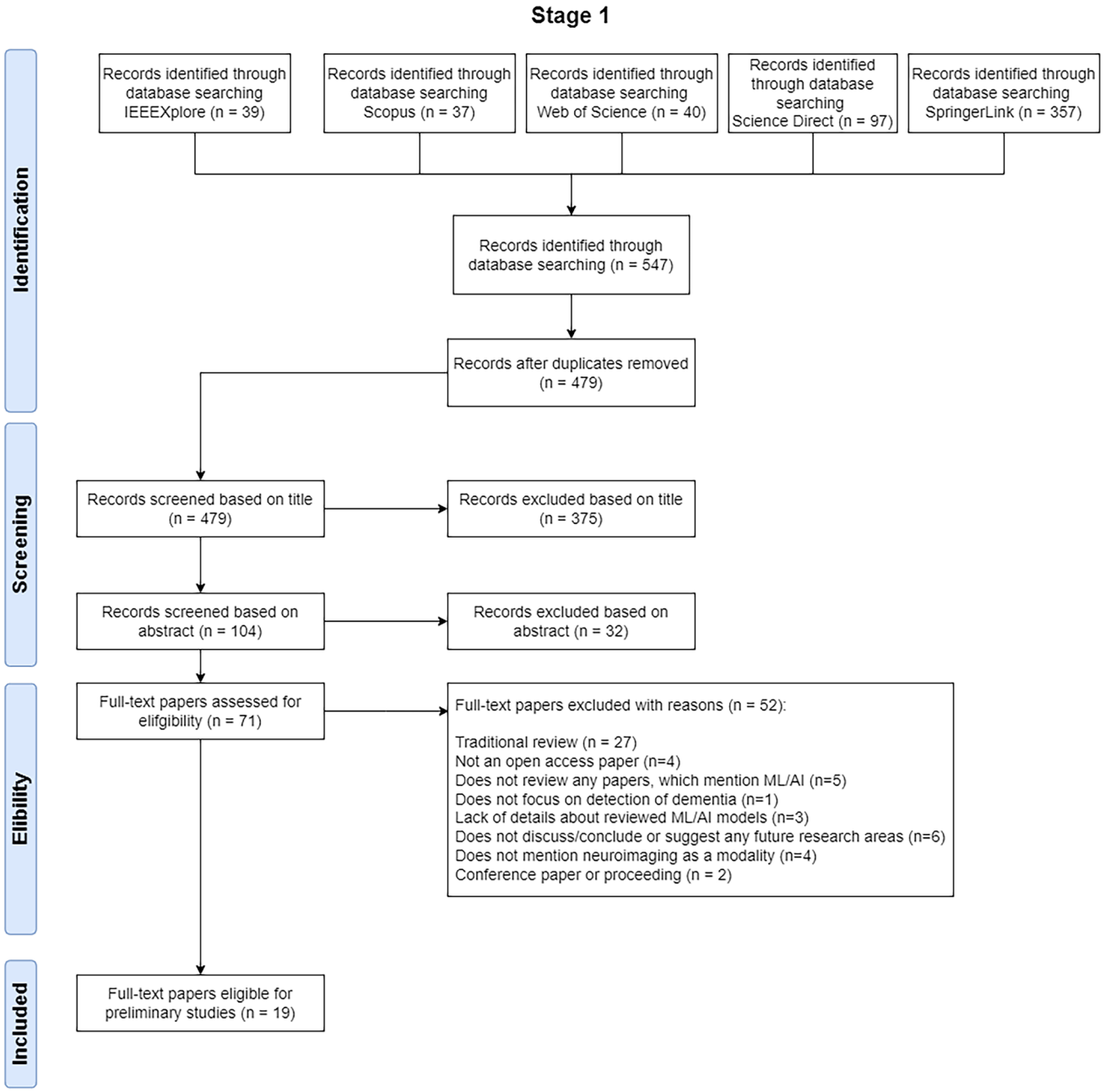

Fig. 2 PRISMA methodology chart for 1st stage of systematic literature review

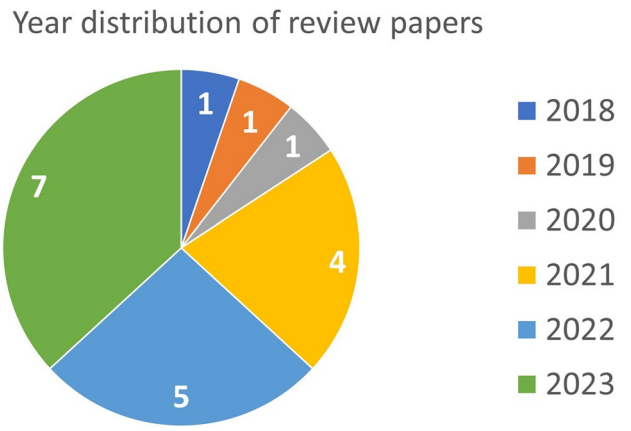

Fig. 3 Year distribution of selected review papers in the first stage of the study

Country of origin of publication first author was also collected during review to obtain overview of contributions from research community around the globe. Geographical distribution can be seen in Fig. 4. Clearly India contributes the most articles.

Distribution of scientific journals of literature reviews can be seen in Table 7. Most popular journals are "Artificial Intelligence in Medicine", "Archives of Computational Methods in Engineering" and "Sensors". Most of the journals in one way or another are related to the research area this study is investigating, except for more generic "Processes" and "IEEE Access" and "Sensors". The highest citations median (160) is observed from journal "Artificial Intelligence in Medicine". High number of references included in articles, particularly in literature reviews, shows if the study is of higher quality (more studies reviewed is equal to better insights of research area).

4.3.2 Stage 2 overview

During the second stage, 2311 articles were identified through database screening. After removing duplicates, reviewing titles and abstracts, 79 articles were selected for full text eligibility assessment. After full text review an additional 39 articles were eliminated and only 40 articles met inclusion criteria. High level overview of performed steps is illustrated in Fig. 5.

Like the first stage, all selected articles were published during the timeframe of 2018 to 2023. Figure 6 illustrates year distribution of articles which were reviewed for eligibility during the second stage of the research. Same observation can be seen as in first stage, majority of the articles are published in the last 3 years. In the figure we can also see that the popularity of the research area is high.

Distribution of countries for the second research step was also observed and is illustrated in Fig. 7. The most contributing countries to the research area are China and India, where authors from China published majority of the selected articles (24 articles).

During the second research stage a list of all journals from scientific articles which were reviewed for eligibility was generated. The most popular journals where researchers published findings from experiments are "Computers in Biology and Medicine" (8 articles), "Biomedical Signal Processing and Control" (3 articles), "Medical Image Analysis" (3 articles) and "Neural Computing and Applications" (3 articles). The full list of journals is provided in Table 8. Highest citations median (65) is from journal "Medical Image Analysis", following by journal "Computers in Biology and Medicine" with citation median 56.5. From the quality assessment we can see, that mostly high quality articles were included in the study.

The full list of selected papers with references, publication year, dementia type is provided in the Table 9.

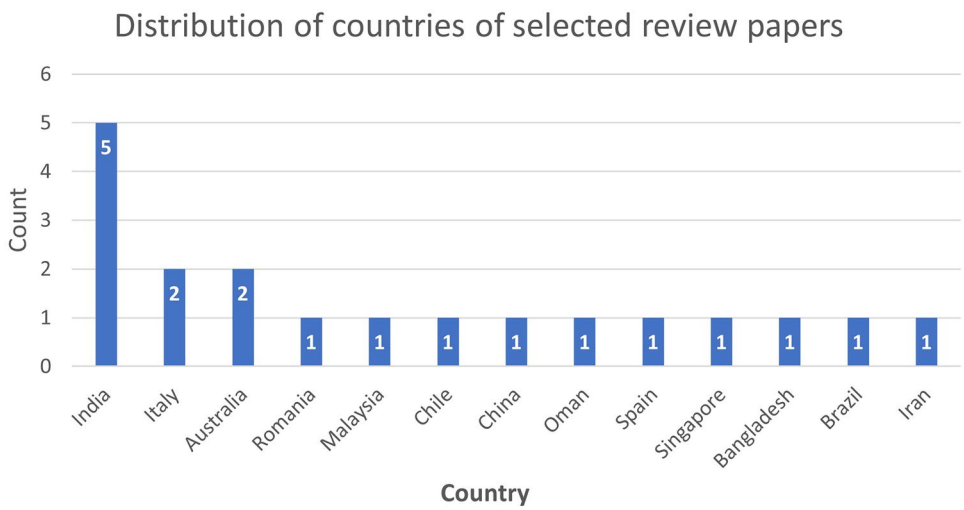

Fig. 4 Distribution of countries of selected review papers in the first stage of the study

Table 7 List of names of journals from papers selected in the first stage of the study and quality assessment

Journal	Number of citations in publications	Citations median
Cognitive Computation	[18] - 179, [19] - 157	168
Artificial Intelligence in Medicine	[4] - 198, [12] - 122	160
Archives of Computational Methods in Engineering	[8] - 185, [7] - 107	146
Sensors	[15] - 156, [10] - 123	139.5
Journal of Biomedical Informatics	[5] - 254	-
IEEE Access	[13] - 204	-
Alzheimer's Research and Therapy	[16] - 149	-
Computers in Biology and Medicine	[11] - 129	-
Bioengineering	[6] - 73	-
IEEE Reviews in Biological Engineering	[14] - 94	-
Health Information Science and Systems	[21] - 77	-
Frontiers in Computational Neuroscience	[17] - 73	-
Processes	[9] - 72	-
International Journal of Intelligent Networks	[22] - 67	-
Brain Informatics	[20] - 47	-

5 Results and discussion

In this section research findings will be presented in a form of answers to defined research questions from sub-section 4.1.1.

5.1 Is there any existing work in this area of research (literature reviews)? What conclusions about future research areas and existing problems other researchers identified?

All the related works (19 articles) found in the research area are represented by Table 1. The summary of issues researchers mentioned is displayed in the Table 10 (only the common issues are included and are sorted by number of mentions in decreasing order).

Existing problems:

Each common issue is discussed in the list below:

1. Explainability

   The biggest problem in the research area currently is the explainability of the results which are produced by ML/DL method. Most of the ML/DL methods are "blackboxes" which takes in an input and outputs some sort of decision, but there is no way to explain the solution proposed, which makes it nearly impossible for medical personnel to trust the ML/DL based methods. This problem mostly exists in areas where critical decisions are made, for example, a method which yields a diagnosis whether person has MCI or not based on data, that was given for the method. In a positive classification case, it is crucial that medical personnel could examine why the decision was imposed. In the literature we can find many different approaches used to increase transparency and explainability of models, like Shap values [106] or Grad-CAM [107] and others.
2. Overfitting small datasets

   One of the biggest problems in this research area is the datasets with limited data. Collecting data with multiple modalities is expensive and therefore most of the datasets are too small to train a DL model. However, researchers are still experimenting and there is a potential that good results showing methods have collected data from experiments where models have overfitted these small datasets, because of small variance in data.
3. Noisy, poor-quality data

   Most of the researchers in dementia detection are dependent on publicly available datasets. One clear observation from the collected data is that those datasets are usually low quality, noisy and require additional pre-processing steps to clean and prepare data for training/validation/testing. This signals the need for standard framework to pre-process data, which could be applied for most used data modalities. There are multiple different neuroimaging pre-processing tools, for example FSL [108] or FreeSurfer [109] or other modalities - [110] for genetic data, which makes the pre-processing easier, but there is a lack of one tool which would do it all or integrate different software solutions for multi-modal data pre-processing.
4. Resource inefficient models

   Deep learning methods tend to be resource hungry due to deep and large neural networks. Training such models requires a lot of hardware resources and is

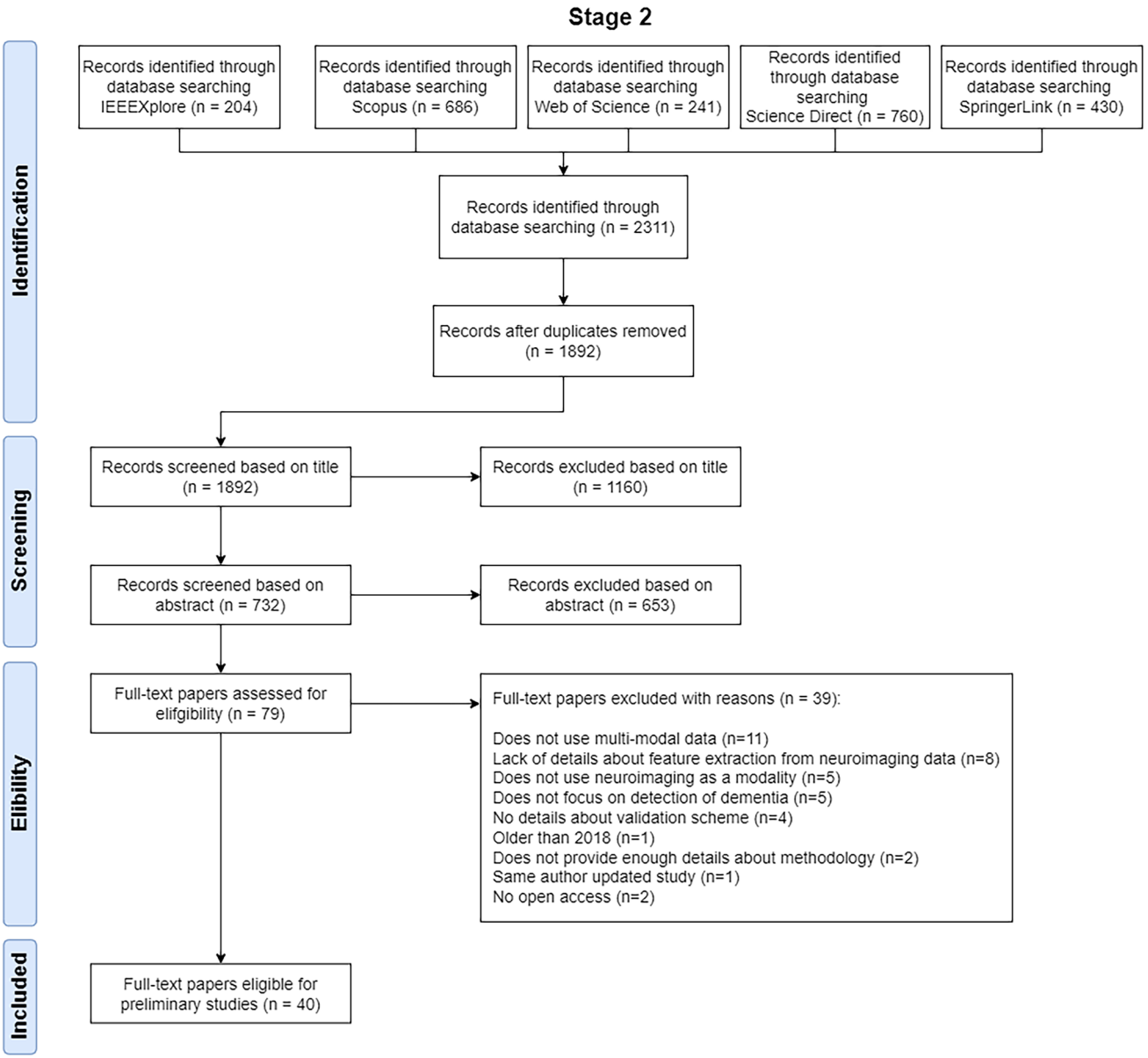

Fig. 5 PRISMA methodology chart for 2nd stage of systematic literature review

usually slow. Therefore, there is a need for resource efficient models, which could be trained with limited hardware capabilities and the training time would be comparable to machine learning methods.

5. Data imbalance

   Another problem with datasets is class imbalance. Preparing datasets is the most important step in the process of training a DL/ML model. In classification, it is crucial that each individual class has the same number of samples. If we have imbalanced dataset, there is a huge possibility that our classifier will learn more features from the dominant class and will overfit the classes with less data samples. This degrades model generalizability.

6. Lack of multi-modal methods

   In clinical practice medical personnel always try to evaluate as many different data modalities as they can before giving a diagnosis, whether it is neuropsychological evaluation with cognitive tests or evaluating protein data in blood from laboratory tests or reviewing neuroimaging data from structure/function magnetic resonance imaging. It is crucial that researchers shift towards real-world scenarios and employ methods which can work with multiple data modalities. Addi-

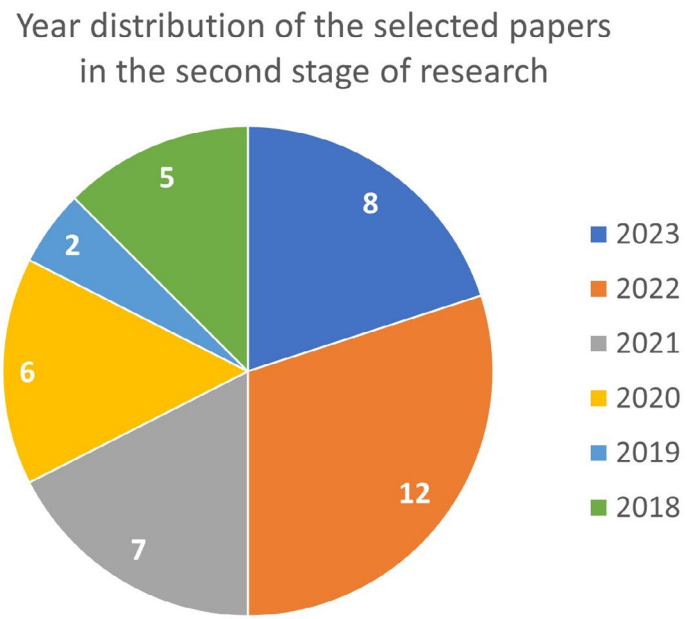

Fig. 6 Year distribution of selected papers in the second stage of the study

tional data always helps to improve not only the performance of models, but also robustness.

7. Data leakage

   Data leakage is a problem which occurs not only in this research area, but in the whole DL/ML industry. It is related to small datasets with a limited number of samples available. Leakage is a use of testing data in training. Such models do not generalize well in the domain and produce false results. Because of this it is impossible to validate whether such models are proper solutions.

8. No standard benchmark

   There is no standard evaluation benchmark in the research area of early dementia detection, which complicates the process of comparing different solutions. There are a couple of evaluation metrics used, for example, sensitivity or specificity, but no approved and standardized benchmark available. Having the standard benchmark would significantly help researchers compare their solutions with existing ones, and potentially support the need for more accurate, higher performing solutions to be higher in the rank.

9. Long time to label the data

   This problem is particularly sensitive in classification, object detection and segmentation. Preparing data for supervised learning models requires a lot of effort and human labor. In some cases, due to rush, human error or subjectiveness, prepared data can be biased,

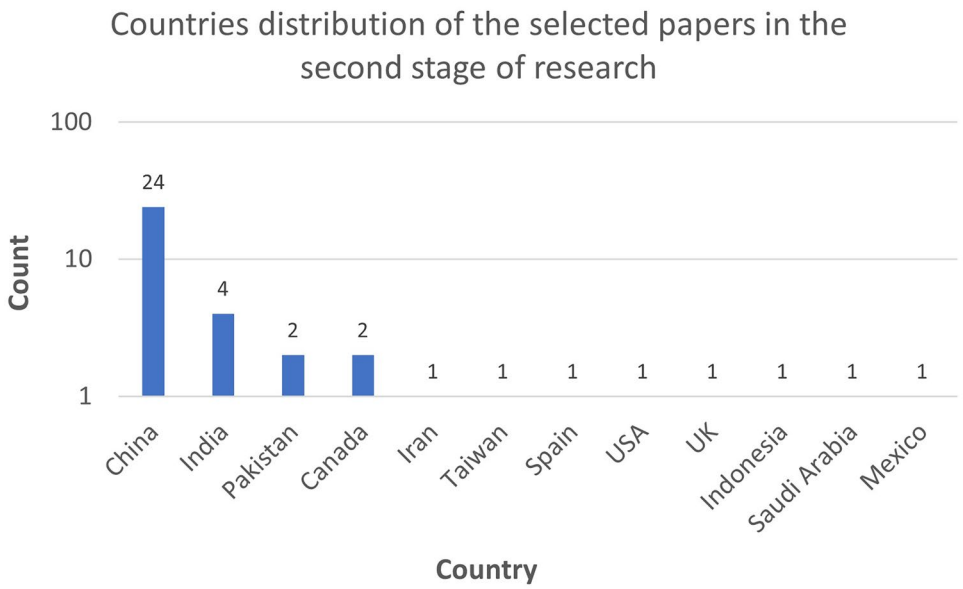

Fig. 7 Distribution of countries of selected papers in the second stage of the study

Table 8 List of names of journals from papers selected in the second stage of the study and quality assessment

Journal	Number of citations in publications	Citations median
Medical Image Analysis	[66] - 67, [67] - 65, [68] - 58	65
Computers in Biology and Medicine	[69] - 87, [70] - 81, [71] - 60, [72] - 59, [73] - 54, [74] - 48, [75] - 46, [76] - 31	56.5
Biomedical Signal Processing and Control	[77] - 71, [78] - 52, [79] - 37	52
Frontiers in Neuroinformatics	[80] - 54, [81] - 40	47
Neural Computing and Applications	[82] - 97, [83] - 45, [84] - 23	45
Neurocomputing	[85] - 52, [86] - 33	42.5
Journal of Biomedical Informatics	[87] - 55	-
IEEE Access	[88] - 48	-
Physics in Medicine and Biology	[89] - 47	-
IEEE Multimedia	[90] - 20	-
The Visual Computer	[91] - 25	-
International Journal of Imaging Systems and Technology	[92] - 38	-
IEEE Journal of Biomedical and Health Informatics	[93] - 48	-
International Journal of Intelligent Engineering and Systems	[94] - 33	-
Neuroinformatics	[95] - 34	-
International Journal of Signal and Imaging Systems Engineering	[96] - 49	-
BMC Bioinformatics	[97] - 49	-
IEEE/ACM Transactions on Computational Biology and Bioinformatics	[98] - 87	-
Journal of Medical and Biological Engineering	[99] - 52	-
Information Fusion	[100] - 51	-
Knowledge Based Systems	[101] - 51	-
Computerized Medical Imaging and Graphics	[102] - 42	-
IET Image Processing	[103] - 28	-
International Journal of Advanced Computer Science and Applications	[104] - 42	-
Mathematics	[105] - 37	-

and this degrades performance of models trained with such datasets.

10. Hard to select relevant features

    In ML scenarios human expert is responsible for feature extraction and preparation for models. For example, in neuroimaging, researchers tend to spend more time on feature engineering than to use DL methods. There are segmentation tools in the FLS and FreeSurfer packages, which allow to extract Region of Interest (ROI) data from brain scans. Then the problem occurs, how to select which regions features to use in training. There is a potential that various feature selection methos will exclude features, which may be related to the regions that are more participating in disease progression than others.

11. Lack of early dementia detection methods

    Detecting dementia early is important for patients as they can prepare for future better or even reverse some of cognitive decline effects [111]. However, majority of researchers focus on detection of dementias that already progressed into diseases like Alzheimer's or Parkinson's, because the differences of brain tissue changes are evident. However, that is not the case for early dementias, where changes are subtle and hard to detect.

12. Lack of studies in some other regions

    This issue arises due to the most common datasets such as [112] being prepared by one country initiative in this case US, but it is known, that there are some anatomical differences in the brain in different ethnic groups and races [113]. DL/ML methods are very sensitive to the differences of dataset domains. For example, if training data is used from one dataset, which has only MRIs collected from asians, but validation set will have MRIs only from americans, algorithm will perform poor due to differences between the latent spaces of the datasets.

Future research areas:

The summary of future areas that researchers mentioned is displayed in the Table 11 (only the common future areas are included and sorted by number of mentions in decreasing order).

Table 9 List of selected papers in the second stage of research

#	Reference	Year	Dementia detection type
1	Yuan et al. [98]	2021	Stable Mild Cognitive Impairment (SMCI) vs Progressive Mild Cognitive Impairment (PMCI)
2	Bi et al. [93]	2021	Early Mild Cognitive Impairment (EMCI) vs CN
3	Song et al. [88]	2020	EMCI vs CN, Late Mild Cognitive Impairment (LMCI) vs CN, EMCI vs LMCI
4	Dong et al. [73]	2022	AD vs CN, MCI vs CN, AD vs MCI
5	Jia and Lao [83]	2022	CN vs Significant Memory Concern (SMC), CN vs MCI, SMC vs MCI, SMC vs AD, MCI vs AD, CN vs AD
6	Pahuja and Prasad [76]	2022	PD vs CN
7	Kumari et al. [82]	2022	AD vs CN, MCI vs CN, AD vs MCI
8	Jin et al. [81]	2022	EMCI vs LMCI
9	Angkoso et al. [94]	2022	AD vs MCI vs CN, AD vs MCI, MCI vs CN, AD vs CN
10	Dwivedi et al. [90]	2022	CN vs AD, CN vs MCI, AD vs MCI
11	Liu et al. [66]	2022	Stable Subjective Cognitive Decline (SSDC) vs Progresive Subjective Cognitive Decline (PSDC), PMCI vs SMCI
12	Abdelaziz et al. [87]	2021	CN vs AD, CN vs SMCI, CN vs PMCI
13	Liu et al. [68]	2021	AD vs CN
14	Liu et al. [97]	2020	EMCI vs CN
15	Castellazzi et al. [80]	2020	AD vs Vascular Dementia (VD)
16	Zhu Et al. [85]	2019	AD vs CN, MCI vs CN, PMCI vs SMCI
17	Lahmiri and Shmuel [77]	2019	AD vs CN
18	Hojjati et al. [69]	2018	SMCI vs PMCI
19	Liu et al. [95]	2018	AD vs CN, PMCI vs CN, SMCI vs CN
20	Asim et al. [92]	2018	AD vs CN, AD vs MCI, MCI vs CN
21	Yang et al. [99]	2020	AD vs CN, MCI vs CN
22	Perez-Gonzalez et al. [89]	2021	MCI vs CN
23	Pahuja et al. [96]	2018	PD vs CN
24	Ye et al. [91]	2022	AD vs CN, EMCI vs LMCI
25	Divya et al. [84]	2021	CN vs AD, CN vs MCI, MCI vs AD
26	Segovia et al. [86]	2020	CN vs AD
27	Altaf et al. [78]	2018	AD vs CN, AD vs MCI, MCI vs CN, AD vs MCI vs CN
28	Zhang et al. [100]	2021	AD vs MCI, AD vs CN, MCI vs CN, AD vs MCI vs CN
29	Tu et al. [72]	2022	AD vs CN, SMCI vs PMCI
30	Yan et al. [71]	2022	AD vs CN
31	Hao et al. [67]	2020	AD vs CN, LMCI vs CN, EMCI vs LMCI
32	Kong et al. [79]	2022	AD vs CN, MCI vs CN, AD vs MCI, AD vs MCI vs CN
33	Zhang et al. [101]	2023	AD vs CN, sMCI vs pMCI
34	Leng et al. [75]	2023	AD vs CN, SMC vs CN
35	Zhang et al. [74]	2023	AD vs CN, sMCI vs pMCI
36	Zhang et al. [70]	2023	AD vs CN, MCI vs CN, AD vs MCI vs CN
37	Gao et al. [102]	2023	AD vs CN, sMCI vs pMCI
38	Chen et al. [103]	2023	AD vs CN, MCI vs CN, AD vs MCI
39	Ding et al. [104]	2023	AD vs CN, MCI vs CN
40	Ismail et al. [105]	2023	AD vs CN, MCI vs CN, AD vs MCI, AD vs MCI vs CN

Each common future area is discussed in the list below:

1. Multi-modal solutions

   Most researchers mentioned multi-modal solutions as being a future direction in the research area. The biggest advantage of multi-modal solutions seems to be the performance increase of the models. However, there is a possibility that suboptimal feature fusion techniques will be used, but in general, models with multi-modal data usually yield higher accuracy.
2. Explainability

   As mentioned in the issues, where models lack of explainability, researchers are suggesting, that in the future ML and DL models will become more transpar-

Table 10 Problems existing in research area identified from reviewed literature reviews

#	Problem	Number of mentions
1	Explainability	6
2	Overfitting small datasets	6
3	Noisy, poor-quality data	5
4	Resource inefficient models	5
5	Data imbalance	4
6	Lack of multi-modal methods	3
7	Data leakage	3
8	No standard benchmark	3
9	Long time to label the data	3
10	Hard to select relevant features	3
11	Lack of early dementia detection methods	2
12	Lack of studies in some other regions	2

ent and have an explainability component, which would allow for clinical practitioners to trust methods that can assist in diagnosis of diseases.

3. Usage of transfer learning

   With the limited training data available, there is a potential to use transfer learning technique, which allows to take a trained model on general domain, for example, object recognition and apply it in other domains by freezing majority of the weights in neural networks and retraining only the last few layers, where the decision is being made. This almost always allows models to converge faster to the global minimum of cost function.

4. Smart environments

   Specialized medical equipment, like magnetic resonance tomograph, is extremely expensive and difficult to access, which makes the researchers focus on more cost-effective solutions. One of the proposed future areas is smart environments packed with sensors, trackers and camaras, which would allow us to detect anomalies in behaviors of patients as well as monitor whether disease is progressing and affecting patients. However, such technologies applied to detection problem are limited and would still require additional medical tests to provide diagnosis of disease.

Table 11 Future research areas identified from reviewed literature reviews

#	Future research area	Number of mentions
1	Multi-modal solutions	15
2	Explainability	5
3	Usage of transfer learning	4
4	Smart environments	4
5	Data imputation with GANs	3
6	EEG in consumer environments	2

5. Data imputation with GANs

   Generative Adversarial Networks (GAN) [114] are great at learning latent space of data distribution and can generate realistic samples with features learned from training data. In most of the publicly available datasets, not every modality is available, frequently patients with MRI data do not have PET data. Therefore, there is a potential to employ these models to generate (impute) missing data values and solve the data imbalance problem.

6. EEG data collection in consumer environments

   Electroencephalograph data is another data modality used in early dementia detection. This technology collects brain signals with the help of electrodes attached to the patient's head. Typically, devices which collect the data are not portable and require medical personnel to administer the process. However, recently more portable devices were introduced, which allows patients to record EEG signals at home [115]. This is potentially a cost-effective solution to detect early dementia, which requires minimal medical personnel attention.

5.2 What types of dementia detection researchers focus their research on?

We have collected detection type data from the second stage in the research. In Table 12 the data with references can be seen.

More than two thirds of the papers (70%) focus on binary Alzheimer's disease detection where the other class is cognitive normal (no symptoms and cognitive decline), 42.5% of papers use binary classification but instead of Alzheimer's, they try to detect MCI. Even smaller fraction of papers (30%) focus on binary classification between Alzheimer's disease and MCI. What is interesting is that only 15% of papers used multi-class classification. Based on the data in sub-section 5.3, multi-class classification is a challenging problem and requires more focus from researchers.

5.3 What are state-of-the-art methods in detection of early dementia in the field of AI/ML/DL?

All the collected AI/ML/DL methods comparison is displayed in Table 13, where results are grouped by dementia detection type. We compare methods by reported accuracy, sensitivity, and specificity, where results are sorted by accuracy in each group. In binary AD versus CN, there are a couple reported solutions, which reach 100 percent accuracy, sensitivity, and specificity. Both solutions are at the top of the table, and both use machine learning methods.

In the group of MCI vs CN, we have highest accuracy reaching 94% reported in [90] article. Researcher of this

Table 12 Dementia detection types researchers investigate

Dementia detection type	References	#
AD vs CN	[67, 68, 70–75, 77–79, 82–87, 90–92, 94, 95, 99–105]	70% (28/40)
MCI vs CN	[70, 73, 78, 79, 82–85, 89, 90, 92, 94, 99, 100, 103–105]	42.5% (17/40)
AD vs MCI	[73, 78, 79, 82–84, 90, 92, 94, 100, 103, 105]	30% (12/40)
SMCI vs PMCI	[66, 69, 72, 74, 85, 98, 101, 102]	20% (8/40)
AD vs MCI vs CN	[70, 78, 79, 94, 100, 105]	15% (6/40)
EMCI vs LMCI	[67, 81, 88, 91]	10% (4/40)
EMCI vs CN	[88, 93, 97]	7.5% (3/40)
LMCI vs CN	[67, 88]	5% (2/40)
PD vs CN	[76, 96]	5% (2/40)
CN vs PMCI	[87, 95]	5% (2/40)
CN vs SMCI	[87, 95]	5% (2/40)
SMC vs CN	[75, 83]	5% (2/40)
SMC vs MCI	[83]	2.5% (1/40)
SMC vs AD	[83]	2.5% (1/40)
SSCD vs PSCD	[66]	2.5% (1/40)
AD vs VD	[80]	2.5% (1/40)
BA (Brain-Atrophy) vs CN	[94]	2.5% (1/40)

article used Support Vector Machine (SVM) and ResNet methods to produce quite good results in this classification group. The same author in the same paper also reports the highest accuracy achieved in the AD vs MCI classification group using the same methods.

In the SMCI vs PMCI classification group popular ML method SVM [69] is in the lead achieving 97 percent accuracy.

In EMCI vs LMCI classification group, DL method GAN is in the lead achieving 87.5 percent accuracy. Comparing this group to the previous ones, it looks like it's a more challenging task and still requires research efforts to reach better results.

In the only multi-class classification task AD vs MCI vs CN both ML solutions have better results. For example, SVM based solution - [105] reaching 92.3 accuracy. DL solutions fall short, with 3D CNN [79] only reaching 87.5 percent accuracy. Only 5 papers out of 40 used multi-class classification task, which means that this task is very challenging and it's hard to reach better results. This problem requires more attention from researchers as well as EMCI vs CN and LMCI vs CN, which also do not show excellent results (EMCI vs CN, highest accuracy 86.2% by ML methods [93] and LMCI vs CN by DL method [88] with 88.7% accuracy.

PD vs CN classification seems to be showing perfect results with highest achieved 98.1% accuracy by ML methods [96], although the dataset is relatively small only 82 samples for each class, which could be that models did overfit during training.

In PMCI vs CN task we can see that DL methods CNN are dominant and reach highest 97.3% accuracy. In SMCI vs CN we can see that DL method also reach highest accuracy 93.1% and in both task the same paper reports these results [87].

In SMC vs CN task we see the highest accuracy reaching 91.3%. Other classification categories have only one article sample related, where SMC vs MCI 94.4%, SMC vs AD 94.4%, SSCD vs PSCD 72.1%, AD vs VD 85.2%, BA vs CN 94.2% accuracies.

Most of the selected articles (75.29%) use ML methods to provide a diagnosis, probably because of relatively small datasets used in the research and 24.71% use DL methods. Results are depicted in Fig. 8.

Grouping the data by the method type yields more interesting results. Figure 9 shows the distribution of machine learning methods used by the researchers. The most popular method is SVM and the second most popular method is RF, which are the most common and basic ML classifiers in the field.

Figure 10 shows the most popular deep learning methods used by researchers. The most popular methods are GANs, that are very good at data generation, 3D 3D CNN and CNN, that are very common choice when dealing with high-dimensional data. Last most common network type is Graph Convolutional Network (GCN) - graph based networks with convolution layers in nodes.

5.4 What modalities are being used in combination with neuroimaging to detect early dementia?

We have collected data from both stages in the study of all modalities that were mentioned, which are being used in one way or another to detect different dementias. The high-level overview can be seen in Fig. 11. There are four main

Table 13 Comparison of AI/ML/DL methods used by the researchers

Detection type	AI/ML/DL methods	Dataset	ACC	SEN	SPE	Ref
AD vs CN	Hyperparameter Tuning Random Forest Ensemble Classifier (HPT-RFE), SVM, DBN (Deep Belief Network)	ADNI (18 AD, 19 CN)	100%	100%	100%	[82]
	Latent Dirichlet Allocation (LDA), K-Neasers Neighbours (KNN), Naïve Bayes, SVM	ADNI (35 AD, 35 CN)	100%	100%	100%	[77]
	Ensemble of CNN	ADNI (45 AD, 56 CN)	99%	-	-	[104]
	CNN	ADNI (186 AD, 226 CN)	98.2%	78.7%	98.7%	[87]
	Fully Connected Network (FCN) with Pyramid Squeeze (PS) attention, Multi Layer Perceptron (MLP)	ADNI (309 AD, 241 CN)	98.1%	96.8%	95.9%	[71]
	3D CNN	ADNI (45 AD, 57 CN)	97.9%	98.6%	98.4%	[103]
	SVM, KNN, Decision Tree (DT), Ensemble with AdaBoost	ADNI (92 AD, 90 CN)	97.8%	100%	95.6%	[78]
	3D CNN	ADNI (184 AD, 254 CN)	97.6%	97.2%	98.2%	[75]
	ResNet, Robust Energy-based Least Squares Twin Support Vector Machine (RELS-TSVM)	ADNI (100 AD, 100 CN)	97%	97%	97%	[90]
	Random Forest Ensemble (RFE), Genetic Algorithm (GA), Logistic Regression (LR), SVM, Random Forest (RF), Extreme Gradient Boosting (XGB)	ADNI (171 AD, 347 CN)	96.8%	92.8%	98.7%	[84]
	DT, AdaBoost, Gradient Boosting Classifier (GBC), Random Forest Classifier (RFC), LDA	ADNI (100 AD, 100 CN)	96.7%	96.7%	100%	[94]
	Graph Neural Network (GNN)	ADNI (215 AD, 246 CN)	96.68%	99.1%	94.4%	[74]
	CNN, Artifial Neural Network (ANN)	ADNI (78 AD, 100 CN)	96%	97%	93%	[72]
	3D ResNet	ADNI (215AD, 246CN)	94.6%	92.1%	97.2%	[101]
	Hybrid 3D CNN Transformer, GAN	ADNI (352 AD, 427 CN)	94.4%	93%	95.5%	[102]
	SVM, MOGOA	ADNI (511 AD, 535 CN)	94.4%	95%	94%	[105]
	Principal Component Analysis (PCA), Radial Basis Function (RBF)-SVM	ADNI (100 AD, 100 CN)	94%	95%	93%	[92]
	Multi Kernel Learning (MKL)-SVM	ADNI (38 AD, 40 CN)	94%	-	-	[100]
	MKL-SVM, RF, Multi Task Learning (MTL)	ADNI (201 AD, 263 CN)	93.7%	95.1%	91.8%	[67]
	3D CNN	ADNI (111 AD, 130 CN)	93.2%	91.4%	95.4%	[79]
AD vs CN	3D CNN	ADNI (93 AD, 100 CN)	93.2%	92.5%	93.9%	[95]
	GAN with attention, Partial Multi-view Clustering (PVC), Unified Embeding Alignment Framework (UEAF), Generative Adversarial Imputation Network (GAIN)	ADNI (160 AD, 210 CN)	93.2%	92.9%	93.4%	[91]
	RF, MOGOA	ADNI (511 AD, 535 CN)	92.8%	94.2%	90.6 %	[105]
	3DMR-PCANet, 3DResNet-10, SVM	ADNI (34 AD, 50 CN)	92%	100%	80%	[83]
	3D CNN	ADNI (129 AD, 110 CN)	91%	91%	91%	[70]
	SVM	ADNI (52 AD, 52 CN)	90.7%	90.1%	88.1%	[73]
	Hybrid 3D CNN Transformer, GNN	OASIS-3 (174 AD, 171 CN)	88.4%	84.6 %	92.3 %	[102]
	Sample weighting based Multi-modal Rank Minimization (SPMRM), MTL, MKL	ADNI (160 AD, 211 CN)	88%	94.3%	80%	[85]
	Softmax classifier, Multi-Objective Grasshopper Optimization Algorithm (MOGOA)	ADNI (511 AD, 535 CN)	87.7%	90%	85.6%	[105]
	SVM	ADNI (20 AD, 32 CN)	86%	83.3%	90.3%	[99]

Table 13 (continued)

Detection type	AI/ML/DL methods	Dataset	ACC	SEN	SPE	Ref
	Auto-encoder	ADNI (185 AD, 90 CN)	83.6%	-	-	[68]
	Partial Least Square (PLS), SVM	Closed source (21 AD, 22 CN)	82.1%	80.9%	89%	[86]
MCI vs CN	ResNet, RELS-TSVM	ADNI (100 MCI, 100 CN)	94%	97%	91%	[90]
	SVM, MOGOA	ADNI (571 MCI, 535 CN)	93.2%	96%	89.2%	[105]
	3DMR-PCANet, 3DResNet-10, SVM	ADNI (50 CN, 35 MCI)	92%	93.3%	90%	[83]
	SVM, KNN, DT, Ensemble with AdaBoost	ADNI (105 MCI, 90 CN)	91.8%	90%	93.3%	[78]
	Softmax classifier, MOGOA	ADNI (571 MCI, 535 CN)	91.5%	93.6%	89.4%	[105]
	RF, Extra Trees Classifier (ETC), SVM, ANN, Gaussian Process (GP)-RBF	ADNI (30 MCI, 30 CN) + Closed Source (11 MCI, 12 CN)	91.3%	-	-	[89]
	HPT-RFE, SVM, DBN	ADNI (65 MCI, 19 CN)	91%	100%	60%	[82]
	RFE, GA, LR, SVM, RF, XGB	ADNI (558 MCI, 347 CN)	89.3%	95.1%	80%	[84]
	RF	ADNI (571 MCI, 535 CN)	89%	91%	88%	[105]
	3D CNN	ADNI (125 MCI, 57 CN)	87.8%	92.4%	92.6%	[103]
	DT, AdaBoost, GBC, RFC, LDA	ADNI (100 MCI, 100 CN)	86.7%	85%	88.3%	[94]
	3D CNN	ADNI (129 MCI, 130 CN)	86.5%	94.3%	81.6%	[79]
	SPMRM, MTL, MKL	ADNI (542 MCI, 211 CN)	84.1%	94.3%	55.2%	[85]
	Ensemble of CNN	ADNI (123 MCI, 56 CN)	81.6%	-	-	[104]
	MKL-SVM	ADNI (42 MCI, 40 CN)	81%	-	-	[100]
	SVM	ADNI (27 MCI, 32 CN)	80.2%	70.4%	89.9%	[99]
	PCA, RBF-SVM	ADNI (100 MCI, 100 CN)	76.5%	78%	75%	[92]
	SVM	ADNI (52 MCI, 52 CN)	74.7%	73.6%	74.4%	[73]
	3D CNN	ADNI (125 MCI, 110 CN)	71.2%	74.6%	67.9%	[70]
AD vs MCI	ResNet, RELS-TSVM	ADNI (100 AD, 100 MCI)	97.5%	96%	99%	[90]
	HPT-RFE, SVM, DBN	ADNI (18 AD, 65 MCI)	95%	100%	80%	[82]
	3DMR-PCANet, 3DResNet-10, SVM	ADNI (35 MCI, 34 AD)	95%	90%	100%	[83]
	3D CNN	ADNI (45 AD, 125 MCI)	92.8%	97.2%	96.2%	[103]
	RFE, GA, LR, SVM, RF, XGB	ADNI (558 MCI, 171 AD)	91.4%	96.5%	98.1%	[84]
	SVM, MOGOA	ADNI (511 AD, 535 CN)	90%	89.2%	93.3%	[105]
	Softmax classifier, MOGOA	ADNI (511 AD, 535 CN)	89.4%	91%	88%	[105]
	MKL-SVM	ADNI (38 AD, 42 MCI)	89%	-	-	[100]
	3D CNN	ADNI (111 AD, 129 MCI)	85.6%	81.2%	95.5%	[79]
	SVM, KNN, DT, Ensemble with AdaBoost	ADNI (92 AD, 105 MCI)	85.3%	94.2%	75%	[78]
	DT, AdaBoost, GBC, RFC, LDA	ADNI (100 AD, 100 MCI)	84.2%	83.3%	85%	[94]
	RF, MOGOA	ADNI (511 AD, 535 CN)	83%	85%	78.2%	[105]
	PCA, RBF-SVM	ADNI (100AD, 100 MCI)	75.5%	71%	80%	[92]
	SVM	ADNI (52 AD, 52 MCI)	75.5%	77.5%	78.5%	[73]
SMCI vs PMCI	SVM	ADNI (62 SMCI, 18 PMCI)	97%	100%	95%	[69]
	CNN, ANN	ADNI (117 SMCI, 53 PMCI)	87%	91%	87%	[72]
	SVM, RF	ADNI (167 SMCI, 24 PMCI)	85.5%	76.2%	88.7%	[98]
	SPMRM, MTL, MKL	ADNI (56 SMCI, 43 PMCI)	78.8%	74.4%	83.9%	[85]
	GAN with Representation Learning	ADNI (234 CN, 629 MCI)	78%	-	-	[66]
	Hybrid 3D CNN Transformer, GAN	ADNI (342 sMCI, 234 pMCI)	77.8%	75.4%	79.6%	[102]
	3D ResNet	ADNI (238 sMCI, 151 pMCI)	77.1%	68.1%	81.8%	[101]
	GNN	ADNI (211 sMCI, 120 pMCI)	77%	51.9%	89.37%	[74]
EMCI vs LMCI	GAN with attention, PVC, UEAF, GAIN	ADNI (272 EMCI, 187 LMCI)	87.5%	93.5%	80.1%	[91]
	Graph Convolutional Network (GCN)	ADNI (40 LMCI, 77 EMCI)	86.3%	83.1%	92.5%	[88]
	GAN	ADNI (124 EMCI, 133 LMCI)	85.2%	79.7%	91.1%	[81]

Table 13 (continued)

Detection type	AI/ML/DL methods	Dataset	ACC	SEN	SPE	Ref
	MKL-SVM, RF, MTL	ADNI (272 EMCI, 187 LMCI)	73.8%	90.5%	48.5%	[67]
AD vs MCI vs CN	SVM, MOGOA	(ADNI 511 AD, 571 MCI, 535 CN)	92.3%	-	-	[105]
	Softmax classifier, MOGOA	(ADNI 511 AD, 571 MCI, 535 CN)	90.1%	-	-	[105]
	RF, MOGOA	(ADNI 511 AD, 571 MCI, 535 CN)	89%	-	-	[105]
	MKL-SVM	ADNI (38 AD, 42 MCI, 40 CN)	88%	-	-	[100]
	3D CNN	ADNI (111 AD, 129 MCI, 130 CN)	87.5%	-	-	[79]
	DT, AdaBoost, GBC, RFC, LDA	ADNI (100 AD, 100 MCI, 100 CN)	86.7%	-	-	[94]
	SVM, KNN, DT, Ensemble with AdaBoost	ADNI (90 CN, 105 MCI, 92 AD)	79.8%	-	-	[78]
	3D CNN	ADNI (129 AD, 125 MCI, 110 CN)	64%	-	-	[70]
EMCI vs CN	Genetic-Evolutionary Random Forest (GERF), SVM, RF	ADNI (37 EMCI, 36 CN)	86.2%	-	-	[93]
	Multi-Task Feature Selection (MTFS)-gLASSO, GCN	ADNI (105 EMCI, 105 CN)	84.1%	86.5%	81.3%	[97]
	GCN	ADNI (77 EMCI, 97 CN)	82.7%	77.6%	89%	[88]
LMCI vs CN	GCN	ADNI (40 LMCI, 97 CN)	88.7%	83.5%	97.5%	[88]
	MKL-SVM, RF, MTL	ADNI (187 MCI, 211 CN)	78.4%	85.8%	70.1%	[67]
PD vs CN	GA, Self-adaptive Resource Allocation Network (SRAN), Extreme Learning Machine (ELM), SVM	PPMI (82 PD, 82 CN)	98.1%	98.7%	97.7%	[96]
	Softmax classifier, CNN	PPMI (59 CN, 73 PD)	90.4%	86%	93.5%	[76]
PMCI vs CN	CNN	ADNI (166 SMCI, 226 CN)	97.3%	97.8%	96.7%	[87]
	3D CNN	ADNI (76 PMCI, 100 CN)	82.9%	81%	84.3%	[95]
SMCI vs CN	CNN	ADNI (225 SMCI, 226 CN)	93.1%	92.6%	93.5%	[87]
	3D CNN	ADNI (128 PMCI, 100 CN)	64%	63%	67.3%	[95]
SMC vs CN	3DMR-PCANet, 3DResNet-10, SVM	ADNI (50 CN, 26 SMC)	91.3%	100%	75%	[83]
	3D CNN	ADNI (98 SMC, 254 CN)	81.6%	72.1%	85.2%	[75]
SMC vs MCI	3DMR-PCANet, 3DResNet-10, SVM	ADNI (26 SMC, 35 MCI)	94.4%	100%	87.5%	[83]
SMC vs AD	3DMR-PCANet, 3DResNet-10, SVM	ADNI (26 SMC, 34 AD)	94.4%	87.5%	100%	[83]
SSCD vs PSCD	GAN with Representation Learning	CLAS (76 SSCD, 52 PSCD)	72.1%	75%	72.1%	[66]
AD vs VD	ANN, SVM, daptive Neuro Fuzzy Inference System (ANFIS)	Closed Source (22 AD, 27 VD, 15 mixed)	85.2%	82%	88.5%	[80]
Brain Atrophy (BA) vs CN	DT, AdaBoost, GBC, RFC, LDA	Closed Source (20 BA, 20 CN)	94.2%	-	-	[94]

ACC Accuracy, *SEN* Sensitivity, *SPE* Specificity

categories of modalities: neuroimaging, neuropsychological/psychiatric evaluation, clinical data, and others. The neuroimaging group of modalities contains MRI, DWI, Single-Photon Emission Computed Tomography (SPECT), and PET. There are a few other subtypes of MRI, like T1-weighted (T1w), T2-weighted (T2w), functional MRI (fMRI), and PET, like FDG-PET, Flute/PiB-PET, Amyloid PET.

The neuropsychological/psychiatric evaluation group of modalities consists of all different identified cognitive function evaluation tests used in clinical practice in this study. Another group of modalities is biomarkers. It includes such modalities as blood, blood plasma or Cerebrospinal Fluid (CSF), saliva and others. All modalities in the second stage of research collected can be seen in Fig. 12.

Most popular modalities are T1w MRI (35 of 40 papers used), FDG-PET (22 of 40 papers used), MMSE (7 of 40 papers used), fMRI (6 of 40 papers used), genetic data (6 of 40 papers used), demogpahic data (6 of 40 papers used).

All combinations of modalities that papers used are displayed in Table 14. From the table it is clear that the most popular combination of modalities is two neuroimaging modalities (T1w MRI and FDG-PET), where one third of the articles collected used this combination.

Grouping all combinations by the category of modality would give a better understanding of what combinations researchers use in multi-modal solutions. Grouped data is provided in Table 15. Many papers (55%) used only neuroimaging modalities in proposed early dementia detection methods. It can be observed that the other combinations of groups of modalities with neuroimaging are cognitive data (15%) or genetic data (10%). Less popular are biomarkers and demographic data.

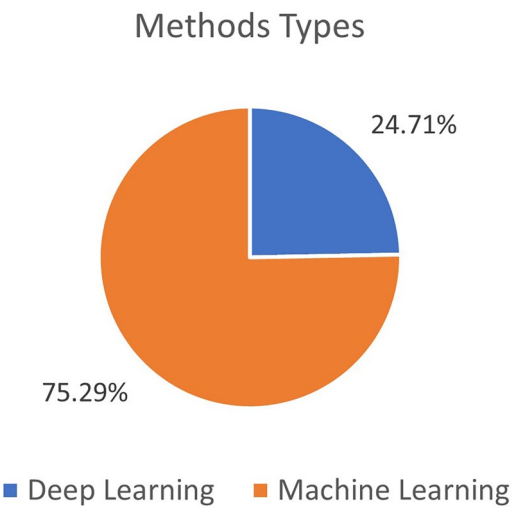

Fig. 8 Types of AI methods used by researchers

Grouping combinations of modalities by the count of modalities used shows in Table 16 an interesting observation, that most (62.5%) researchers only used two modalities in their research to detect dementia and only minority used more than two modalities. There is even one case [82], where 9 modalities were used (3 neuroimaging, 5 cognitive tests and demographic data).

5.5 How is model performance evaluated and validated?

5.5.1 Evaluation

We have collected data from research articles about different model performance evaluation metrics used. We can see the results in Table 17.

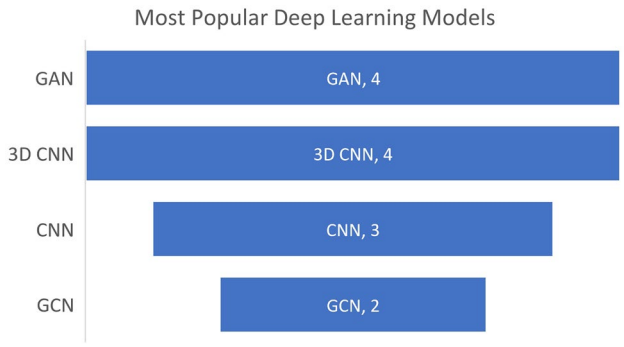

Fig. 10 Popular deep learning methods used by researchers

We can see that all papers used accuracy, most used specificity (82.5%) and sensitivity (82.5%), almost half used area under curve (45%) metrics to evaluate developed models against others. There are other metrics not so frequently measured like: F1 score (17.5%) and precision (17.5%), as well as rarely used Matthew's correlation coefficient (2.5%), negative predictive value (2.5%) and geometric mean of sensitivity and specificity (2.5%).

Below is the list of metrics, where we provide details about how the metric is calculated and what it allows to evaluate:

1. Accuracy

   In binary classification accuracy is describing how many samples model classified correctly both true negative and true positive samples from the whole population. True positive is a sample, which had a positive value associated with it and model correctly assigned the class it belongs to. For example, patient has dementia and model correctly detects it. True negative is a sample, which had a negative value associated with it and model again correctly assigns the class to it. For exam-

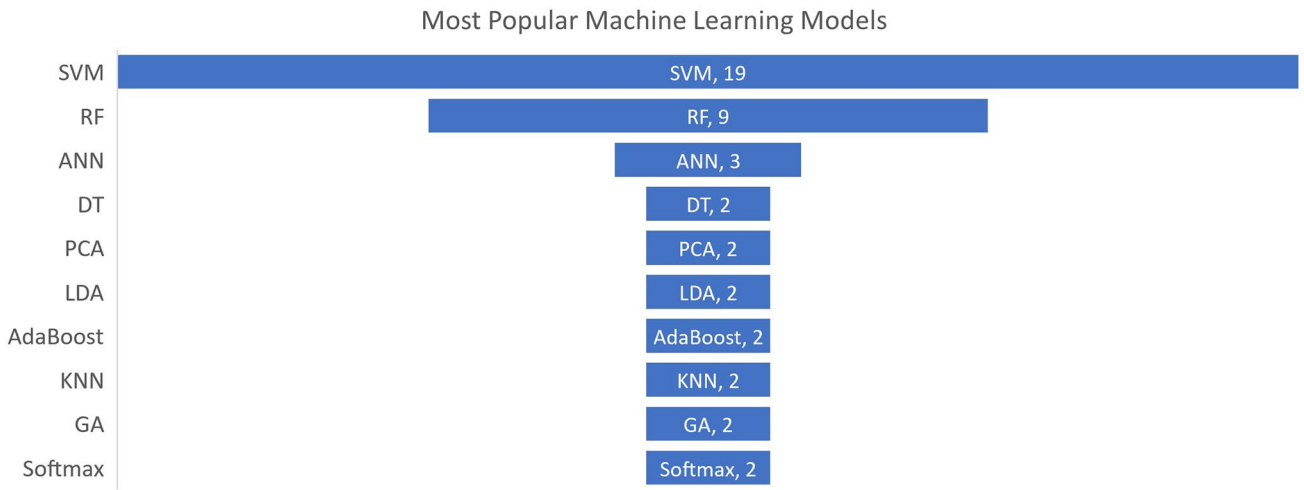

Fig. 9 Popular machine learning methods used by researchers

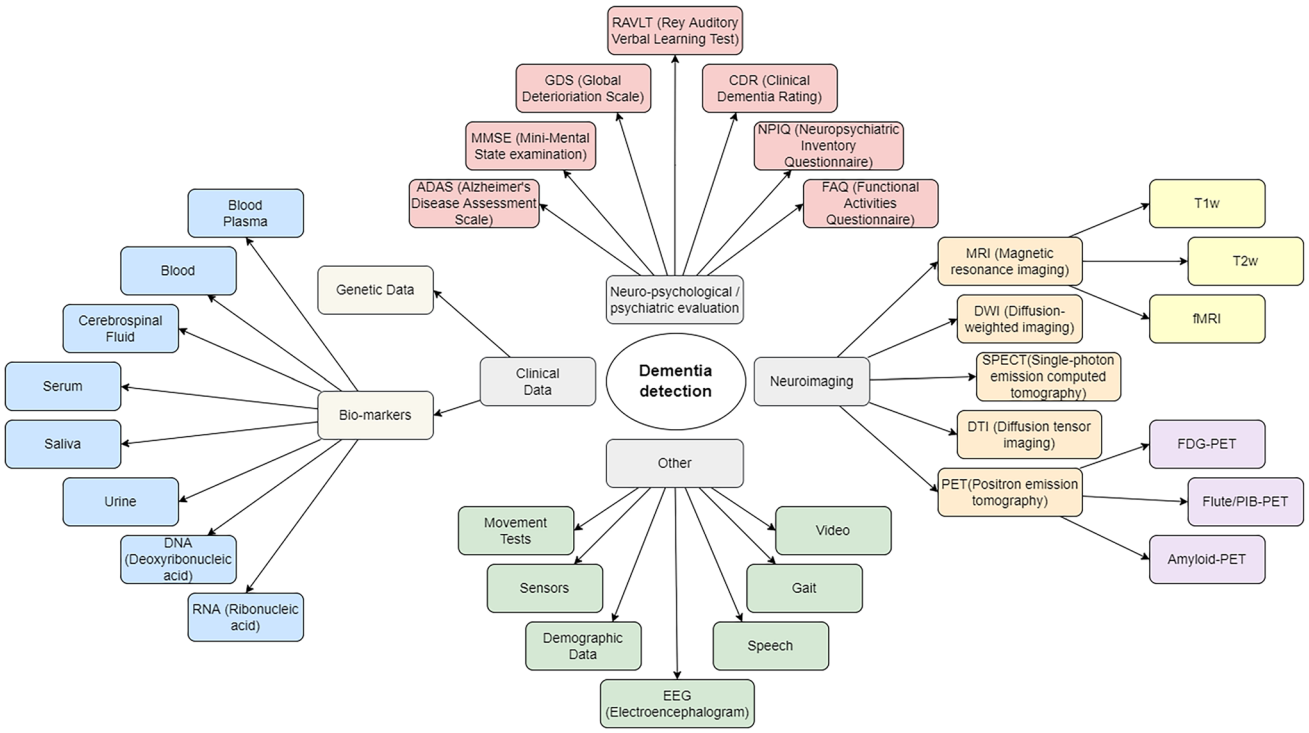

Fig. 11 All in this study identified modalities that are being used in dementia detection

ple, patient does not have dementia and models does not detect it. There are false positive and false negative samples, where false positive is when model incorrectly assigns a negative class to a positive sample (patient has dementia, but model failed to detect it) and false negative, where vice versa (patient does not have dementia, but model detected it). Therefore, accuracy can be expressed as an Eqs. (1) and (2) below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP - true positive, TN - true negative, FP - false positive and FN - false negative. In Multi-class classification accuracy is just a rate of how many classifications were correct from the overall and can be expressed as an equation below:

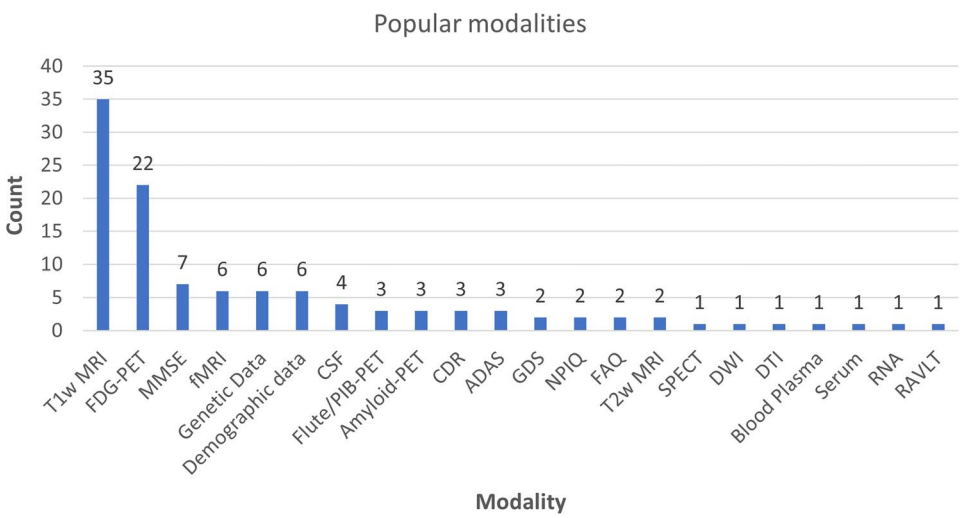

Fig. 12 Modalities collected in second stage of research

Table 14 Combinations of modalities researchers used

Combination of modalities	References	#
T1w MRI + FDG-PET	[66–68, 73, 75, 79, 81, 90, 92, 95, 100, 101, 105]	32.5% (13/40)
T1w MRI + fMRI	[69, 83, 97]	7.5% (3/40)
T1w MRI + Genetic	[98]	2.5% (1/40)
fMRI + Genetic	[93]	2.5% (1/40)
fMRI + Demographic	[88]	2.5% (1/40)
T1w MRI + SPECT + CSF	[76]	2.5% (1/40)
T1w MRI + FDG-PET + Flute/PIB-PET + MMSE + GDS + CDR + FAQ + NPIQ + Demographic	[82]	2.5% (1/40)
T1w MRI + FDG-PET + Genetic + CDR + ADAS + RAVLT + Demographic	[103]	2.5% (1/40)
T1w MRI + T2w MRI	[94]	2.5% (1/40)
T1w MRI + T2w MRI + FDG-PET	[102]	2.5% (1/40)
T1w MRI + FDG-PET + MMSE	[74]	2.5% (1/40)
T1w MRI + FDG-PET + CSF	[70]	2.5% (1/40)
T1w MRI + FDG-PET + Genetic	[87]	2.5% (1/40)
fMRI + DTI	[80]	2.5% (1/40)
T1w MRI + FDG-PET + Flute/PIB-PET + Amyloid PET + Genetic	[85]	2.5% (1/40)
T1w + ADAS	[77]	2.5% (1/40)
T1w + MMSE + ADAS	[104]	
FDG-PET + Flute/PIB-PET	[99]	2.5% (1/40)
T1w MRI + DWI + MMSE + CDR	[89]	2.5% (1/40)
T1w MRI + Plasma + Serum + CSF + RNA	[96]	2.5% (1/40)
T1w MRI + FDG-PET + Amyloid PET	[91]	2.5% (1/40)
T1w MRI + MMSE	[84]	2.5% (1/40)
T1w MRI + Amyloid PET	[86]	2.5% (1/40)
T1w MRI + FAQ + NPIQ + GDS	[78]	2.5% (1/40)
T1w MRI + Demographic + Genetic + MMSE	[72]	2.5% (1/40)
T1w MRI + Demographic + MMSE	[71]	2.5% (1/40)

$$Accuracy = \frac{\#\ of\ correct\ classifications}{\#\ of\ all\ classifications} \tag{2}$$

2. Sensitivity and specificity

Sensitivity and specificity are frequently calculated, when it is important that models are reliable in terms of decisions they make. Sensitivity represents the true positive rate, which is a number of true positive samples divided by the whole population, where specificity is a true negative rate, which is just a number of true negative samples divided by the whole population. Equation for sensitivity (3) and specificity (4) are shown below:

$$Sensitivity(TPR) = \frac{TP}{TP + FN} \tag{3}$$

Table 15 Combinations of category groups of modalities

Combination	References	#
Neuroimaging	[66–69, 73, 75, 79–81, 83, 86, 90–92, 94, 95, 97, 99–102, 105]	55% (22/40)
Neuroimaging + Cognitive tests	[74, 77, 78, 84, 89, 104]	15% (6/40)
Neuroimaging + Genetic	[85, 87, 93, 98]	10% (4/40)
Neuroimaging + Biomarkers	[70, 76, 96]	7.5% (3/40)
Neuroimaging + Demographic + Cognitive tests	[71, 74, 82]	7.5% (3/40)
Neuroimaging + Demographic + Genetic + Cognitive tests	[72, 103]	5% (2/40)
Neuroimaging + Demographic	[88]	2.5% (1/40)

Table 16 Count of modalities researchers used

Count of modalities used	References	#
2 modalities	[66–69, 73, 75, 77, 79–81, 83, 84, 86, 88, 90, 92–95, 97–101, 105]	62.5% (25/40)
3 modalities	[70, 71, 74, 76, 87, 91, 102, 104]	20% (8/40)
4 modalities	[72, 74, 78, 89]	10% (4/40)
5 modalities	[85, 96]	5% (2/40)
7 modalities	[103]	2.5% (1/40)
9 modalities	[82]	2.5% (1/40)

$$Specificity(TNR) = \frac{TN}{TN + FP} \tag{4}$$

3. Negative Predictive Value and Precision (Positive Prediction Value)

   Negative and Positive Predictive Value (Precision) metrics are very similar to Sensitivity and Specificity. The only difference is that these metrics calculate probabilities whether model decision positive or negative is really positive or negative. Equation for Negative Predictive Value (5) and Positive Predictive Value (6) are provided below:

$$Negative\ Predictive\ Value = \frac{TN}{TN + FN} \tag{5}$$

$$Positive\ Predictive\ Value = \frac{TP}{TP + FP} \tag{6}$$

4. Area Under Curve

   In classification, Area Under Curve (AUC) is frequently used to determine the performance of model and it is calculated as an integral of Receiver Operating Characteristic Curve (ROC) and x axis. ROC curve represents how well model classifies data at each sensitivity and specificity rate point. AUC allows to aggregate the performance values. Therefore, it provides an insight whether model classifies random value more positively or negatively. The value of AUC equal to 1 means the model is 100% correct. Equation (7) for AUC is represented below:

$$AUC\ of\ ROC = \int_0^1 Sensitivity(Specificity^{-1}(x))dx \tag{7}$$

5. F1 Score

   The F1 metric is a harmonic mean of positive prediction value (precision) and sensitivity (also known as recall). It is a common performance metric to capture how well model can detect true positive cases and be accurate with the decision. The Eq. (8) for F1 score is shown below:

$$F1 = 2 * \frac{PPV * Sensitivity}{PPV + Sensitivity} \tag{8}$$

   where PPV - positive predictive value.

6. Geometric Mean of Sensitivity and Specificity

   The geometric mean is another metric sometimes used by researchers. It is defined as a square root of the product of all values. In this sense, geometric mean finds the central tendency of sensitivity and specificity or the "compromise" of these two metrics, which can be another performance indicator.

7. Matthew's correlation coefficient

   Matthew's correlation coefficient is an extended version of F1 score, which also captures the true negative rate in the calculations. The equation for the metric (9) is shown below:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TN + FN)(TN + FP)(TN + FN)}} \tag{9}$$

5.5.2 Validation

In terms of validation of model performance, we have also collected data in the second stage of study from collected papers. Results are provided in Table 18.

Table 17 Evaluation metrics used in articles

Evaluation metric	References	#
Accuracy	[66–104]	100% (40/40)
Sensitivity	[66, 67, 69–88, 90–92, 94–99, 101–103]	82.5% (33/40)
Specificity	[66, 67, 69–88, 90–92, 94–99, 101–103]	82.5% (33/40)
Area Under Curve	[66, 67, 69, 70, 72–75, 81, 83, 85, 88, 89, 91, 95, 98, 101, 102]	45% (18/40)
F1 Score	[66, 70, 71, 76, 83, 87, 94]	17.5% (7/40)
Precision (Positive Predictive Value)	[69, 80, 87, 94, 98, 100, 103]	17.5% (7/40)
Geometric Mean of Sensitivity and Specificity	[76]	2.5% (1/40)
Negative Predictive Value	[80]	2.5% (1/40)
Matthew's correlation coefficient	[71]	2.5% (1/40)

Table 18 Validation types used by researchers

Validation type			References	#
Cross validation	Leave-One-Out		[88, 92]	75% (30/40)
	k-fold	10-fold	[67, 68, 72, 73, 77, 79, 80, 84–87, 90, 91, 95, 99, 105]	
		9-fold	[69]	
		5-fold	[70, 71, 74, 75, 78, 94, 97, 100, 102–104]	
Random Split			[76, 81–83, 93, 96, 101]	17.5% (7/40)
Separate test set			[66, 89, 98]	7.5% (3/40)

Most of the papers (75%) use cross-validation to validate model results, 17.5% of papers used random subsampling of dataset and only 7.5% used separate test set to validate the results. K-fold validation type is used to reduce bias due to available small datasets. Data in public datasets is very limited, therefore there is a need to try to overcome this issue by introducing cross-validation. The most popular cross-validation type is k-fold. This validation type can be understood essentially as partitioning original dataset into k number of partitions where we use only one of the partitions for validation and remaining for training. Validation is then repeated k times, so we get k models and their validation results. In the end, we can average the k models results and get an overall estimate of performance of the developed model.

While random subsampling is technique that exists, it is not ideal to use it because of risk to introduce bias. Selecting samples from original dataset for training and validation may be biased simply because of randomness of the subsampling method. Some samples could be more closely related in the latent feature space than others and if these samples get used in both training and validation, the generalizability of model decreases. The best possible validation is to use unseen data or separate dataset, which has similar distribution to the one used in training. This way, we can prevent bias introduction and see how well the model generalizes in the trained domain.

5.6 What datasets researchers used in their studies?

In the collected data from the second stage, we have also noted the datasets that researchers used. Collected data is provided in Table 19. We can see that 87.5% of articles used Alzheimer's Disease Neuroimaging Initiative (ADNI), 10% used closed source dataset which was provided by the third party like a hospital, which accepted to participate in research. Only 2 articles used Parkinson's Progression Markers Initiative (PPMI) dataset. Only 1 article used Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL), Chinese Longitudinal Aging Study (CLAS) and Open Access Series of Imaging Studies (OASIS) datasets.

Table 19 Different datasets used by the researchers

Dataset	References	#
ADNI	[66–75, 77–79, 81–85, 87–95, 97–104]	87.5% (35/40)
Closed source	[80, 86, 89, 94]	10% (4/40)
PPMI	[76, 96]	5% (2/40)
AIBL	[66]	2.5% (1/40)
CLAS	[66]	2.5% (1/40)
OASIS	[102]	2.5% (1/40)

1. ADNI
   Alzheimer's Disease Neuroimaging Initiative (ADNI) [112] dataset was first introduced in 2004. The dataset focuses on Alzheimer's disease and its prodromal stage research (MCI). It contains many different modalities like: T1w MRI, FDG-PET, Flute/Fib-PET, CSF, genetic data, neuropsychological tests, and demographics. It is evidently one of the highest contributing datasets available in the research area. All patients are from United States of America.
2. PPMI
   Parkinson's Progression Markers Initiative (PPMI) [116] dataset is like ADNI, but more focused on Parkinson's disease. It was first introduced in 2010 and contains a few different clinical, neuroimaging, genetic and biomarkers data. The patients are from 11 different countries.
3. AIBL
   The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) [117] is another dataset, which contains data and focuses on Alzheimer's disease like ADNI. The dataset was first announced in 2006. All patients are from Australia.
4. CLAS
   Chinese Longitudinal Aging Study (CLAS) [118] dataset was first introduced in 2011 and it contains demographic, neuropsychiatric, genetic data, biomarkers, and T1w MRI scans of patients from China.
5. OASIS
   The Open Access Series of Imaging Studies (OASIS) [119] is a multimodal neurimaging dataset, which was first released in 2007. It's purpose it to provide open access to neuroimaging data. It contains longitudinal multimodal neuroimaging, clinical, cognitive, and

biomarker data of patiets. There are 4 versions of the dataset. The newest one is OASIS-4, which focuses on memory disorders and dementia.

5.7 What are the commonly used pre-processing and feature extraction from neuroimages techniques?

We have collected from the second stage in research all pre-processing and feature extraction techniques used by the researchers. Results are listed in Table 20. We also collected the software which was used and grouped it by the neuroimaging modalities.

Pre-processing and feature extraction of structural MR images usually contains the same steps as PET images, where fMRI has additional pre-processing steps to average out motion and timings due to the procedure of the scan, where scans are taken in a long period of time and it is impossible for patient to stay still during the whole procedure. The same applies to diffusion neuroimaging (DWI and DTI). The only distinguishable differences between fMRI and DWI, DTI are types of features, which are extracted. For fMRI (in this study) researchers calculated average time series signal for segmented brain regions, where for diffusion images they were calculating more specific features like eigenvalues and Fractional Anisotropy (FA). All the pre-processing, feature extraction techniques and software involved in the process are briefly described below:

Pre-processing techniques:

1. Intensity normalization

   Intensity normalization also known as bias field correction is a technique which allows to eliminate bias field, which is a "low-frequency and very smooth signal that corrupts MRI images specially those produced by old MRI machines." [120]. Using images which are not bias field corrected will probably yield bad results, due to gray color intensities in pixels being disrupted due to this signal.
2. Spacial normalization

   Spacial normalization or registration is a step, which allows to align one scan with another due to human brain/head sizes and shapes being different. This step involves aligning scan with the template scan (usually an average template obtained specifically for the study) or a standard template like MNI (McConnel Brain Imaging) or ICBM (International Consortium for Brain Mapping) [121], which maps the location of brain in the scan to the location in the template. This way, all human brains in different scans appear in the same location.
3. Skull stripping

   Skull stripping or brain extraction is a procedure where unwanted tissues, skull, eyes etc. are removed from the scan to produce a volume that contains only the brain tissue. This allows to reduce the scan complexity and size, which makes the other processing steps faster due to smaller size of data in the scan.
4. Noise reduction or smoothing

   This step is optional, but sometimes performed to remove noise from the images by applying smoothing filters like anisotropic diffusion, Gaussian filter, median filter, or other statistical filters.
5. Transformations

   Transformations can be applied to neuroimages to convert to different coordinates (Talairach), Mean Regional Homogeneity (mReHo) that "measure the regional synchronization degree of fMRI time course" [122], Discrete Wavelet can be used to fuse the different signals in data [90]. All the transformations allow us to later extract different information from the images/signals.

Table 20 Common pre-processing and feature extraction techniques identified in selected articles

Data modality	Pre-processing	Feature extraction techniques	Software
T1w MRI	Intensity normalization. Spatial normalization. Skull stripping. Noise reduction or smoothing. Transformations (Mean Regional Homogeneity, Discrete Wavelet, Talairach). Segmentation.	Average intensity for each voxel in segmented regions of interest. Surface extraction. Feature extraction with Neural Networks (3DMR-PCANet, 3DResNet-10, CNN, 3D CNN, GAN, FCN).	FreeSurfer FSL MIPAV
FDG-PET, Flute/PiB-PET, Amyloid-PET			FreeSurfer FSL
fMRI	Time volumes removal. Slice timing correction. Motion correction. Brain extraction. Spatial normalization. Smoothing. Filtering. Transformations (Mean Regional Homogeneity). Segmentation	Average time series signal for segmented ROI	SPM12 SPM8 DPABI GRETNA AFNI
DWI, DTI	Intensity normalization. Spacial normalization. Motion correction. Brain extraction. Denoise	Eigenvalues computation for each voxel. Tract-based spatial statistics (TBSS) for Fractional Anisotropy (FA) and Mean Diffusivity (MD).	FreeSurfer FSL AFNI

6. Segmentation
   Segmentation step is one of the most frequent steps involved in the processing pipeline, when dealing with hand-crafted features. Segmenting images into Gray Matter (GM), White Matter (WM) or Cerebrospinal Fluid (CSF) layers allows to express these features later as numbers by calculating voxel intensities. There are anatomical atlases, which are templates for pre-processed brain to be subdivided into ROI (Region of Interest), for example, AAL (Automated Anatomical Labelling) atlas [123]. Then each region can be analyzed separately or used to extract features.
7. Motion correction
   Due to the long fMRI capturing procedure it is impossible for patients to stay still the whole time. Therefore, in the images this will look as if the patient moved their head from slice to slice. This can be fixed by using a reference image, for example, in the first slice and then capturing in subsequent slices whether the brain has been translated in the space from reference, then motion can be fixed by translating the brain volume to the reference slice location.
8. Filtering
   There are different filtering techniques that researchers use to clean and separate signals in fMRI data like linear detrending, high/low pass filtering or other temporal filtering techniques. Essentially, filter selection depends on the purpose of the pre-processing step, whether we want to remove low frequencies (high pass filter) or high frequencies (low pass filter), or to remove linear drifts in signal (linear detrend). Filtering data allows researchers to analyze the signals in different ways.

Feature extraction techniques:

1. Voxel average intensity
   Voxel is an atomic unit in 3D space. Typically, segmented brain regions are analyzed by measuring voxel intensity values in each ROI. Then the average value is saved as a feature.
2. Surface extraction
   In some cases, researchers used brain surface extraction tools to generate surface meshes to estimate cortical thickness as a feature in detection of dementia. Or extracts a Gray Matter (GM) or White Matter (WM) features, that can be used as features.
3. Features from neural networks
   These types of feature extraction techniques are the least common among researchers as previously mentioned. Typically, some type of CNN is used to capture high level features from images or 3D volumes of brains. For example, 3DMR-PCANet, 3DResNet-10, CNN, 3D CNN, GAN, FCN.
4. Average time series
   This average time series signal is exactly the same as average voxel intensity values. The only difference is that this time series signal represents average voxel intensities changing in time due to fMRI data being 4D (having temporal component).
5. Eigenvalues
   Diffusion Tensor Imaging allows to capture the directionality and magnitude of water diffusion. Diffusion tensors are diagonal, that contains three nonzero elements, which are called eigenvalues [124].
6. Tract-based spacial statistics (TBSS) for Fractional Anisotropy (FA) and Mean Diffusivity (MD)
   Tract-based spacial statistics is a whole process of preprocessing diffusion neuroimaging data, from spacial normalization to segmentation to acquiring data from fractional anisotropy or mean diffusivity metrics at each voxel. Fractional anisotropy is a value, which is calculated from the eigenvalues. It measures the degree of anisotropy of a diffusion at each voxel in the scan. This metric ranges from 0 to 1, where 0 represents isotropic diffusion and 1 highly directional. Mean diffusivity is another diffusion degree describing metric, which does not have directional component. This metric is calculated as an average of all eigenvalues.

Software:

1. FreeSurfer [109] - open-source tool to analyze and visualize structural, functional and diffusion neuroimaging data.
2. FSL [108] - library of tools to analyze structural, functional and diffusion neuroimaging data.
3. SPM12 and SPM8 [125] - are libraries to analyze and visualize functional neuroimaging data.
4. MIPAV [126] - is a toolbox which contains tool to analyze and visualize structural neuroimaging data.
5. DPABI [127] - library of tools to process and analyze functional neuroimaging data.
6. GRETNA [128] - another library to process and analyze functional neuroimaging data.
7. AFNI [129] - toolbox to analyze and visualize functional and diffusion neuroimaging data.

5.8 What are the key issues in detection of early dementia?

Based on the data collected from the papers, we think these are the issues we can see in the research are, that requires attention from the researchers:

1. Explainability
   Only 4 papers from 40 try to look at explainability issue and propose some kind of solution for it, for example [93] mentions the way to detect disease related brain-regions with optimal feature analysis methods, [99] tries to analyze and find the most affected brain region by AD, [79] paper which proposes a neuroimage fusion

techniques, which gives more context on the results from the model, and [75] uses Grad-CAM [107] to show to what regions in image model focuses the most. This issue is one of the drawbacks of created methods, because in clinical practice doctors need to assess the diagnosis and the lack of interpretability of the ML/DL methods, which essentially are "blackboxes" does not make it easier. This key issue has not been solved yet, and our findings meet with the ones mentioned in reviewed literature reviews. To solve the explainability issue we would have to conduct a separate study, but as a guidance, if we are dealing with CNNs it is possible to use previously mentioned Grad-CAM or Shap values [106], which allows to see what features were most used for particular class. For neural networks it is possible to utilize decision trees, where solution could be traced from tree roots.

2. Small datasets

   Publicly accessible datasets are usually full of different modalities, but there is a small amount of patients in that dataset, who go through many different diagnostic methods and therefore have a small amount of modalities available for researchers. This is particularly sensitive for DL methods, which tend to overfit small datasets. To overcome this, researchers could collect all available public datasets and make a union. This could improve the variaty of data as different patients have different anatomy.

3. Same dataset for validation

   Using the same dataset for model validation might introduce bias and falsely inflate the results of the model's performance, therefore many researchers use cross-validation as a solution to this problem. However, the most preferred way would be to use a completely separate dataset for validation, this eliminates the risk of bias introduction.

4. Selection of ROI

   When working with neuroimages and extracting features typically researchers ROI (Region of Interest) average voxel intensities. There is usually a step in the process after feature extraction, to do a feature selection, which would reduce the amount of features for classifier to deal with. However, there is a possibility that feature selection method would introduce bias and might not select regions, which could be disease related. Most papers that use neuroimages extract handcrafted features (65%), which raises this ROI selection issue. The distribution of types of features is depicted in Fig. 13. To overcome this selection bias, it would be beneficial to cross-validate solutions with multiple different collections of ROI. Then we could compare and analyze why some collection of ROI is performing better or worse than others.

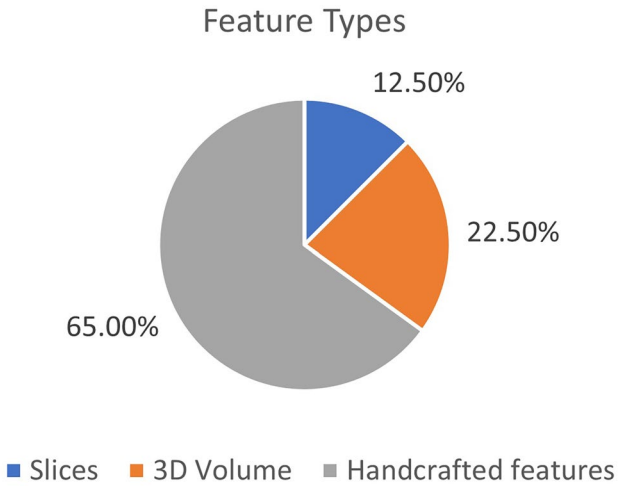

Fig. 13 Feature types extracted from neuroimages distribution

5. Imbalanced datasets

   Some of the researchers used imbalanced datasets in their research. Data imbalance means that one class contains more samples than the other. This potentially increases bias into the model, because it might learn more features from one class than the other. This reduces generalizability and is preferred to always have the same number of samples in each data class.

5.9 What are the potential future research areas?

Based on this study, these are the main research areas, which should get more attention in the future:

1. Separate datasets for validation

   As mentioned in the key issues Section 5.8, having separate dataset for validation is the most preferred way in the classification task, because it allows to eliminate the potential of bias introduction. We think the researchers should employ separate datasets for the validation of created models instead of cross-validation and random split.

2. Explainability

   It is clear at this point, that in order for AI to be more adopted in the clinical practice, researchers should spend more time in their research to improve the explainability of the models decision, because usually it is extremely hard to understand why some decision is made by the model if there is no interpretability component in it. As this issues is not solved yet, we believe it will receive increasing amount of attention from researchers in future.

3. Data imputation

   Usually in open source datasets, there is a small amount of samples, which contain all of the selected data modalities in the study. Therefore, researchers

should look more into methods that can deal with missing data, or use data imputation techniques, like training GAN models to generate missing data modalities. This key issue is still present as it was also discovered in the reviewed literature reviews.

4. Multi-class classification

   This task is not receiving enough attention at this moment. From this study we identified that only a small amount (5 out of 40) of articles try to investigate the multi-class classification problem. We believe that models developed to detect multiple dementia types or sub-types, for example, early or late MCI would be more beneficial. The amount of work required to create, train and validate one model versus a few when it comes to classifying different dementia types - significantly reduces. However, training multi-class classifier comes with its own challenges: one being the requirement for more distinct samples for each class to learn differences better. The problem is challenging and should receive more attention in the future from researchers.

5. More dementia detection types

   In the study we observed that only 2 papers investigated Parkinson's disease and 1 paper Vascular dementia. There are plenty of other dementia types, which were not captured by this study like Frontotemporal dementia, Huntington's disease or HIV dementia. The study is limited, but also the amount of papers investigating these dementias. Therefore, we think in future other dementia types could be investigated instead of Alzheimer's disease, which seems to be already reaching perfect detection accuracies.

5.10 Limitations

A single person performed this study. Therefore, no cross-validation of the results was performed. There could be a potential bias involved in the selection of the articles eligible for the study, data extraction errors, which could have occurred while extracting main key points from the selected studies. We tried to include as many papers from different scientific databases as we could to include all the relevant studies. However, there is a possibility that some studies were not included in the query results.

6 Conclusion

This study investigated the early dementia detection problem from the multi-modal perspective with the focus on neuroimaging being used as one of the modalities. We reviewed 19 related literature reviews and 40 selected articles in the study. We used PRISMA methodology to query 5 databases and select papers to promote reproduction of the study. We defined the dementia detection problem domain, extracted main issues and future research ares from past literature reviews, looked at what dementia types researchers focus their studies, what are state-of-the-art methods in the different dementia detection groups by comparing methods by accuracy, specificity and sensitivity, investigated different modalities combinations used and how models are being evaluated and validated, gathered datasets utilized in the research, common pre-processing and feature extraction from neuroimages techniques, defined key issues that we think are important in this research area as well as potential future research areas. Key findings of the research: (1) Alzheimer's disease and MCI are the most researched dementia types in the field, 70% and 42.5% articles investigated them respectively; (2) typical choice for dementia detection is ML methods, 75.29% of papers used, where SVM and RF are the most popular types, 47.5% and 22.5% of papers used them respectively; (3) the most popular modalities combination is T1w + FDG-PET, 32.5% of articles used it; (4) accuracy, sensitivity and specificity are the main evaluation metrics used by the researchers; (5) k-fold validation is being used the most, 75% of articles used it, due to relatively small training and validation datasets; (6) ADNI is the most used dataset by researchers, 87.5% of articles used it; (7) intensity and spacial normalization, skull stripping and segmentation are the most common pre-processing techniques for neuroimages; (8) voxel average intensities are being used the most as features in classification extracted from neuroimages; (9) explainability still persists as one of the main issues in adoption of developed methods in clinical practise, because our findings match proposed by previous research; (10) there is a lack of research for other types of dementias, for example Vascular dementia, Frontotemporal dementia, Parkinson's disease and Huntington's disease. We sincerely hope that this study will help other researchers to focus their studies on the key aspects learned from our study.

Author contributions Conceptualization, Rytis Maskeliunas; methodology, Rytis Maskeliunas and Robertas Damaševičius; validation, Robertas Damaševičius, Rytis Maskeliunas and Ovidijus Grigas; formal analysis, Robertas Damaševičius, Rytis Maskeliunas and Ovidijus Grigas; investigation, Robertas Damaševičius, Rytis Maskeliunas and Ovidijus Grigas; resources, Rytis Maskeliunas; writing—original draft preparation, Ovidijus Grigas; writing—review and editing, Rytis Maskeliunas and Robertas Damaševičius; visualization, Ovidijus Grigas; supervision, Rytis Maskeliunas.

Funding No funding was used to support this study.

Data availability There is no data associated with the article.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Conflict of interest The authors declare no conflict of interest.

References

1. World Health Organization. Dementia. 2021. https://www.who.int/en/news-room/fact-sheets/detail/dementia. Accessed 3 Mar 2023.
2. NIH. What is dementia? Symptoms, types, and diagnosis. 2022. https://www.nia.nih.gov/health/what-is-dementia. Accessed 3 Mar 2023.
3. U.S Department of Health and Human Services (HHS). What is dementia? 2022. https://www.alzheimers.gov/alzheimers-dementias/what-is-dementia. Accessed 3 Mar 2023.
4. Tăuţan A-M, Ionescu B, Santarnecchi E. Artificial intelligence in neurodegenerative diseases: a review of available tools with a focus on machine learning techniques. Artif Intell Med. 2021;117:102081. https://doi.org/10.1016/j.artmed.2021.102081.
5. Li R, Wang X, Lawler K, Garg S, Bai Q, Alty J. Applications of artificial intelligence to aid early detection of dementia: a scoping review on current capabilities and future directions. J Biomed Inform. 2022;127:104030. https://doi.org/10.1016/j.jbi.2022.104030.
6. Battineni G, Chintalapudi N, Hossain MA, Losco G, Ruocco C, Sagaro GG, Traini E, Nittari G, Amenta F. Artificial intelligence models in the diagnosis of adult-onset dementia disorders: a review. Bioengineering. 2022;9(8):370. https://doi.org/10.3390/bioengineering9080370.
7. Saravanan S, Ramkumar K, Adalarasu K, Sivanandam V, Kumar SR, Stalin S, Amirtharajan R. A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson's disease. Arch Comput Meth Eng. 2022;29(6):3639–53. https://doi.org/10.1007/s11831-022-09710-1.
8. Goyal P, Rani R, Singh K. State-of-the-art machine learning techniques for diagnosis of Alzheimer's disease from MR-images: a systematic review. Arch Comput Meth Eng. 2021;29(5):2737–80. https://doi.org/10.1007/s11831-021-09674-8.
9. Billeci L, Badolato A, Bachi L, Tonacci A. Machine learning for the classification of Alzheimer's disease and its prodromal stage using brain diffusion tensor imaging data: a systematic review. Processes. 2020;8(9):1071. https://doi.org/10.3390/pr8091071.
10. Loh HW, Hong W, Ooi CP, Chakraborty S, Barua PD, Deo RC, Soar J, Palmer EE, Acharya UR. Application of deep learning models for automated identification of Parkinson's disease: a review (2011–2021). Sensors. 2021;21(21):7034. https://doi.org/10.3390/s21217034.
11. Fathi S, Ahmadi M, Dehnad A. Early diagnosis of Alzheimer's disease based on deep learning: a systematic review. Comput Biol Med. 2022;146:105634. https://doi.org/10.1016/j.compbiomed.2022.105634.
12. Pereira CR, Pereira DR, Weber SA, Hook C, de Albuquerque VHC, Papa JP. A survey on computer-assisted Parkinson's disease diagnosis. Artif Intell Med. 2019;95:48–63. https://doi.org/10.1016/j.artmed.2018.08.007.
13. Khan P, Kader MF, Islam SMR, Rahman AB, Kamal MS, Toha MU, Kwak K-S. Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. IEEE Access. 2021;9:37622–55. https://doi.org/10.1109/access.2021.3062484.
14. Ahmed MR, Zhang Y, Feng Z, Lo B, Inan OT, Liao H. Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. IEEE Rev Biomed Eng. 2019;12:19–33. https://doi.org/10.1109/rbme.2018.2886237.
15. Agarwal D, Marques G, de la Torre-Díez I, Martin MAF, Zapiraín BG, Rodríguez FM. Transfer learning for Alzheimer's disease through neuroimaging biomarkers: a systematic review. Sensors. 2021;21(21):7259. https://doi.org/10.3390/s21217259.
16. Blanco K, Salcidua S, Orellana P, Sauma-Pérez T, León T, Steinmetz LCL, Ibañez A, Duran-Aniotz C, de la Cruz R. Systematic review: fluid biomarkers and machine learning methods to improve the diagnosis from mild cognitive impairment to Alzheimer's disease. Alzheimers Res Ther. 2023;15(1). https://doi.org/10.1186/s13195-023-01304-8.
17. Zhao Z, Chuah JH, Lai KW, Chow C-O, Gochoo M, Dhanalakshmi S, Wang N, Bao W, Wu X. Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: a review. Front Comput Neurosci. 2023;17. https://doi.org/10.3389/fncom.2023.1038636.
18. Viswan V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen F. Explainable artificial intelligence in Alzheimer's disease classification: a systematic review. Cogn Comput. 2023. https://doi.org/10.1007/s12559-023-10192-x.
19. Aggarwal N, Saini BS, Gupta S. Role of artificial intelligence techniques and neuroimaging modalities in detection of Parkinson's disease: a systematic review. Cogn Comput. 2023. https://doi.org/10.1007/s12559-023-10175-y.
20. Arya AD, Verma SS, Chakarabarti P, Chakrabarti T, Elngar AA, Kamali A-M, Nami M. A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. Brain Inform. 2023;10(1). https://doi.org/10.1186/s40708-023-00195-7.
21. Huang G, Li R, Bai Q, Alty J. Multimodal learning of clinically accessible tests to aid diagnosis of neurodegenerative disorders: a scoping review. Health Inf Sci Syst. 2023;11(1). https://doi.org/10.1007/s13755-023-00231-0.
22. Muhammed Niyas K, Thiyagarajan P. A systematic review on early prediction of mild cognitive impairment to Alzheimer's using machine learning algorithms. Int J Intell Netw. 2023;4:74–88. https://doi.org/10.1016/j.ijin.2023.03.004.
23. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, Brayne C, Burns A, Cohen-Mansfield J, Cooper C, Costafreda SG, Dias A, Fox N, Gitlin LN, Howard R, Kales HC, Kivimäki M, Larson EB, Ogunniyi A, Orgeta V, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N. Dementia prevention, intervention, and care: 2020 report of the lancet commission. The Lancet. 2020;396(10248):413–46. https://doi.org/10.1016/s0140-6736(20)30367-6.
24. van de Vorst IE, Koek HL, de Vries R, Bots ML, Reitsma JB, Vaartjes I. Effect of vascular risk factors and diseases on mortality in individuals with dementia: a systematic review and meta-analysis. J Am Geriatr Soc. 2015;64(1):37–46. https://doi.org/10.1111/jgs.13835.
25. Anstey KJ, Cherbuin N, Budge M, Young J. Body mass index in midlife and late-life as a risk factor for dementia: a meta-analysis of prospective studies. Obes Rev. 2011;12(5):e426–43. https://doi.org/10.1111/j.1467-789x.2010.00825.x.
26. Anstey KJ, von Sanden C, Salim A, O'Kearney R. Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. Am J Epidemiol. 2007;166(4):367–78. https://doi.org/10.1093/aje/kwm116.
27. Gardener H, Wright CB, Rundek T, Sacco RL. Brain health and shared risk factors for dementia and stroke. Nat Rev Neurol. 2015;11(11):651–7. https://doi.org/10.1038/nrneurol.2015.195.
28. Morris JC. The clinical dementia rating (CDR). Neurology. 1993;43(11):2412.2-2412-a. https://doi.org/10.1212/wnl.43.11.2412-a.

29. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. J Psychiatr Res. 1975;12(3):189–98. https://doi.org/10.1016/0022-3956(75)90026-6.
30. Cullen B, O'Neill B, Evans JJ, Coen RF, Lawlor BA. A review of screening tests for cognitive impairment. J Neurol Neurosurg Psychiatry. 2007;78(8):790–9. https://doi.org/10.1136/jnnp.2006.095414.
31. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale-cognitive subscale (ADAS-COG): Modifications and responsiveness in pre-dementia populations. A narrative review. J Alzheimers Dis. 2018;63(2):423–44. https://doi.org/10.3233/JAD-170991.
32. King JH, Gfeller JD, Davis HP. Detecting simulated memory impairment with the Rey auditory verbal learning test: implications of base rates and study generalizability. J Clin Exp Neuropsychol. 1998;20(5):603–12. https://doi.org/10.1076/jcen.20.5.603.1124.
33. Brañez-Condorena A, Soriano-Moreno DR, Navarro-Flores A, Solis-Chimoy B, Diaz-Barrera ME, Taype-Rondan A. Accuracy of the geriatric depression scale (GDS)-4 and GDS-5 for the screening of depression among older adults: a systematic review and meta-analysis. PLoS ONE. 2021;16(7):e0253899. https://doi.org/10.1371/journal.pone.0253899.
34. González DA, Gonzales MM, Resch ZJ, Sullivan AC, Soble JR. Comprehensive evaluation of the functional activities questionnaire (FAQ) and its reliability and validity. Assessment. 2021;29(4):748–63. https://doi.org/10.1177/1073191121991215.
35. Kaufer DI, Cummings JL, Ketchel P, Smith V, MacMillan A, Shelley T, Lopez OL, DeKosky ST. Validation of the NPI-Q, a brief clinical form of the neuropsychiatric inventory. J Neuropsychiatry Clin Neurosci. 2000;12(2):233–9. https://doi.org/10.1176/jnp.12.2.233.
36. NHS. Symptoms of dementia. 2022. https://www.nhs.uk/conditions/dementia/symptoms/. Accessed 3 Mar 2023.
37. Castellani RJ, Rolston RK, Smith MA. Alzheimer disease. Dis Mon. 2010;56(9):484–546. https://doi.org/10.1016/j.disamonth.2010.06.001.
38. O'Brien JT, Thomas A. Vascular dementia. The Lancet. 2015;386(10004):1698–706. https://doi.org/10.1016/s0140-6736(15)00463-8.
39. Walker Z, Possin KL, Boeve BF, Aarsland D. Lewy body dementias. The Lancet. 2015;386(10004):1683–97. https://doi.org/10.1016/s0140-6736(15)00462-6.
40. Parkinson's Foundation. Parkinson's disease vs. Parkinsonism. 2022. https://www.parkinson.org/library/fact-sheets/parkinsonism. Accessed 3 Mar 2023.
41. Bang J, Spina S, Miller BL. Frontotemporal dementia. The Lancet. 2015;386(10004):1672–82. https://doi.org/10.1016/s0140-6736(15)00461-4.
42. Walker FO. Huntington's disease. The Lancet. 2007;369(9557):218–28. https://doi.org/10.1016/s0140-6736(07)60111-1.
43. McArthur JC. HIV dementia: an evolving disease. J Neuroimmunol. 2004;157(1–2):3–10. https://doi.org/10.1016/j.jneuroim.2004.08.042.
44. Nath A, Schiess N, Venkatesan A, Rumbaugh J, Sacktor N, Mcarthur J. Evolution of HIV dementia with HIV infection. Int Rev Psychiatry. 2008;20(1):25–31. https://doi.org/10.1080/09540260701861930.
45. Medspace. FDA clears 5-minute test for early dementia. 2022. https://www.medscape.com/viewarticle/961277. Accessed 3 Mar 2023.
46. Winder Z, Sudduth TL, Anderson S, Patel E, Neltner J, Martin BJ, Snyder KE, Abner EL, Jicha GA, Nelson PT, Wilcock DM. Examining the association between blood-based biomarkers and human post mortem neuropathology in the University of Kentucky Alzheimer's disease research center autopsy cohort. Alzheimers Dement. 2022;19(1):67–78. https://doi.org/10.1002/alz.12639.
47. Alzheimer's Society. Genetic testing for dementia. 2022. https://www.alzheimers.org.uk/about-dementia/risk-factors-and-prevention/genetic-testing-dementia'. Accessed 3 Mar 2023.
48. Knopman DS, Petersen RC, Jack CR. A brief history of "Alzheimer disease''. Neurology. 2019;92(22):1053–9. https://doi.org/10.1212/wnl.0000000000007583.
49. Dembitsky VM, Dzhemileva L, Gloriozova T, D'yakonov V. Natural and synthetic drugs used for the treatment of the dementia. Biochem Biophys Res Commun. 2020;524(3):772–83. https://doi.org/10.1016/j.bbrc.2020.01.123.
50. Hafdi M, Hoevenaar-Blom MP, Richard E. Multi-domain interventions for the prevention of dementia and cognitive decline. Cochrane Database Syst Rev. 2021;2021(11). https://doi.org/10.1002/14651858.cd013572.pub2.
51. Amieva H, Robert PH, Grandoulier A-S, Meillon C, Rotrou JD, Andrieu S, Berr C, Desgranges B, Dubois B, Girtanner C, Joël M-E, Lavallart B, Nourhashemi F, Pasquier F, Rainfray M, Touchon J, Chêne G, Dartigues J-F. Group and individual cognitive therapies in Alzheimer's disease: the ETNA3 randomized trial. Int Psychogeriatr. 2015;28(5):707–17. https://doi.org/10.1017/s1041610215001830.
52. Cammisuli DM, Cipriani G, Giusti EM, Castelnuovo G. Effects of reminiscence therapy on cognition, depression and quality of life in elderly people with Alzheimer's disease: a systematic review of randomized controlled trials. J Clin Med. 2022;11(19):5752. https://doi.org/10.3390/jcm11195752.
53. Spector A, Thorgrimsen L, Woods B, Royan L, Davies S, Butterworth M, Orrell M. Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia. Br J Psychiatry. 2003;183(3):248–54. https://doi.org/10.1192/bjp.183.3.248.
54. Taylor J-P, McKeith IG, Burn DJ, Boeve BF, Weintraub D, Bamford C, Allan LM, Thomas AJ, O'Brien JT. New evidence on the management of Lewy body dementia. Lancet Neurol. 2020;19(2):157–69. https://doi.org/10.1016/s1474-4422(19)30153-x.
55. Moreno-Morales C, Calero R, Moreno-Morales P, Pintado C. Music therapy in the treatment of dementia: a systematic review and meta-analysis. Front Med. 2020;7. https://doi.org/10.3389/fmed.2020.00160.
56. Pradhan AU, Uwishema O, Onyeaka H, Adanur I, Dost B. A review of stem cell therapy: an emerging treatment for dementia in Alzheimer's and Parkinson's disease. Brain Behav. 2022;12(9). https://doi.org/10.1002/brb3.2740.
57. NHS. What are the treatments for dementia? 2022. https://www.nhs.uk/conditions/dementia/treatment/. Accessed 3 Mar 2023.
58. Orgeta V, McDonald KR, Poliakoff E, Hindle JV, Clare L, Leroi I. Cognitive training interventions for dementia and mild cognitive impairment in Parkinson's disease. Cochrane Database Syst Rev. 2020;2020(2). https://doi.org/10.1002/14651858.cd011961.pub2.
59. Alzheimer Europe. Prevalence of dementia in Europe. 2019. https://www.Alzheimer-europe.org/dementia/prevalence-dementia-europe. Accessed 3 Mar 2023.
60. Gustavsson A, Norton N, Fast T, Frölich L, Georges J, Holzapfel D, Kirabali T, Krolak-Salmon P, Rossini PM, Ferretti MT, Lanman L, Chadha AS, van der Flier WM. Global estimates on the number of persons across the Alzheimer's disease continuum. Alzheimers Dement. 2022. https://doi.org/10.1002/alz.12694.
61. Fymat A. Dementia: a review. Journal of Clinical Psychiatry and Neuroscience. 2018;1(3):27–34. https://www.pulsus.com/scholarly-articles/dementia-a-review-5127.html. Accessed 3 Mar 2023.
62. Alzheimer's Association. Mild cognitive impairment (MCI). 2022. https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment. Accessed 3 Mar 2023.

63. Alzheimer's Association. Alzheimer's disease facts and figures. 2022. https://www.alz.org/alzheimers-dementia/facts-figures. Accessed 3 Mar 2023.
64. World Health Organization. The top 10 causes of death. 2019. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 3 Mar 2023.
65. Jia J, Wei C, Chen S, Li F, Tang Y, Qin W, Zhao L, Jin H, Xu H, Wang F, Zhou A, Zuo X, Wu L, Han Y, Han Y, Huang L, Wang Q, Li D, Chu C, Shi L, Gong M, Du Y, Zhang J, Zhang J, Zhou C, Lv J, Lv Y, Xie H, Ji Y, Li F, Yu E, Luo B, Wang Y, Yang S, Qu Q, Guo Q, Liang F, Zhang J, Tan L, Shen L, Zhang K, Zhang J, Peng D, Tang M, Lv P, Fang B, Chu L, Jia L, Gauthier S. The cost of Alzheimer's disease in China and re-estimation of costs worldwide. Alzheimers Dement. 2018;14(4):483–91. https://doi.org/10.1016/j.jalz.2017.12.006.
66. Liu Y, Yue L, Xiao S, Yang W, Shen D, Liu M. Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages. Med Image Anal. 2022;75:102266. https://doi.org/10.1016/j.media.2021.102266.
67. Hao X, Bao Y, Guo Y, Yu M, Zhang D, Risacher SL, Saykin AJ, Yao X, Shen L. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. Med Image Anal. 2020;60:101625. https://doi.org/10.1016/j.media.2019.101625.
68. Liu Y, Fan L, Zhang C, Zhou T, Xiao Z, Geng L, Shen D. Incomplete multi-modal representation learning for Alzheimer's disease diagnosis. Med Image Anal. 2021;69:101953. https://doi.org/10.1016/j.media.2020.101953.
69. Hojjati SH, Ebrahimzadeh A, Khazaee A, Babajani-Feremi A. Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. Comput Biol Med. 2018;102:30–9. https://doi.org/10.1016/j.compbiomed.2018.09.004.
70. Zhang J, He X, Liu Y, Cai Q, Chen H, Qing L. Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data. Comput Biol Med. 2023;162:107050. https://doi.org/10.1016/j.compbiomed.2023.107050.
71. Yan B, Li Y, Li L, Yang X, Qiang Li T, Yang G, Jiang M. Quantifying the impact of pyramid squeeze attention mechanism and filtering approaches on Alzheimer's disease classification. Comput Biol Med. 2022;148:105944. https://doi.org/10.1016/j.compbiomed.2022.105944.
72. Tu Y, Lin S, Qiao J, Zhuang Y, Zhang P. Alzheimer's disease diagnosis via multimodal feature fusion. Comput Biol Med. 2022;148:105901. https://doi.org/10.1016/j.compbiomed.2022.105901.
73. Dong A, Zhang G, Liu J, Wei Z. Latent feature representation learning for Alzheimer's disease classification. Comput Biol Med. 2022;150:106116. https://doi.org/10.1016/j.compbiomed.2022.106116.
74. Zhang Y, He X, Chan YH, Teng Q, Rajapakse JC. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from SMRI and pet scans. Comput Biol Med. 2023;164:107328. https://doi.org/10.1016/j.compbiomed.2023.107328.
75. Leng Y, Cui W, Peng Y, Yan C, Cao Y, Yan Z, Chen S, Jiang X, Zheng J. Multimodal cross enhanced fusion network for diagnosis of Alzheimer's disease and subjective memory complaints. Comput Biol Med. 2023;157:106788. https://doi.org/10.1016/j.compbiomed.2023.106788.
76. Pahuja G, Prasad B. Deep learning architectures for Parkinson's disease detection by using multi-modal features. Comput Biol Med. 2022;146:105610. https://doi.org/10.1016/j.compbiomed.2022.105610.
77. Lahmiri S, Shmuel A. Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease. Biomed Signal Process Control. 2019;52:414–9. https://doi.org/10.1016/j.bspc.2018.08.009.
78. Altaf T, Anwar SM, Gul N, Majeed MN, Majid M. Multi-class Alzheimer's disease classification using image and clinical features. Biomed Signal Process Control. 2018;43:64–74. https://doi.org/10.1016/j.bspc.2018.02.019.
79. Kong Z, Zhang M, Zhu W, Yi Y, Wang T, Zhang B. Multi-modal data Alzheimer's disease detection based on 3D convolution. Biomed Signal Process Control. 2022;75:103565. https://doi.org/10.1016/j.bspc.2022.103565.
80. Castellazzi G, Cuzzoni MG, Ramusino MC, Martinelli D, Denaro F, Ricciardi A, Vitali P, Anzalone N, Bernini S, Palesi F, Sinforiani E, Costa A, Micieli G, D'Angelo E, Magenes G, Wheeler-Kingshott CAMG. A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features. Front Neuroinform. 2020;14. https://doi.org/10.3389/fninf.2020.00025.
81. Jin L, Zhao K, Zhao Y, Che T, Li S. A hybrid deep learning method for early and late mild cognitive impairment diagnosis with incomplete multimodal data. Front Neuroinform. 2022;16. https://doi.org/10.3389/fninf.2022.843566.
82. Kumari R, Nigam A, Pushkar S. An efficient combination of quadruple biomarkers in binary classification using ensemble machine learning technique for early onset of Alzheimer disease. Neural Comput Appl. 2022;34(14):11865–84. https://doi.org/10.1007/s00521-022-07076-w.
83. Jia H, Lao H. Deep learning and multimodal feature fusion for the aided diagnosis of Alzheimer's disease. Neural Comput Appl. 2022;34(22):19585–98. https://doi.org/10.1007/s00521-022-07501-0.
84. Divya R, Shantha Selva Kumari R, Alzheimer's Disease Neuroimaging Initiative. Genetic algorithm with logistic regression feature selection for Alzheimer's disease classification. Neural Comput Appl. 2021;33(14):8435–44. https://doi.org/10.1007/s00521-020-05596-x.
85. Zhu Q, Yuan N, Huang J, Hao X, Zhang D. Multi-modal AD classification via self-paced latent correlation analysis. Neurocomputing. 2019;355:143–54. https://doi.org/10.1016/j.neucom.2019.04.066.
86. Segovia F, Ramírez J, Castillo-Barnes D, Salas-Gonzalez D, Gómez-Río M, Sopena-Novales P, Phillips C, Zhang Y, Górriz J. Multivariate analysis of dual-point amyloid PET intended to assist the diagnosis of Alzheimer's disease. Neurocomputing. 2020;417:1–9. https://doi.org/10.1016/j.neucom.2020.06.081.
87. Abdelaziz M, Wang T, Elazab A. Alzheimer's disease diagnosis framework from incomplete multimodal data using convolutional neural networks. J Biomed Inform. 2021;121:103863. https://doi.org/10.1016/j.jbi.2021.103863.
88. Song X, Elazab A, Zhang Y. Classification of mild cognitive impairment based on a combined high-order network and graph convolutional network. IEEE Access. 2020;8:42816–27. https://doi.org/10.1109/access.2020.2974997.
89. Perez-Gonzalez J, Jiménez-Ángeles L, Saavedra KR, Morales EB, Medina-Bañuelos V. Mild cognitive impairment classification using combined structural and diffusion imaging biomarkers. Phys Med Biol. 2021;66(15):155010. https://doi.org/10.1088/1361-6560/ac0e77.
90. Dwivedi S, Goel T, Tanveer M, Murugan R, Sharma R. Multimodal fusion-based deep learning network for effective diagnosis of Alzheimer's disease. IEEE Multimedia. 2022;29(2):45–55. https://doi.org/10.1109/mmul.2022.3156471.
91. Ye H, Zhu Q, Yao Y, Jin Y, Zhang D. Pairwise feature-based generative adversarial network for incomplete multi-modal Alzheimer's disease diagnosis. Vis Comput. 2022. https://doi.org/10.1007/s00371-021-02354-5.

92. Asim Y, Raza B, Malik AK, Rathore S, Hussain L, Iftikhar MA. A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning. Int J Imaging Syst Technol. 2018;28(2):113–23. https://doi.org/10.1002/ima.22263.
93. Bi X-A, Zhou W, Li L, Xing Z. Detecting risk gene and pathogenic brain region in EMCI using a novel GERF algorithm based on brain imaging and genetic data. IEEE J Biomed Health Inform. 2021;25(8):3019–28. https://doi.org/10.1109/jbhi.2021.3067798.
94. Angkoso CV, Tjahyaningtijas HPA, Adrianto Y, Sensusiati AD, Purnama IKE, Purnomo MH. Multi-features fusion in multi-plane MRI images for Alzheimer's disease classification. Int J Intell Eng Syst. 2022;15(4). https://doi.org/10.22266/ijies2022.0831.17.
95. Liu M, Cheng D, Wang K, Wang Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics. 2018;16(3–4):295–308. https://doi.org/10.1007/s12021-018-9370-4.
96. Pahuja G, Nagabhushan T, Prasad B, Pushkarna R. Early detection of Parkinson's disease through multimodal features using machine learning approaches. Int J Signal Imaging Syst Eng. 2018;11(1):31. https://doi.org/10.1504/ijsise.2018.090605.
97. Liu J, Tan G, Lan W, Wang J. Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. BMC Bioinformatics. 2020;21(S6). https://doi.org/10.1186/s12859-020-3437-6.
98. Yuan S, Li H, Wu J, Sun X. Classification of mild cognitive impairment with multimodal data using both labeled and unlabeled samples. IEEE/ACM Trans Comput Biol Bioinf. 2021;18(6):2281–90. https://doi.org/10.1109/tcbb.2021.3053061.
99. Yang B-H, Chen J-C, Chou W-H, Huang W-S, Fuh J-L, Liu R, Wu C-H. Classification of Alzheimer's disease from 18f-FDG and 11c-PiB PET imaging biomarkers using support vector machine. J Med Biol Eng. 2020;40(4):545–54. https://doi.org/10.1007/s40846-020-00548-1.
100. Zhang Y, Wang S, Xia K, Jiang Y, Qian P. Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. Inf Fusion. 2021;66:170–83. https://doi.org/10.1016/j.inffus.2020.09.002.
101. Zhang Y, He X, Liu Y, Ong CZL, Liu Y, Teng Q. An end-to-end multimodal 3D CNN framework with multi-level features for the prediction of mild cognitive impairment. Knowl-Based Syst. 2023;281:111064. https://doi.org/10.1016/j.knosys.2023.111064.
102. Gao X, Shi F, Shen D, Liu M. Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease. Comput Med Imaging Graph. 2023;110:102303. https://doi.org/10.1016/j.compmedimag.2023.102303.
103. Chen H, Guo H, Xing L, Chen D, Yuan T, Zhang Y, Zhang X. Multimodal predictive classification of Alzheimer's disease based on attention-combined fusion network: integrated neuroimaging modalities and medical examination data. IET Image Proc. 2023;17(11):3153–64. https://doi.org/10.1049/ipr2.12841.
104. Ding D, Zeng X, Wang X, Zhang J. AMIM: an adaptive weighted multimodal integration model for Alzheimer's disease classification. Int J Adv Comput Sci Appl. 2023;14(1). https://doi.org/10.14569/IJACSA.2023.01401108
105. Ismail WN, Fathimathul Rajeena PP, Ali MAS. A meta-heuristic multi-objective optimization method for Alzheimer's disease detection based on multi-modal data. Mathematics. 2023;11(4):957. https://doi.org/10.3390/math11040957.
106. Holzinger A, Saranti A, Molnar C, Biecek P, Samek W. Explainable AI methods - a brief overview. In: xxAI - Beyond Explainable AI. Springer International Publishing; 2022. p. 13–38. https://doi.org/10.1007/978-3-031-04083-2_2.
107. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int J Comput Vision. 2019;128(2):336–59. https://doi.org/10.1007/s11263-019-01228-7.
108. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, Luca MD, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, Stefano ND, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004;23:S208–19. https://doi.org/10.1016/j.neuroimage.2004.07.051.
109. The General Hospital Corporation. Freesurfer. 2023. https://surfer.nmr.mgh.harvard.edu/. Accessed 3 Mar 2023.
110. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75. https://doi.org/10.1086/519795.
111. Tripathi M, Vibha D. Reversible dementias. Indian Journal of Psychiatry. 2009;51(Suppl 1):S52-5.
112. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). Alzheimers Dement. 2005;1(1):55–66. https://doi.org/10.1016/j.jalz.2005.06.003.
113. Chee MWL, Zheng H, Goh JOS, Park D, Sutton BP. Brain structure in young and old East Asians and Westerners: comparisons of structural volume and cortical thickness. J Cogn Neurosci. 2011;23(5):1065–79. https://doi.org/10.1162/jocn.2010.21513.
114. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. arXiv:1406.2661 [Preprint]. 2014. Available from: https://arxiv.org/abs/1406.2661. Accessed 3 Mar 2023.
115. Biondi A, Laiou P, Bruno E, Viana PF, Schreuder M, Hart W, Nurse E, Pal DK, Richardson MP. Remote and long-term self-monitoring of electroencephalographic and noninvasive measurable variables at home in patients with epilepsy (EEG@HOME): Protocol for an observational study. JMIR Research Protocols. 2021;10(3):e25309. https://doi.org/10.2196/25309.
116. Marek K, Chowdhury S, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, Simuni T, Jennings D, Tanner CM, Trojanowski JQ, Shaw LM, Seibyl J, Schuff N, Singleton A, Kieburtz K, Toga AW, Mollenhauer B, Galasko D, Chahine LM, Weintraub D, Foroud T, Tosun-Turgut D, Poston K, Arnedo V, Frasier M, Sherer T. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. Ann Clin Transl Neurol. 2018;5(12):1460–77. https://doi.org/10.1002/acn3.644.
117. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, Masters C, Milner A, Pike K, Rowe C, Savage G, Szoeke C, Taddei K, Villemagne V, Woodward M, Ames D. The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr. 2009;21(4):672–87. https://doi.org/10.1017/S1041610209009405.
118. Xiao S, Lewis M, Mellor D, McCabe M, Byrne L, Wang T, Wang J, Zhu M, Cheng Y, Yang C, Dong S. The China longitudinal ageing study: overview of the demographic, psychosocial and cognitive data of the Shanghai sample. J Ment Health. 2016;25(2):131–6. https://doi.org/10.3109/09638237.2015.1124385.
119. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci. 2007;19(9):1498–507. https://doi.org/10.1162/jocn.2007.19.9.1498.
120. Juntu J, Sijbers J, Dyck D, Gielen J. Bias field correction for MRI images. In: Advances in Soft Computing. Springer, Berlin Heidelberg. p. 543–51. https://doi.org/10.1007/3-540-32390-2_64.

121. UCLA Brain Mapping Center. ICBM template. 2023. http://www.bmap.ucla.edu/portfolio/atlases/ICBM_Template/. Accessed 3 Mar 2023.
122. Jia H, Wang Y, Duan Y, Xiao H. Alzheimer's disease classification based on image transformation and features fusion. Comput Math Methods Med. 2021;2021:1–11. https://doi.org/10.1155/2021/9624269.
123. Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M. Automated anatomical labelling Atlas 3. Neuroimage. 2020;206:116189. https://doi.org/10.1016/j.neuroimage.2019.116189.
124. Magnetic Resonance - Technology Information Portal. Eigenvalues. 2023. https://www.mr-tip.com/serv1.php?type=db1&dbs=Eigenvalues. Accessed 3 Mar 2023.
125. Centre for Human Neuroimaging. SPM. 2023. https://www.fil.ion.ucl.ac.uk/spm/. Accessed 3 Mar 2023.
126. Center for Information Technology. MIPAV. 2023. https://mipav.cit.nih.gov/. Accessed 3 Mar 2023.
127. Yan C-G, Wang X-D, Zuo X-N, Zang Y-F. DPABI: data processing and analysis for (resting-state) brain imaging. Neuroinformatics. 2016;14(3):339–51. https://doi.org/10.1007/s12021-016-9299-4.
128. Wang J, Wang X, Xia M, Liao X, Evans A, He Y. GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. Front Hum Neurosci. 2015;9. https://doi.org/10.3389/fnhum.2015.00386.
129. National Institutes of Health, Medical College of Wisconsin. MIPAV. 2023. https://afni.nimh.nih.gov/. Accessed 3 Mar 2023.