REVIEW

Machine learning for classification and prediction of brain diseases: recent advances and upcoming challenges

Ninon Burgos[a,b,c,d,e] *and Olivier Colliot*[a,b,c,d,e]

Purpose of review

Machine learning is an artificial intelligence technique that allows computers to perform a task without being explicitly programmed. Machine learning can be used to assist diagnosis and prognosis of brain disorders. Although the earliest articles date from more than ten years ago, research increases at a very fast pace.

Recent findings

Recent works using machine learning for diagnosis have moved from classification of a given disease versus controls to differential diagnosis. Intense research has been devoted to the prediction of the future patient state. Although a lot of earlier works focused on neuroimaging as data source, the current trend is on the integration of multimodal data. In terms of targeted diseases, dementia remains dominant but approaches have been developed for a wide variety of neurological and psychiatric diseases.

Summary

Machine learning is extremely promising for assisting diagnosis and prognosis in brain disorders. Nevertheless, we argue that key challenges remain to be addressed by the community for bringing these tools in clinical routine: good practices regarding validation and reproducible research need to be more widely adopted; extensive generalization studies are required; interpretable models are needed to overcome the limitations of black-box approaches.

Keywords

artificial intelligence, classification, machine learning, prediction, translational research

INTRODUCTION

Artificial intelligence has witnessed tremendous progress in the past decade. As in other fields of medicine, there is a major interest in using artificial intelligence for assisting management of brain disorders. Artificial intelligence applications include assisting diagnosis, providing prognosis information, and predicting response to treatment. These techniques have potential value both in expert centers and in community-based practice. In expert centers, they may enrich the set of information which is available to the clinician, for instance providing accurate prognostic information which is currently out of reach in most diseases. In community-based practice, one can expect that artificial intelligence will assist the early detection of diseases and their referral to expert centers.

Among artificial intelligence approaches, the one which has led to the most impressive advances of the past years is machine learning. Machine learning allows a computer to perform a task (for instance, finding the diagnosis of a patient) without an explicit implementation of the underlying procedure. Instead, the computer will learn by examining a set of data called the training set. It is given a generic model, which parameters are adjustable. The optimal values of the parameters are then automatically estimated from the training data.

The first landmark publications on the use of machine learning in brain disorders date from more than ten years [1–3]. However, the field has made important progress since then. Computer-aided

---

[a]Paris Brain Institute, ICM, [b]Inserm, U 1127, [c]CNRS, UMR 7225, [d]Sorbonne Université and [e]Inria Paris, Aramis project-team, Paris, France

Correspondence to Olivier Colliot, PhD, ICM – Paris Brain Institute, ARAMIS Lab, Pitié-Salpêtrière Hospital, 47-83, boulevard de l'Hôpital, 75651 Paris Cedex 13, France. E-mail: olivier.colliot@upmc.fr

Curr Opin Neurol 2020, 33:439–450

DOI:10.1097/WCO.0000000000000838

KEY POINTS

- Machine learning allows a computer to perform a task, such as finding the diagnosis of a patient, without an explicit implementation of the underlying procedure.
- Computer-aided diagnosis has moved from the discrimination between a single disease and controls to differential diagnosis.
- Machine learning can also predict the future state of a patient (future diagnosis or cognitive/clinical score).
- Recent works are focusing on the integration of multimodal data, including neuroimaging, clinical/cognitive data, genomic, and other measures.
- There is still an important gap between research results and what can be implemented in the clinic, because estimation of performance is not always adequately performed and because the results can be difficult to reproduce, generalize, and interpret.

diagnosis has moved from the discrimination between a single disease and controls to differential diagnosis [4▪,5,6,7,8]. In addition to diagnosis, models for predicting the subsequent evolution of patients have been developed [9▪,10▪,11▪,12,13]. Most of the initial work focused on neuroimaging as the data source, because it is inherently digital and databases are easy to access. Recent works are focusing on the integration of multimodal data, including clinical/cognitive data, genomic, and other measures [9▪,10▪,14–17].

Nevertheless, there is still an important gap between research results and what can be implemented in the clinic. Estimation of performance is not always adequately performed in publications [9▪,18,19▪,20▪▪]. The results are sometimes difficult, if not impossible, to reproduce by others. The ability of the technique to generalize from highly controlled research data to routine data can be questioned. Finally, interpretability, that is, the ability to understand why the machine learning model takes a given decision, is a key issue [21▪,22,23].

In this article, we review recent progresses in the use of machine learning for management of brain diseases. We first briefly introduce the main concepts in a way that is accessible to neurologists. We then describe the main applications that have been developed. Moreover, we highlight recent advances on the integration of multimodal data, including neuroimaging, genetic, and clinical/cognitive data. Finally, we discuss the challenges that remain to be addressed for translation to the clinic. Note that this is not an exhaustive review but a presentation and discussion of recent data and articles.

MACHINE LEARNING CONCEPTS

Although a detailed introduction to machine learning is beyond the scope of the present article, it seems useful to clarify the main concepts and terms. These concepts are summarized in Fig. 1 and the main terms are defined in Table 1. Machine learning is an artificial intelligence technique in which the computer learns from data or experience. Machine learning techniques look for statistical patterns in the data. Other artificial intelligence techniques (e.g., symbolic approaches) exist but most of the recent successes have been based on machine learning.

In this review, we will consider machine learning techniques that aim to predict an *output* y from an *input* x. The input is typically a set of data characterizing a patient (e.g., an MRI scan, a set of cognitive test results, a set of genetic variants, etc.). The output can be a diagnostic category (e.g., Alzheimer's disease versus Lewy-body dementia versus vascular dementia) and one then deals with a classification problem. The output can also be a clinical/cognitive score or even an MRI image and one deals with a regression problem.

Learning aims at finding the best *model* f *that maps* x *to* y. This is done by analyzing a set of data, called the training set, to find a model that minimizes the error between the predicted output and the true output. Solely minimizing the error is often nonoptimal as it would produce a perfect prediction on the training set but a poor one on new data, a phenomenon called *overfitting*. Other machine learning pitfalls and their solutions are presented in Table 2.

Many choices are possible for the type of model f. One can cite different machine learning techniques, including for instance, logistic and linear regression (possibly with penalties), support vector machines (SVM), random forests, and deep learning. Deep learning deserves a specific mention as it led to some of the most impressive recent advances. In deep learning, the model f is made of a very large set of artificial neurons, organized into layers learning a hierarchy of representations. But keep in mind that there is more to artificial intelligence than machine learning and there is more to machine learning than deep learning.

MACHINE LEARNING FOR CLASSIFICATION AND PREDICTION

This section describes the main recent applications of machine learning to automatic classification and prediction of neurological diseases. The reviewed studies are summarized in Table 3.

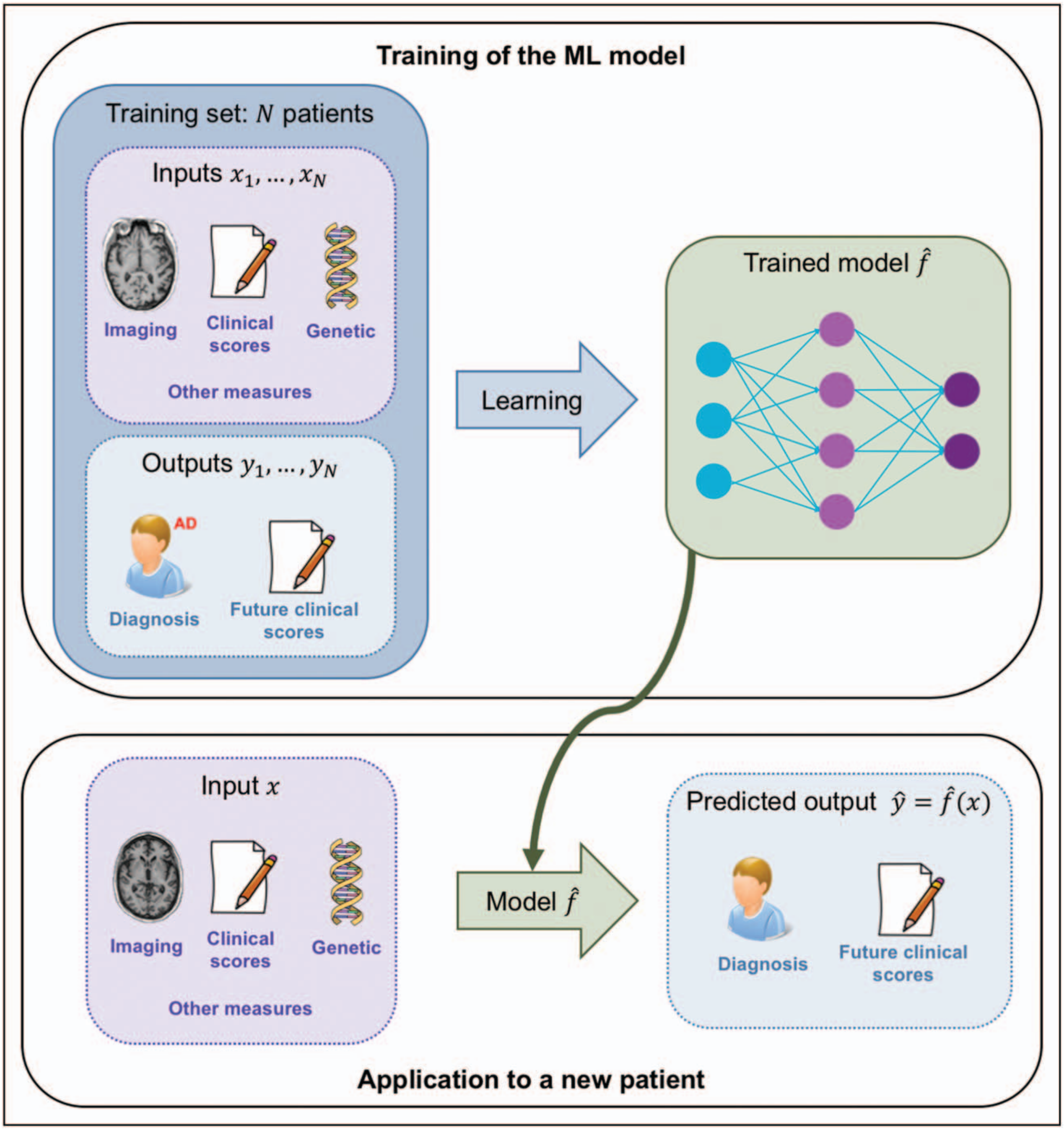
Training of the ML model
Training set: N patients
Inputs $x_1, \dots, x_N$
Imaging
Clinical scores
Genetic
Other measures
Outputs $y_1, \dots, y_N$
AD
Diagnosis
Future clinical scores
Learning
Trained model $\hat{f}$
Input x
Imaging
Clinical scores
Genetic
Other measures
Model $\hat{f}$
Predicted output $\hat{y} = \hat{f}(x)$
Diagnosis
Future clinical scores
Application to a new patient

FIGURE 1. Machine learning concepts. There are two main phases in building and evaluating a machine learning model. The training (upper panel) aims at building the model. For that, one uses a training set comprising both inputs and outputs for a set of N patients. Input data can be of any kind, for instance neuroimaging scans (MRI, PET), clinical/cognitive scores, genotyping, or the combination of those. Output data can also be of various kinds, for instance a diagnostic category, the future value of a clinical/cognitive score or even a neuroimaging scan of the patient. The learning phase estimated the function $\hat{f}$ that best transforms the input to the output data across the different patients of the training set. In the second phase, the estimated model is applied to new input data to predict the output.

Assisting diagnosis

The most common use of machine learning is probably computer-assisted diagnosis. Early works have tackled automatic classification of Alzheimer's disease [2,3] and schizophrenia [1] from anatomical MRI data. Since then, hundreds of articles have proposed automatic classification approaches for different brain disorders and based on different types of data. Some recent works include classification of epilepsy with hippocampal sclerosis [24▪],

Table 1. Main terms used in machine learning

- **Machine learning**.Techniques that can make a computer perform a task without being explicitly programmed for. Instead, the computer learns a model by examining a set of examples called the training set
- **Model**. The mathematical function that transforms the inputs (e.g., cognitive scores, imaging data) into outputs (e.g., a diagnosis)
- **Supervised learning**. Type of learning in which the training data must contain both inputs and outputs. The model is trained by examining examples for which the desired output is provided
- **Unsupervised learning**. Type of learning in which the training data only contain inputs. This is for example used for finding disease subtypes which are currently unknown
- **Classification**. Supervised learning technique in which the outputs are classes (e.g., diagnostic classes)
- **Regression**. Supervised learning techniques in which the outputs are continuous values (e.g., cognitive scores)
- **Support vector machine (SVM)**. A supervised machine learning technique which is often a good choice for high-dimension/small sample size problems. It can be linear or nonlinear. Is mainly used for classification although extensions to regression exist
- **Random forests**. A supervised machine learning technique which assembles a large number of decision trees. Can be used both for classification and regression
- **Deep learning**. Type of machine learning in which the model is a set of artificial neurons which are arranged into a large number of layers. The number of layers is the depth of the model. This is a large family of models which include both supervised and unsupervised techniques
- **Convolutional neural network**. A deep learning technique in which a given neuron is connected to its neighbors, performing a mathematical operation called convolution. This technique is mainly used for imaging data
- **Hyperparameters**. Parameters that modify the behavior of the model or of the training procedures. They are called hyper-parameters by contrast with the model parameters. They are not learned but are set or tuned
- **Training set/validation set/test set**. The training set is the set of subjects/patients which is used to train the model. The test set is the set of patients which is used to evaluate the performance. Often, one also adds a validation set which allows to determine when to stop the training procedure or to tune hyperparameters. This is not needed if the training procedure can be stopped automatically or if there are no hyperparameters. It is crucial that the three sets are disjoint to have an unbiased evaluation of the performance
- **Cross-validation**. A procedure in which the training and validation/testing set are iteratively exchanged. For instance, one can split the dataset into five-folds and, at each of the five iterations, four of the folds are used as the training set and one-fold is used as the testing set
- **Data leakage**. A bad practice in which information from the training set is also in the testing/validation set. This leads to an over-optimistic evaluation of performance<
- **Generalization**. The ability of a model training on a given dataset to perform well on another dataset
- **Features**. Characteristics used by the model to perform the task. The characteristics can be directly the input data or can be extracted from the input data. For example, when using anatomical MRI data, one can extract the volumes of different anatomical regions of the brain
- **Feature extraction**. The procedure that computes the features from the input data
- **Feature selection**. A procedure to select which of the features will be actually used as inputs of the model. One aims to select the features that will be the most predictive. This can for instance be done using univariate statistical tests or more complex multivariate procedures
- **Feature transformation**. A procedure to reduce the number of features by embedding them into a lower dimensional space
- **Performance metrics**. Measures used to assess the performance of the model. Classical examples for classification include accuracy, balanced accuracy, area under the curve, sensitivity, specificity, positive and negative values. Classical examples for regression include mean squared error and mean absolute error
- **Data augmentation**. Family of techniques that allow to increase the size of the training set through the generation of new training samples from existing ones. This can be done for instance by applying simple transformations (symmetry, translation, etc.) to the original data or using complex generative models

Table 2. Pitfalls and solutions

- **Small size of training set**. Having too few training samples often leads to a model with low performance. Data augmentation (see Table 1) may be used to mitigate this issue
- **Small size of the testing set**. Having too few testing samples will lead to unreliable estimates of the performance. The only solution to this is to have more samples in the testing set
- **Overfitting**. Overfitting corresponds to a situation where the model is fitting the training set too well and will poorly generalize to the test set: the model has learnt the training samples 'by heart.' Overfitting may typically occur when the model is too complex or when there are too many output features. Possible solutions include: adding some regularization to the model, reducing the number of features by performing feature selection or feature transformation, applying specific techniques for deep learning models (early stopping, dropout)
- **Data leakage**. Data leakage is a bad practice in which information from the training set is also in the testing set. This leads to an over-optimistic evaluation of performance. Possible sources of data leakage include: wrong split of the training and testing sets (for instance, having some visits of a given patient in the training set and some in the testing set), testing different models and hyperparameters using the testing set. The testing set should be separated from the very beginning of the study and left untouched until the final evaluation of the model
- **Inadequate performance metrics**. It is important to choose metrics that are adapted to the task at-hand. For instance, accuracy is inadequate when dealing with unbalanced datasets (different number of patients in each group) and balanced accuracy should be preferred

Table 3. Summary of the reviewed studies

	First author	Source title	Publication year	Disease	Database	Number of patients per class	Modalities used as features	Classification/regression algorithm
[17]	Peng *et al.*	*MICCAI*	2016	Alzheimer's disease	ADNI[b]	93 MCI, 49 Alzheimer's disease, 47 healthy control	T1w MRI, FDG PET, genetic	MKL
[21■]	Böhle *et al.*	*Front Aging Neurosci*	2019	Alzheimer's disease	ADNI[b]	193 Alzheimer's disease, 151 healthy control	T1w MRI	CNN
[22]	Ding *et al.*	*Radiology*	2019	Alzheimer's disease	ADNI[b], local	413 MCI, 243 Alzheimer's disease, 386 healthy control	FDG PET	CNN
[23]	Thibeau-Sutre *et al.*	*SPIE MI*	2020	Alzheimer's disease	ADNI[b], AIBL[c]	412 Alzheimer's disease, 759 healthy control,	T1w MRI	CNN
[32]	Li *et al.*	*Alzheimers Dement*	2018	Alzheimer's disease	ADNI[b]	511 sMCI, 292 pMCI	T1w MRI, FDG PET, CSF, genetic, neuropsychological tests, demographics	Cox proportional hazards model
[47]	Li *et al.*	*Alzheimers Dement*	2019	Alzheimer's disease	ADNI[b], AIBL[c]	862 MCI, 417 Alzheimer's disease, 867 healthy control	T1w MRI, genetic, neuropsychological tests, demographics	LASSO regularized Cox proportional hazards model
[3]	Gerardin *et al.*	*NeuroImage*	2009	Alzheimer's disease, MCI	Local	23 aMCI, 23 Alzheimer's disease	T1w MRI	Linear SVM
[9■]	Ansart *et al.*[a]	*Preprint*	2019	Alzheimer's disease, MCI	–	–	–	–
[10■]	Samper-Gonzalez *et al.*	*SPIE MI*	2019	Alzheimer's disease, MCI	ADNI[b]	507 MCI, 126 Alzheimer's disease, 115 healthy control	T1w MRI, FDG PET, genetic, neuropsychological tests	Random forests, linear SVM
[11■]	Marinescu *et al.*	*Preprint*	2018	Alzheimer's disease, MCI	ADNI[b]	1095 MCI, 257 Alzheimer's disease, 646 healthy control	T1w MRI, DWI, FDG PET, amyloid PET, tau PET, CSF, neuropsychological tests, genetic, demographics	–
[15]	Sørensen *et al.*	*J Neurosci Methods*	2018	Alzheimer's disease, MCI	ADNI[b]	100 MCI, 100 pMCI, 100 Alzheimer's disease, 100 healthy control	T1w MRI, neuropsychological tests, demographics	Ensemble linear and RBF SVM
[16]	Lu and Colliot	*SPIE MI*	2020	Alzheimer's disease, MCI	ADNI[b]	154 sMCI, 172 pMCI	T1w MRI, genetic	Cox proportional hazards model
[19■]	Wen *et al.*	*Medical Image Analysis*	2020	Alzheimer's disease, MCI	ADNI[b], AIBL[c], OASIS[d]	880 MCI, 490 Alzheimer's disease, 835 healthy control	T1w MRI	CNN
[33]	Koval *et al.*	*Front Neurol*	2018	Alzheimer's disease, MCI	ADNI[b]	154 pMCI	T1w MRI	Mixed-effects model
[34]	Schiratti *et al.*	*J Mach Learn Res*	2017	Alzheimer's disease, MCI	ADNI[b]	248 pMCI	Neuropsychological tests	Mixed-effects model
[35]	Moradi *et al.*	*NeuroImage*	2015	Alzheimer's disease, MCI	ADNI[b]	100 sMCI, 164 pMCI, 130 uMCI, 200 Alzheimer's disease, 231 healthy control	T1w MRI, neuropsychological tests	Low-density separation
[37]	Da *et al.*	*NeuroImage Clin*	2014	Alzheimer's disease, MCI	ADNI[b]	381 MCI, 200 Alzheimer's disease, 232 healthy control	T1w MRI, CSF, genetic, neuropsychological tests	Cox proportional hazards model
[38]	Gupta *et al.*	*Front Comput Neurosci*	2019	Alzheimer's disease, MCI	ADNI[b]	36 sMCI, 46 pMCI, 38 Alzheimer's disease, 38 healthy control	T1w MRI, FDG PET, CSF, genetic	Nonlinear SVM
[39]	Khanna *et al.*	*Sci Rep*	2018	Alzheimer's disease, MCI	ADNI[b]	609 MCI, 315 healthy control	T1w MRI, FDG PET, genetic, neuropsychological tests, diagnosis, demographics	Gradient boosting machine
[40]	Varatharajah *et al.*	*Sci Rep*	2019	Alzheimer's disease, MCI	ADNI[b]	96 sMCI, 39 pMCI	T1w MRI, FDG PET, amyloid PET, CSF, genetic, neuropsychological tests, demographics	SVM, MKL, GLM with elastic-net regularization
[41■■]	Samper-González *et al.*	*NeuroImage*	2018	Alzheimer's disease, MCI	ADNI[b], AIBL[c], OASIS[d]	962 MCI, 514 Alzheimer's disease, 953 healthy control	T1w MRI, FDG PET	Linear SVM, LR with L2 regularization, random forests

Table 3 (*Continued*)

	First author	Source title	Publication year	Disease	Database	Number of patients per class	Modalities used as features	Classification/regression algorithm
[43]	Rathore *et al.*[a]	*NeuroImage*	2017	Alzheimer's disease, MCI	–	–	–	–
[5]	Bouts *et al.*	*J Alzheimers Dis*	2018	Alzheimer's disease, bvFTD	Local	30 Alzheimer's disease, 23 bvFTD, 35 healthy control	T1w MRI, DWI, rs-fMRI	Elastic net regression
[7]	Zheng *et al.*	*Front Neurol*	2019	Alzheimer's disease, VaD	Local	58 Alzheimer's disease, 35 VaD	T1w MRI, FLAIR	kNN, LR, random forests, linear SVM, RBF SVM
[4■]	Morin *et al.*	*J Alzheimers Dis*	2020	Alzheimer's disease, CBD, depression, bvFTD, LBD, lvPPA, svPPA	Local	31 CBD, 24 depression, 34 early Alzheimer's disease, 39 FTD, 22 LBD, 49 late Alzheimer's disease, 23 lvPPA, 17 svPPA, 12 SCD	T1w MRI	Linear SVM, univariate classification
[2]	Klöppel *et al.*	*Brain*	2008	Alzheimer's disease, depression, FTD, LBD	ADNI[b], AIBL[c], local	388 Alzheimer's disease, 61 depression, 39 FTD, 23 LBD, 586 healthy control	T1w MRI	Linear SVM
[6]	Tong *et al.*	*NeuroImage Clin*	2017	Alzheimer's disease, FTLD, LBD, VaD	Amsterdam dementia cohort[e]	219 Alzheimer's disease, 92 FTLD, 47 LBD, 24 VaD, 118 SMC	T1w MRI, FLAIR, CSF	RUSBoost
[28]	Meyer *et al.*	*NeuroImage Clin*	2017	bvFTD	German consortium for FTLD[f], local	52 bvFTD, 52 healthy control	T1w MRI	SVM
[27]	Feis *et al.*	*NeuroImage Clin*	2018	FTD	Local	55 presymptomatic FTD mutation carriers, 48 familial healthy control	T1w MRI, DWI, rs-fMRI	Elastic net regression
[46]	Bisenius *et al.*	*NeuroImage Clin*	2017	lvPPA, nfvPPA, svPPA	German consortium for FTLD[f]	16 nfvPPA, 17 svPPA, 11 lvPPA, 20 healthy control	T1w MRI	Linear SVM
[14]	Qiu *et al.*	*Alzheimers Dement Diagn Assess Dis Monit*	2018	MCI	National Alzheimer's Coordinating Center[g]	83 MCI, 303 healthy control	T1w MRI, FLAIR, neuropsychological tests	CNN, ANN
[45■■]	Bouts *et al.*	*Hum Brain Mapp*	2019	MCI	Rotterdam study[h], local	48 MCI, 77 Alzheimer's disease, 790 healthy control	T1w MRI, DWI	Elastic net regression
[31]	van der Burgh *et al.*	*NeuroImage Clin*	2017	Amyotrophic lateral sclerosis	Local	135 sporadic ALS (52 short, 52 medium and 31 long survivors)	T1w MRI, DWI, clinical data	ANN
[8]	Péran *et al.*	*Mov Disord*	2018	Parkinson's disease, multiple system atrophy	Local	26 Parkinson's disease, 29 MSA, 26 healthy control	T1w MRI, T2 relaxometry, DWI	LR
[13]	Zhang *et al.*	*NeuroImage Clin*	2019	Multiple sclerosis	Local	65 CIS to multiple sclerosis converters, 18 CIS non converters	T1w MRI, FLAIR	Random forests
[25]	Saccà *et al.*	*Brain Imaging Behav*	2019	Multiple sclerosis	Local	18 multiple sclerosis, 19 healthy control	rs-fMRI	Random forests, RBF SVM, Naïve Bayes, kNN, ANN
[26]	Zurita *et al.*	*NeuroImage Clin*	2018	MS	Local	104 RRMS, 46 healthy control	DWI, rs-fMRI	Linear SVM
[24■]	Chen *et al.*	*Brain Imaging Behav*	2019	Epilepsy	Local	16 left MTLE-HS, 6 right MTLE-HS, 15 healthy control	T1w MRI	RBF SVM
[12]	Ho *et al.*	*IEEE Trans Med Imaging*	2019	Stroke	Local	85 TSS $<$ 4.5 h, 46 TSS $\geq$ 4.5 h	PWI, DWI, FLAIR	logistic regression, random forests, gradient boosted regression tree, SVM, stepwise multilinear regression, CNN
[20■■]	Pulini *et al.*[a]	*Biol Psychiatry Cogn Neurosci Neuroimaging*	2019	ADHD	–	–	–	–
[30]	Chen *et al.*	*Artif Intell Med*	2020	ADHD	ADHD-200 consortium[i]	248 ADHD, 299 healthy control	rs-fMRI	Subspace learning
[1]	Fan *et al.*	*MICCAI*	2005	Schizophrenia	Unknown	23 schizophrenia, 38 healthy control	T1w MRI	RBF SVM

Table 3 (*Continued*)

	First author	Source title	Publication year	Disease	Database	Number of patients per class	Modalities used as features	Classification/regression algorithm
[29]	Zeng *et al.*	*EBioMedicine*	2018	Schizophrenia	Local, OpenfMRI[j]	474 schizophrenia, 607 healthy control	rs-fMRI	Linear SVM
[44■]	Cai *et al.*	*Hum Brain Mapp*	2020	Schizophrenia	Local	85 schizophrenia, 78 healthy control	rs-fMRI	Linear discriminant analysis

Diseases: ADHD, Attention deficit hyperactivity disorder; ALS, amyotrophic lateral sclerosis; aMCI, amnestic MCI; bvFTD, behavioral variant FTD; CBD, corticobasal degeneration; CIS, clinically isolated syndrome; FTD, frontotemporal dementia; FTLD, frontotemporal lobar degeneration; LBD, Lewy body dementia; lvPPA, logopenic variant PPA; MCI, mild cognitive impairment; MSA, multiple system atrophy; MTLE-HS, mesial temporal lobe epilepsy with hippocampal sclerosis; nfvPPA, nonfluent variant PPA; pMCI, progressive MCI; PPA, primary progressive aphasia; RRMS, relapsing-remitting MS; SCD, subjective cognitive decline; SMC, subjective memory complaints; sMCI, stable MCI; svPPA, semantic variant PPA; TSS, time-since-stroke; uMCI, unknown MCI; VaD, vascular dementia. *Modalities*: CSF, Cerebrospinal fluid; DWI, diffusion weighted imaging; FDG, ^{18}F fluorodeoxyglucose; FLAIR, fluid-attenuated inversion recovery; PET, positron emission tomography; PWI, perfusion weighted imaging; rs-fMRI, resting state functional magnetic resonance imaging; T1w, T1-weighted. *Algorithms*: ANN, artificial neural network; CNN, convolutional neural network; GLM, generalized linear model; kNN, *k*-nearest neighbors; MKL, multiple kernel learning; RBF, radial basis function; SVM, support vector machine.

[a] Review.

[b] Alzheimer's Disease Neuroimaging Initiative (http://adni.loni.usc.edu).

[c] Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (https://aibl.csiro.au).

[d] Open Access Series of Imaging Studies (https://www.oasis-brains.org).

[e] https://www.alzheimercentrum.nl/wetenschap/amsterdam-dementia-cohort.

[f] http://www.ftld.de.

[g] https://www.alz.washington.edu.

[h] http://www.erasmus-epidemiology.nl/research/ergo.htm.

[i] http://fcon_1000.projects.nitrc.org/indi/adhd200/.

[j] https://openfmri.org.

multiple sclerosis [25,26], frontotemporal dementia (FTD) [27,28], schizophrenia [29], and attention deficit hyperactivity disorder [30]. In most of these works, the classification distinguishes patients with a given disease from healthy controls. This can have value for assisting in the detection of diseases which are difficult to detect and diagnose, in particular outside of expert clinical centers, and if the tool is sensitive at an early disease stage.

However, comparison to healthy controls often does not correspond to a clinically realistic situation, where difficult diagnoses are between different diseases that may present similarly. This is the case for instance for distinguishing between Parkinson's disease and Parkisonian syndromes, or between different types of dementia. Péran *et al.* [8] were able to discriminate between Parkinson's disease and multiple system atrophy with very high accuracy (area under the receiver operating characteristics curve (AUC) $>95\%$) using different MRI techniques (anatomical, diffusion, T2$*$ relaxometry). Some studies have looked at differentiation between two types of dementia including Alzheimer's disease from FTD [5] and Alzheimer's disease from vascular dementia [7]. Tong *et al.* [6] studied differential diagnoses between the four most common dementias (Alzheimer's disease, FTD, vascular dementia, dementia with Lewy bodies) and patients with subjective memory complaints (SMC), and achieved high five-class accuracy around 70% based on T1 and FLAIR MRI as well as cerebrospinal fluid biomarkers. Morin *et al.* [4■] included eight different cognitive conditions (Alzheimer's disease, FTD, dementia with Lewy bodies, logopenic, and semantic primary progressive aphasias, cortical-basal syndrome, depression, SMC). Both studies found high accuracies for diseases which have clear MRI alteration patterns (Alzheimer's disease, FTD, semantic dementia) but not for others (e.g., dementia with Lewy bodies). These tools have potential clinical utility for difficult differential diagnoses (for instance, early-onset Alzheimer's disease versus FTD). Importantly, they were assessed using clinical routine MRI data making their application realistic.

Predicting evolution

Although the previous approaches propose to classify patients based on their current data, other works aim at predicting the future state of a patient based on its baseline data (or based on longitudinal data from visits before that of the predicted outcome). Here, the output can be a future diagnosis (e.g., future diagnosis of dementia within patients with mild impairment at baseline) or future relevant measures (e.g., the future value of a cognitive/clinical score).

One way to frame the problem of predicting a future state is as a classification task. One sets a temporal horizon (e.g., 24 months) and aims at discriminating between patients who reached this state (e.g., became demented) before or at this temporal horizon and those who did not. This task has been widely addressed in the case of predicting progression to Alzheimer's disease among patients with mild cognitive impairment (MCI) at baseline. A recent review identified 172 articles on that specific topic [9▪]. The best AUCs obtained by well-powered studies are typically around 0.80–0.85. Predictive studies in other neurological disorders are less frequent. Zhang *et al.* [13] predicted progression from clinically isolated syndrome to multiple sclerosis using characteristics of MRI white matter lesions with a balanced accuracy of 72%. Instead of predicting the future, one can also use machine learning to go back in time and estimate the date at which the disease started. For instance, Ho *et al.* [12] distinguished patients with time-since-stroke lower or higher than 4.5 h from diffusion, perfusion, and FLAIR MRI.

Instead of fixing a temporal horizon, one may aim to determine the time at which the event of interest will occur. Such works are not overly common thus far but are receiving increased interest. In the case of Alzheimer's disease, this is one of the aims of the TADPOLE challenge [11▪] in which participating researchers must predict future diagnoses, cognitive test values (Alzheimer's Disease Assessment Scale-Cognitive Subscale) and MRI measures (ventricular volume) at each month over five years. Predicting time-of-event can be done using classification techniques. For instance, this has been applied to predict survival in amyotrophic lateral sclerosis from MRI [31]. But this requires to set arbitrary times and prevents from using censored data. A more adapted framework is that of survival analysis, a classical statistical technique which was used to predict the time of progression to Alzheimer's disease [32] and was extended to high-dimensional MRI and genetic data in [16]. Finally, generative models allow predicting a full sequence of future measures (e.g., the sequence of future cognitive scores or brain images) [33,34].

USING MULTIMODAL DATA: IMAGING, GENETIC, CLINICAL

Without doubt neuroimaging is the modality which has been the most widely used in machine learning works. However, it seems more than natural that a more comprehensive characterization of the patient would lead to better predictions.

Combining imaging and clinical/cognitive data

The most natural candidates for use as input of machine learning methods are clinical and cognitive scores. They are the core of diagnosis in many situations and are inexpensive to acquire. Of course, using them if the predicted outcome is clinical diagnosis at the same time point would lead to a circular analysis. However, there are of interest in all other situations (e.g., when a future clinical diagnosis is predicted or when the outcome is based on postmortem examinations). An important result of a recent systematic review on Alzheimer's disease prediction [9▪] was that using clinical/cognitive data significantly improved predictions compared with not including them, whereas this was not the case of anatomical MRI, even though the latter had been the subject of intense research. Some studies have combined different imaging modalities with clinical/cognitive scores [14,15,31,35]. A recent large-scale study showed that clinical/cognitive scores were better than T1 MRI or ^{18}F fluorodeoxyglucose positron emission tomography at predicting progression to Alzheimer's disease in patients with MCI and that the combination of all modalities further improved the results, with up to 89% AUC [10▪].

Combining imaging and genetic data

Genetic factors modulate the disease risk and the evolution. Association studies between imaging and genetic data have exploded in the past decade [36] but their combination for disease prediction has been more limited. Several studies combined brain imaging with a single gene (e.g., apolipoprotein E in Alzheimer's disease) for disease prediction [37,38]. Current works aim at integrating imaging with multiple genes or genome-wide information [16,17,39] as well as with gene expression data [40]. Nevertheless, the identification of relevant genetic variables requires very high sample sizes (typically 10 000–100 000). Major progresses in the identification of genetic variants and polygenic risk scores have stemmed from the analysis of large population cohorts (e.g., UKBiobank) or metaanalysis (e.g., ENIGMA). Combining these results with smaller samples on specific brain diseases may lead to improved predictions.

RELIABLE MACHINE LEARNING: THE PATH TO THE CLINIC

Even though a large number of methods are being developed to assist diagnosis, only a few are translated to the clinic. This can be explained by different

Table 4. Challenges for accelerating clinical translation

•**Rigorous validation**. Bad practices remain too common in the litterature. Data leakage (see Table 1) can be insidious. This leads to over-optimistic evaluations of performance
•**Reproducibility**. Code should be shared. Data also whenever possible. Standardized community practices need to be adopted for data management and preprocessing
•**Generalizability**. Most studies are based on highly controlled research data. More studies with routine clinical data are needed. Generalization to widely variable settings (different hospitals, different acquisition devices, different populations) is critical
•**Interpretability**. Ability to understand the decision of a machine learning method is key for safe use and for adoption by clinicians. Methodological advances are needed in this area
•**Workflow integration**. This aspect has been overlooked by most studies. It is crucial to identify at which steps of the clinical workflow machine learning should be used and how. We also need to reflect on its place in the patient's journey through the healthcare system

factors, summarized in Table 4. Note that the present review does not cover regulatory aspects (e.g., FDA clearance) and those related to the data itself (data ownership, security, and privacy).

Reproducibility

Reproducibility is defined as the ability to reproduce results based on the same data and methodology. This differs from replication, which is the ability to confirm results on independent data. Key elements of reproducible research include data sharing, fully automatic data manipulation and sharing of code. Without these elements, results cannot be reproduced, a step essential to guarantee the robustness of a technique. Initiatives have emerged to improve the reproducibility of machine learning and deep learning approaches applied to neuroimaging. Samper-González *et al.* [41▪▪] and Wen *et al.* [19▪] proposed a reproducible framework for the evaluation of Alzheimer's disease classification methods that comprise data management tools that rely on a community standard [42]; image preprocessing and feature extraction pipelines; standard classification algorithms and convolutional neural network models; and rigorous validation procedures. These tools are available in the open-source software platform Clinica (www.clinica.run).

Data leakage

Unbiased evaluation of machine learning and deep learning algorithms is critical to assess their potential clinical value. A major source of bias is data leakage, which refers to the use of test data in any part of the training process [18,43]. Several causes of data leakage exist and have been found in published works, as revealed in [19▪,20▪▪]. Not splitting the dataset at the subject-level when defining the training, validation, and test sets can result in data from the same subject to appear in several sets. This problem can occur when patches or slices are extracted from a 3D image, or when images of the same subject have been acquired at multiple time points. Performing procedures, such as feature selection or data augmentation, before the training/validation/test split means, in the case of feature selection, that the test set is used to select the most relevant features. The absence of an independent test set implies that the same data have been used to select the optimal hyperparameters of the method and evaluate the performance.

Generalizability

Assessing the ability of an approach to generalize from highly controlled research data of a certain cohort to another cohort, and more generally to routine data, is an essential step to ensure translation to the clinic. However, this step is rarely reached. Cai *et al.* [44▪] assessed the generalizability of an approach for the classification of schizophrenia based on resting-state functional MRI data. They showed that the predictive model did not successfully generalize to a novel dataset when directly applied and that an additional step enabling the model to adapt to the new dataset was necessary. A reduction in accuracy when the model is applied to a new dataset was also observed in [29]. Bouts *et al.* [45▪▪] pushed the analysis further by assessing whether an MRI-based classification method trained to detect MCI on a clinical cohort could be used on a general population. Even though the model could detect MCI better than chance, the classification performance was moderate and probably insufficient to efficiently assist diagnosis.

Interpretation

The ability to understand why the machine learning and deep learning models take a given decision is a key issue to facilitate their acceptance, know how far they can be trusted and achieve better performance. The main idea for image-based classification methods is to highlight the parts of the image that contribute the most to the decision. This can be

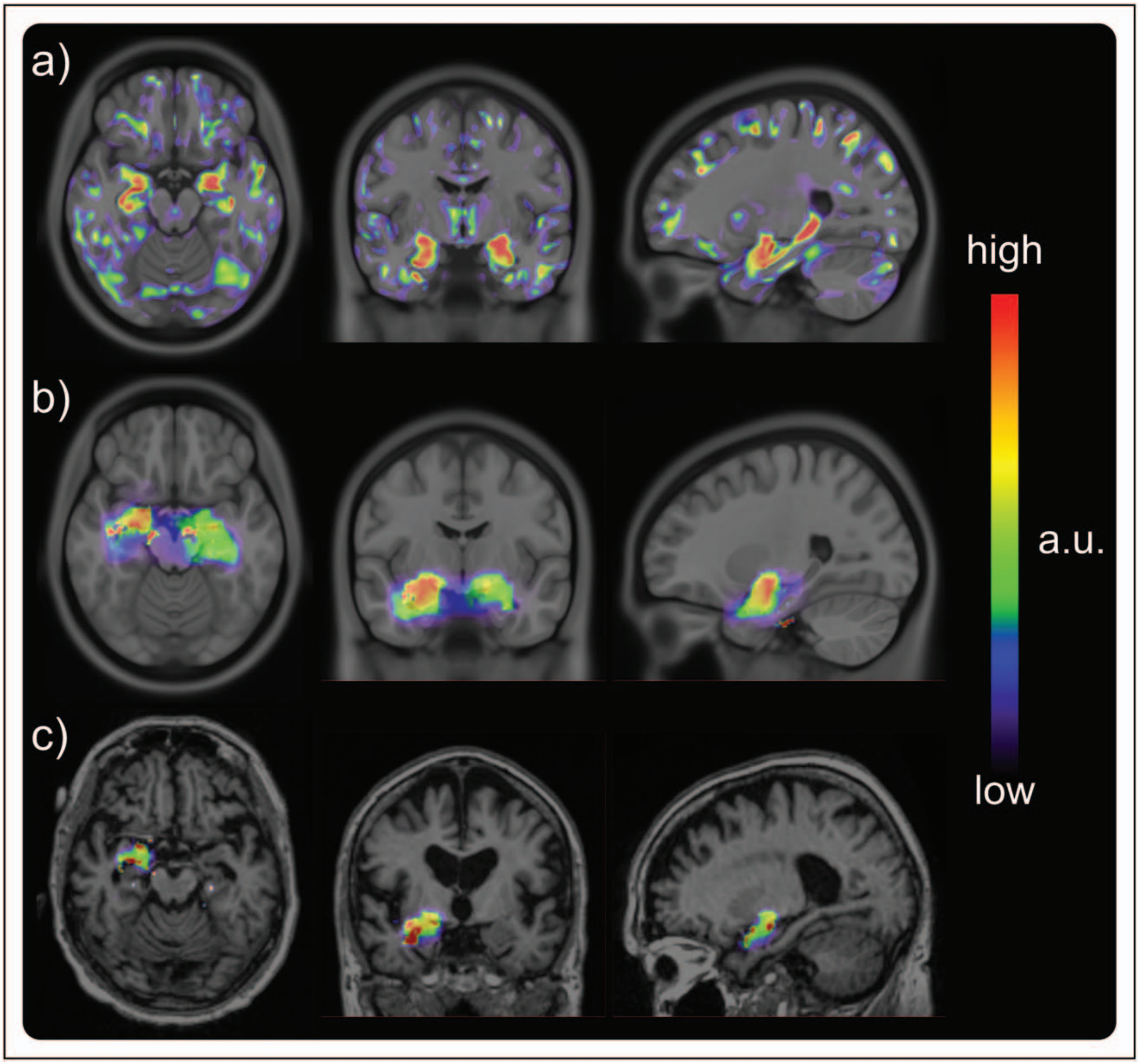

FIGURE 2. Interpretation of machine learning methods. The classifiers were trained to distinguish cognitively normal controls from patients with Alzheimer's disease. Highlighted regions are the ones that contribute the most to the decision. (a) Results obtained with a linear support vector machine (SVM) classifier [41■■]. Only voxels that contribute the most to the classification of subjects as patients (and not as controls) are displayed. (b,c) Results obtained with a convolutional neural network (CNN) classifier and a visualization approach that consists in optimizing a mask that will perturb the trained CNN so it will classify masked images in the wrong class [23]. The mask can be generated using the images of a group of patients (b), here Alzheimer's disease patients, thus highlighting regions that are in general relevant for the Alzheimer's disease classification, or using the image of a single patient (c), here an Alzheimer's disease patient, thus highlighting how the machine learning algorithm took its decision for this specific patient.

applied to machine learning models such as SVM [26,28,46], but also to deep learning models [22,47], see Fig. 2 for example. The level of interpretability greatly varies between the methods employed and new approaches, tailored to the task at hand, are being developed [21■,23].

Workflow integration

Another key question is how machine-learning tools should be integrated in the clinical workflow, in addition, more generally in the patient journey into the healthcare system. There are myriad of steps and tasks at which machine learning can potentially be used. This ranges from early screening for referral to experts centers, to diagnosis of difficult cases and treatment choice. Research is needed on which of these uses would be most beneficial for the patients and the healthcare system. This will need to be conducted for different neurological diseases.

CONCLUSION

Intense research has been conducted on the use of machine learning to assist diagnosis and prognosis of brain diseases. This has led to impressive results in research settings where data is highly controlled. Even though dementia (in particular Alzheimer's disease) is overly represented, promising results have also been obtained in many other diseases, including movement disorders, multiple sclerosis, epilepsy or psychiatric conditions. In spite of this, their translation to the clinic remains difficult. Specific challenges need to be addressed by the community for machine learning to be clinically applied. Of course, good practices of rigorous validation and reproducible research need to be adopted widely. Generalization studies where models are applied to a wide variety of clinical routine data (from different hospitals, different populations, different countries, etc.) are critically needed. Interpretation of machine learning remains an important methodological challenge. Finally, we need to determine how machine learning models can be integrated in the clinical workflow. This requires a tight collaboration between machine learning researchers and clinicians and a cross-dissemination of expertise from both fields.

Acknowledgements

None.

Financial support and sponsorship

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the 'Investissements d'avenir' program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), from the European Union H2020 program (project EuroPOND, grant number 666992), and from the Abeona Foundation (project Brain@Scale).

Conflicts of interest

Disclosure of interests related to the present article: None to disclose.

Disclosure of interests unrelated to the present article: O.C. reports having received consulting fees from AskBio (2020), having received fees for writing a lay audience short article from Expression Santé (2019), having received speaker fees for a lay audience presentation from Palais de la découverte (2017), and that his laboratory has received grants (paid to the institution) from Qynapse (2017–present). Members from his laboratory have cosupervised a PhD thesis with myBrainTechnologies (2016–present). O.C.'s spouse is an employee of myBrainTechnologies (2015–present). O.C. has submitted a patent to the International Bureau of the World Intellectual Property Organization (PCT/IB2016/0526993, Schiratti J-B, Allassonniere S, Colliot O, Durrleman S, A method for determining the temporal progression of a biological phenomenon and associated methods and devices) (2016).

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

■ of special interest
■■ of outstanding interest

1. Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv 2005; 8:1–8.

2. Klöppel S, Stonnington CM, Chu C, *et al.* Automatic classification of MR scans in Alzheimer's disease. Brain J Neurol 2008; 131:681–689.

3. Gerardin E, Chételat G, Chupin M, *et al.* Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. NeuroImage 2009; 47:1476–1486.

4. ■ Morin A, Samper-Gonzalez J, Bertrand A, *et al.* Accuracy of MRI classification algorithms in a tertiary memory center clinical routine cohort. J Alzheimers Dis 2020; 74:1157–1166.

Differential diagnosis study using clinical routine data.

5. Bouts MJ, Möller C, Hafkemeijer A, van Swieten JC, *et al.* Single subject classification of Alzheimer's disease and behavioral variant frontotemporal dementia using anatomical, diffusion tensor, and resting-state functional magnetic resonance imaging. J Alzheimers Dis 2018; 62:1827–1839.

6. Tong T, Ledig C, Guerrero R, *et al.* Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting. NeuroImage Clin 2017; 15:613–624.

7. Zheng Y, Guo H, Zhang L, *et al.* Machine learning-based framework for differential diagnosis between vascular dementia and Alzheimer's disease using structural MRI features. Front Neurol 2019; 10:1097.

8. Péran P, Barbagallo G, Nemmi F, *et al.* MRI supervised and unsupervised classification of Parkinson's disease and multiple system atrophy. Mov Disord 2018; 33:600–608.

9. ■ Ansart M, Epelbaum S, Bassignana G, *et al.* Predicting the progression of mild cognitive impairment using machine learning: a systematic and quantitative review. HAL 2019; Available from: https://hal.archives-ouvertes.fr/hal-02337815.

An exhaustive review on prediction of progression to Alzheimer's disease in patients with MCI, which brings important insights.

10. ■ Samper-Gonzalez J, Burgos N, Bottani S, *et al.* Reproducible evaluation of methods for predicting progression to Alzheimer's disease from clinical and neuroimaging data. In: Medical Imaging 2019: Image Processing International Society for Optics and Photonics; 2019. p. 109490V.

Demonstrates the importance of using clinical and cognitive data as input.

11. ■ Marinescu RV, Oxtoby NP, Young AL, *et al.* TADPOLE challenge: prediction of longitudinal evolution in Alzheimer's disease. ArXiv 2018; ArXiv180503909. Available from: http://arxiv.org/abs/1805.03909.

Well defined challenge which aim is to predict future diagnoses, cognitive test values and MRI measures at each month over five years.

12. Ho KC, Speier W, Zhang H, *et al.* A Machine learning approach for classifying ischemic stroke onset time from imaging. IEEE Trans Med Imaging 2019; 38:1666–1676.

13. Zhang H, Alberts E, Pongratz V, *et al.* Predicting conversion from clinically isolated syndrome to multiple sclerosis–An imaging-based machine learning approach. NeuroImage Clin 2019; 21:101593.

14. Qiu S, Chang GH, Panagia M, *et al.* Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimers Dement Diagn Assess Dis Monit 2018; 10:737–749.

15. Sørensen L, Nielsen M. Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. J Neurosci Methods 2018; 302:66–74.

16. Lu P, Colliot O. Multilevel survival analysis with structured penalties for imaging genetics data. In: Medical imaging 2020: Image Processing International Society for Optics and Photonics; 2020. P. 113130K

17. Peng J, An L, Zhu X, *et al.* Structured sparse kernel learning for imaging genetics based Alzheimer's disease diagnosis. Med Image Comput Comput-Assist Interv MICCAI 2016 – 19th Int Conf Proc 2016; LNCS 9901:70–78.

18. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 2009; 12:535–540.

19. ■ Wen J, Thibeau-Sutre E, Diaz-Melo M, *et al.* Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med Image Anal 2020; 63:101694.

This work highlights the biases present in numerous articles proposing deep-learning methods for the classification of Alzheimer's disease. An open-source framework enabling the reproducible evaluation of such methods is then proposed.

20. ■■ Pulini AA, Kerr WT, Loo SK, Lenartowicz A. Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: effects of sample size and circular analysis. Biol Psychiatry Cogn Neurosci Neuroimaging 2019; 4:108–120.

Review demonstrating that high classification accuracies appear to be inflated by circular analysis (detected in a nonnegligible number of studies) and small sample size.

21. ■ Böhle M, Eitel F, Weygandt M, Ritter K. Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. Front Aging Neurosci 2019; 11:194.

This interpretability approach enables the identification of the brain regions driving the decision when classifying patients with Alzheimer's disease, both at the data set scale and at the individual scale.

22. Ding Y, Sohn JH, Kawczynski MG, *et al.* A deep learning model to predict a diagnosis of Alzheimer disease by using ^{18}F-FDG PET of the brain. Radiology 2019; 290:456–464.

23. Thibeau-Sutre E, Colliot O, Dormont D, Burgos N. Visualization approach to assess the robustness of neural networks for medical image classification In: Medical imaging 2020: Image Processing International Society for Optics and Photonics; 2020. p. 113131J.

24. ■ Chen S, Zhang J, Ruan X, *et al.* Voxel-based morphometry analysis and machine learning based classification in pediatric mesial temporal lobe epilepsy with hippocampal sclerosis. Brain Imaging Behav 2019.

One of the very first studies on the topic.

25. Saccà V, Sarica A, Novellino F, *et al.* Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data. Brain Imaging Behav 2019; 13:1103–1114.

26. Zurita M, Montalba C, Labbé T, *et al.* Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. NeuroImage Clin 2018; 20:724–730.

27. Feis RA, Bouts MJRJ, Panman JL, *et al.* Single-subject classification of presymptomatic frontotemporal dementia mutation carriers using multimodal MRI. NeuroImage Clin 2018; 20:188–196.

28. Meyer S, Mueller K, Stuke K, *et al.* Predicting behavioral variant frontotemporal dementia with pattern classification in multicenter structural MRI data. NeuroImage Clin 2017; 14:656–662.

29. Zeng L-L, Wang H, Hu P, *et al.* Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine 2018; 30:74–85.

30. Chen Y, Tang Y, Wang C, *et al.* ADHD classification by dual subspace learning using resting-state functional connectivity. Artif Intell Med 2020; 103:101786.

31. van der Burgh HK, Schmidt R, Westeneng HJ, *et al.* Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. NeuroImage Clin 2017; 13:361–369.

32. Li K, O'Brien R, Lutz M, Luo S. A prognostic model of Alzheimer's disease relying on multiple longitudinal measures and time-to-event data. Alzheimers Dement J Alzheimers Assoc 2018; 14:644–651.

33. Koval I, Schiratti J-B, Routier A, *et al.* Spatiotemporal propagation of the cortical atrophy: population and individual patterns. Front Neurol 2018; 9:235.

34. Schiratti J-B, Allassonnière S, Colliot O, Durrleman S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. J Mach Learn Res 2017; 18:1–33.

35. Moradi E, Pepe A, Gaser C, *et al.* Alzheimer's disease neuroimaging initiative. machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 2015; 104:398–412.

36. Shen L, Thompson PM. Brain imaging genomics: integrated analysis and machine learning. Proc IEEE Inst Electr Electron Eng 2020; 108: 125–162.

37. Da X, Toledo JB, Zee J, *et al.* Integration and relative value of biomarkers for prediction of MCI to AD progression: Spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. NeuroImage Clin 2014; 4:164–173.

38. Gupta Y, Lama RK, Kwon G-R. Prediction and classification of Alzheimer's disease based on combined features from apolipoprotein-e genotype, cerebrospinal fluid, MR, and FDG-PET imaging biomarkers. Front Comput Neurosci 2019; 13:72.

39. Khanna S, Domingo-Fernández D, Iyappan A, *et al.* Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's disease and reconstruction of relevant biological mechanisms. Sci Rep 2018; 8:1–13.

40. Varatharajah Y, Ramanan VK, Iyer R, *et al.* Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. Sci Rep 2019; 9:2235.

41. ■■ Samper-González J, Burgos N, Bottani S, *et al.* Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. NeuroImage 2018; 183:504–521.

Proposes a framework and sets standards for reproducible research.

42. Gorgolewski KJ, Auer T, Calhoun VD, *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 2016; 3:160044.

43. Rathore S, Habes M, Iftikhar MA, *et al.* A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage 2017; 155:530–548.

44. ■ Cai XL, Xie DJ, Madsen KH, *et al.* Generalizability of machine learning for classification of schizophrenia based on resting-state functional MRI data. Hum Brain Mapp 2020; 41:172–184.

One of the rare studies investigating both within-site and between-site generalizability of a machine learning classification framework.

45. ■■ Bouts MJRJ, van der Grond J, Vernooij MW, *et al.* Detection of mild cognitive impairment in a community-dwelling population using quantitative, multiparametric MRI-based classification. Hum Brain Mapp 2019; 40: 2711–2722.

Study assessing the ability of an MRI-based classifier trained on a clinical cohort to generalize to a general population.

46. Bisenius S, Mueller K, Stuke K, *et al.* Predicting primary progressive aphasias with support vector machine approaches in structural MRI data. NeuroImage Clin 2017; 14:334–343.

47. Li H, Habes M, Wolk DA, Fan Y. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimers Dement 2019; 15:1059–1070.