REVIEwS

Applications of machine learning to diagnosis and treatment of neurodegenerative diseases

*Monika A. Myszczynska 1, Poojitha N. Ojamies2, Alix M. B. Lacoste 3, Daniel Neil2, Amir Saffari2, Richard Mead 1, Guillaume M. Hautbergue 1, Joanna D. Holbrook2 and Laura Ferraiuolo 1*✉

Abstract | Globally, there is a huge unmet need for effective treatments for neurodegenerative diseases. The complexity of the molecular mechanisms underlying neuronal degeneration and the heterogeneity of the patient population present massive challenges to the development of early diagnostic tools and effective treatments for these diseases. Machine learning, a subfield of artificial intelligence, is enabling scientists, clinicians and patients to address some of these

challenges. In this Review, we discuss how machine learning can aid early diagnosis and interpretation of medical images as well as the discovery and development of new therapies. A unifying theme of the different applications of machine learning is the integration of multiple high-dimensional sources of data, which all provide a different view on disease, and the automated derivation of actionable insights.

Omics

A collective term for a field within biological research concerned with the study of ‘omes’; for example, the genome, transcriptome or proteome.

*1Sheffield Institute of Translational Neuroscience, University of Sheffield, Sheffield, UK.*

*2BenevolentAI, London, UK.*

*3BenevolentAI, Brooklyn, NY, USA.*

✉*e-mail: l.ferraiuolo@ sheffield.ac.uk*

https://doi.org/10.1038/ s41582-020-0377-8

First described in 1955 by John McCarthy as “the sci ence and engineering of making intelligent machines”1, artificial intelligence (AI) is a field of computer science research devoted to designing software capable of per forming computations with a sophistication similar to that of human intelligence. AI encompasses a wide range of computational systems and tools that mimic actions that the human brain performs on a daily basis: prob lem solving, reasoning, pattern spotting and knowledge acquisition2. Machine learning and natural language processing are the forms of AI most commonly used in health-care settings as they enable a robust interrogation of datasets in order to identify previously undiscovered patterns and relationships between different features in the data3. These two forms of AI are distinct, but share common features, as natural language processing largely uses machine learning to derive meaning from language. The functions of machine learning could aid diagnosis, development of novel therapies and help improve our understanding of the disease course.

The application of machine learning algorithms to medicine and scientific research has been widely dis cussed in recent years4–9. In the past decade, new tech nologies have enabled rapid accumulation of patient data such as ultrasonography and MRI readouts; omics profiles of biological samples; electronically cap tured clinical, behavioural and activity data; and social media-derived information10–13. These big health datasets are high-dimensional, meaning the number of features

(or variables) recorded per observation can sometimes exceed the total number of observations. For exam ple, gene expression datasets can contain the expres sion levels of ~20,000 genes, whereas obtaining data from 20,000 individuals with a given disease would be extremely challenging. These high-dimensional datasets are often sparse, noisy, cross-sectional and lack statistical power, making it extremely difficult to gain biological insights from these data using traditional data analytical approaches, which look for changes in single variables or perform simple correlations14. These problems with data analysis are further compounded by the integration of diverse data types (for example, imaging, genomics and clinical data) that is necessary to gain an understanding of disease mechanisms. In response to these challenges, advanced machine learning models are increasingly applied to biomedical and health-care data. Traditional computer science derives results from input data through the application of predefined rules, whereas machine learning learns rules and insights from input data directly, thus allowing the application of those rules to make predictions from data in new situations. Machine learning approaches can help overcome the challenge of high-dimensional data by reducing the number of fea tures analysed in favour of the least variable15. Different machine learning algorithms can also be used to integrate data from different sources to increase statistical power.

By the year 2050, an estimated 22% of the global popu lation will be over 60 years of age16. As age is the main

Nature Reviews **| Neurology**

Reviews

**Key points**

• Machine learning and natural language processing are forms of artificial intelligence that enable robustinterrogation of multiple datasetsto identify previously undiscovered patterns and relationshipsin the data.

• Machine learning approaches have been applied to the study of neurodegenerative diseases and show promise in the areas of early diagnosis, prognosis and development of new therapies.

• A substantial number of machine learning algorithms exist, and choosing the correct algorithm to apply to differenttypes of data is crucialto obtain reliable results. • Neuroimaging wasthe first area of neurology to benefitfrom the application of machine learning approachesto improve diagnosis; more recently, application of machine learning methodsto motorfunction and language feature analysis has shown promise in decreasing the time taken to perform clinical assessments. • The application of machine learning to longitudinal patient data collection and electronic health records hasthe potentialto inform prognosis prediction and patient stratification.

• Large collections of curated datasets and robust assessment of machine learning methods will be needed to achieve full integration of machine learning into diagnostic and prognostic neurology practice and the design of future therapeutics.

risk factor for most neurodegenerative disorders, includ

ing Alzheimer disease (AD), Parkinson disease (PD)

and motor neuron disease (MND), countries across the

world are facing an unprecedented economic challenge17.

In 2016 the global cost of caring for individuals with AD

reached an estimated US$946 billion, which was triple

the estimated expenditure in the year 2000 (ref.16). These

costs are expected to rise, as the number of individuals

with AD is likely to reach 115 million worldwide by 2050

(ref.16). These estimates call for changes in the way that

individuals are diagnosed and treated, and highlight the

urgent need for effective therapeutic interventions. In this

context, machine learning could enable data to be used

more efficiently to provide insights into disease mech

anisms, and to help with earlier diagnosis, prognosis,

patient stratification and development of new therapies.

With these goals in mind, many researchers have gath

ered rich, high-dimensional datasets from healthy indi

viduals and individuals with neurodegenerative diseases;

(for example, the size of a brain region on an MRI scan) and the label (for example, a diagnostic category). The algorithm can then apply this model to new, unlabelled datasets to predict the label on the basis of the new input features. Gathering sufficiently large volumes of accurate labels for supervised machine learning can be a challenge19.

Supervised machine learning is divided into clas sification and regression algorithms18. Classification algorithms, such as the example above, predict the cate gorical output (diagnostic category) for each data sam ple (patient). By contrast, regression algorithms predict a real-valued variable (for example, degree of functional impairment measured on a continuous scale) for each data sample. When applied to health-care data, both classification and regression algorithms can define patient endotypes by identifying patterns within the data and clustering areas of similarity together. A practical example of regression approaches would be the subtyp ing of patients into progression endotypes on the basis of algorithms that model motor function decline, disease duration, or slope of progression to form nuanced rep resentations of the progression time series. This regres sion approach contrasts with endotyping on the basis of categories, which might include specific genetic muta tions or the site of disease onset. Most machine learning algorithms have variants to support both classification and regression.

In contrast to supervised machine learning, unsu pervised machine learning algorithms do not require labelled data and are useful for tasks such as clustering data samples into groups, or reducing the dimension ality of datasets by generating a simpler representation of highly complex data20,21. For example, unsupervised clustering algorithms can be used to analyse gene expres sion datasets and identify clusters of patients with shared molecular signatures22. Furthermore, unsupervised clus tering approaches, such as latent variable models, can help identify co-expression modules of genes, which are sets of genes that are likely to be co-regulated or corre

Endotypes

Clusters of individuals within a disease population that share functional and pathological traits.

Molecular signatures

A collection of proteins, genes and their variants that can be used as hallmarks for a given phenotype.

Regularizing

The technique of adding constraints or knowledge within the training process in order to prevent overfitting.

Data leakage

An undesirable process

whereby information is

accidentally shared between the training data and the test data, resulting in test

evaluation scores that are not representative of real-world unseen data.

for example, the Alzheimer’s Disease Neuroimaging Initiative (ADNI), the Allen Brain Atlas and the UK Biobank. In this Review, we highlight the latest devel opments in the use of machine learning to interrogate

neurodegenerative disease-related datasets, includ ing the applications of machine learning to diagnosis, prognosis and development of new therapies.

Machine learning models

Machine learning methods are broadly categorized into supervised, unsupervised and reinforcement learn ing approaches18 (Fig. 1). Supervised machine learning algorithms are currently the methods most commonly applied to neurodegenerative disease-related data and require a labelled dataset from which to learn. Often, these labels require manual curation or expert assess ment; for example, a radiologist is needed to label a set of MRI scan images and a neuropathologist is required to categorize a set of images obtained from post-mortem patient samples. Once this ‘benchmark’ dataset has been labelled, the machine learning algorithm builds a model of the relationship between the input features

spond to common biological mechanisms or pathways. In addition to analysing existing data, unsupervised clus tering algorithms can also be used to make predictions; for example, a model can be trained on a set of historical clinical data to then predict survival from the cluster that a patient is placed in.

Supervised and unsupervised learning approaches can be combined, for example, to form semi-supervised learning methods23. Semi-supervised methods enrich a small set of labelled data with additional unlabelled data, which allows clustering (unsupervised) methods to improve the performance of classification (supervised) methods, as well as regularizing the predictive model with additional data. Similarly, transductive learning methods use the test data as unlabelled data to improve a standard supervised classification approach24,25; these methods do not result in data leakage as the labels are not shared, and can increase performance in problems where low volumes of data are available.

Finally, in reinforcement learning approaches26 a reward or punishment is given to achieve a desired out put. For example, an algorithm might be used to explore

**www.nature.com/nrneurol**

Sample-to-feature ratio (SFR). The number of data points divided by the number of features; for example, gene expression data comprising tens of patients with thousands of gene expression levels would have an SFR of <1.

a new medication regimen for a patient based on the patient’s medical history. During training, in the event of a negative reaction to a new drug or adverse drug– drug interactions, a punishment would be given to the algorithm, whereas a reward would be given for a drug improving the disease course, which is the desired out put. These approaches are being rapidly explored27,28, but are not yet as widely used as supervised and unsupervised learning in the field of neurodegenerative diseases.

Model selection

A substantial number of machine learning algorithms exist and choosing the correct algorithm to apply to a particular type of data is important. With a particular focus on supervised learning, two factors are particularly relevant for selecting the right algorithm: modality (the form that the data is in) and volume (the number of data samples). In terms of volume, for datasets with a low sample-to-feature ratio29 (SFR <10:1), an algorithm will struggle to learn a useful ‘featurization’ in addition to

Reviews

classification (Fig. 2). Data featurization requires the algo rithm to identify and extract characteristics or ‘features’ in the data that enable the subsequent separation of data points into classes. The higher the SFR, the easier it will be for the algorithm to identify features that separate the data points. For example, given the limited availability of post-mortem tissue samples for certain neurodegen erative conditions, a model that needs a high SFR to make meaningful predictions about the data, such as a deep neural network, is unlikely to be able to learn from tens of samples and identify features that can accurately classify patients into different pathological subtypes. For such limited datasets, highly constrained or ‘regularized’ models, such as hierarchical Bayesian models, simplify the task and guide the algorithm by learning only a few parameters for that data.

For larger datasets, support vector machines (SVM)30 or random forests31,32 (Table 1) are typically used. These approaches are more flexible than hierarchical Bayesian models, but require a greater volume of data and are

**Supervised learning Unsupervised learning Reinforcement learning**

• A labelled dataset is provided

• Learning is task-driven

• Algorithm trains to improve outcome over time

Training dataset 

in which known

pathological

features are

labelled

• No labelled dataset is provided and output is unknown

• Learning based on pattern identification and recognition

Heterogeneous

patient population

• Based on trial and error with reward or punishment before repetition

• Algorithm trains to improve outcome over time

Patients and

their medical

history

Train model to recognize

features

Evaluate model on test data

Unsupervised

Multimodal patient data

Choose and apply action

Side effects, adverse

reaction, lack of treatment response,

drug–drug interaction,

Apply model to

unseen data Unseen data Output 

Diagnostic

class 1

clustering on basis of

molecular features

Diagnostic

class 2

Patient subtypes with shared

molecular features

disease

Done? progression No

Yes

Measure outcome,

apply reward or

punishment

Fig. 1 | **Categories of machine learning.** Machine learning methods can be divided into three main categories: supervised learning, unsupervised learning and reinforcement learning. Supervised learning relies on a set of data labelled by a professional to train an algorithm to extract specific disease features. Once trained, the model identifies features of interest in unlabelled datasets to aid diagnosis. Unsupervised learning learns to determine patterns and classes within the dataset without labels and can be particularly useful in identifying molecular signatures that categorize heterogeneous patient disease groups into molecular subtypes. In reinforcement learning models, the process of decision making and the output is a result of knowledge gained from previous experiences. The algorithm is trained on a trial and error basis, with a reward or punishment driving the learning process and skills acquisition over a series of trials. Such an approach is ideal for

automation and robotics, although use in medicine is increasing. Brain scan image credit: M. Andritoiu/Alamy Stock Photos. Nature Reviews **| Neurology**

Reviews

Overfitting

When an algorithm learns the patterns within the training dataset as opposed to the patterns representative of all data.

Classifier

A type of model used to identify the correct category for a data point.

more complex. SVMs map datasets into a space such that two categories (for example, healthy and diseased) are separated as widely as possible. In the 1990s, pioneer work demonstrated that the mathematical theory behind SVM, initially used for two-group classification prob

lems, could be applied to more complex and multiclass datasets33. Random forests use a different approach to SVM. A random forest algorithm constructs numerous different, independent decision trees, which each require a series of binary choices to be made about the data. In this way, each decision tree provides a classification for the input data, and the algorithm then selects the most common output prediction from the different trees. This approach corrects for the overfitting that can occur when using a single decision tree. SVM and random forests were dominant until the more recent development of neural networks.

Artificial neural networks, which include the popular deep neural networks, are widely used to analyse many modalities of data, but particularly image, video and sound data19,34 (Table 1). These networks are designed to simulate the highly connected neural system found in the brain, and are based on the pioneering work of McCulloch and Pitts35 who developed a mathematical model to help bridge the gap between statistics and neuropsychiatry. Similar to a neuron receiving electrical

stimuli from other neurons, a node in an artificial neural network receives inputs from other nodes. Just as a neu ron needs to reach a certain threshold to fire an action potential, the nodes of an artificial neural network sum up the inputs from other nodes and calculate the like lihood that a potential output is true. Artificial neural networks are typically arranged into one or more lay ers of parallel ‘neurons’ that transform their inputs into outputs; networks that contain two or more layers are known as deep neural networks. In each successive layer, the neural network is able to manipulate more abstract representations of the data. Artificial neural networks require even fewer manual data manipulation steps during preprocessing than SVM or random forests and in some cases subsume the choice of classifier into the architecture of the network19. These networks are mostly supervised, but can also be unsupervised.

The use of a particular type of deep neural network, known as a convolutional neural network (CNN)36–38 (Table 1), has led to significant improvements in machine learning performance for medical image interpretation. CNN, which draws inspiration from the human visual system, extracts features at higher and higher levels of abstraction, initially combining local information and eventually integrating large-scale information across the image. This approach enables complex processing, such

a

Classifier

Classifier

Learned

representation

Classifier

Learned

representation

b

**Classifier type:**

Hierarchical Bayesian

models

Random

forest; SVMs

Neural network with sofmax regression

Manual features

Manual features

Manual features

**101 102 103 104 105** Data volume (number of data points)

Data

c

Data Data

Increasing data

Volume

**Data category Suggested machine learning models Example usage in neurodegenerative disorders** Natural language Diagnosis of Alzheimer disease from language data132; Specialized natural language processing tools,

for example, BERT153, SpaCy155, CoreNLP154 and SyntaxNet156

prediction of linguistic decline caused by dementia158

Imaging and volumetric data

Time-series data Graph data

2D and 3D convolutional neural networks19, 70 Alzheimer disease diagnosis72

3D and 4D convolutional neural networks122 Parkinson disease diagnosis from wearable device that measures gait119; activity recognition126

Graph convolutional neural networks178 Prediction of drug adverse effects176; molecular fingerprint prediction177

Dementia classification76; ALS progression prediction General data and 32 Hierarchical Bayesian models83; SVMs76;

multimodal data

random forests32; variational autoencoders73

Fig. 2 | **Determining the best machine learning model for a given problem.** Two factors are particularly relevant to the choice of machine learning model: the data modality (for example, time series or imaging) and the data volume (number of data points). **a**,**b** | At small data volumes, manual features and well-regularized classifiers are needed. As data volume increases, a learned representation that discovers features from data automatically can be used, and eventually a separate, custom classifier on the learned representation is no longer needed. **c** | Choosing the right model to match the data type enables machine learning to extract a meaningful representation from the data, which can be paired with a variety of classifiers. ALS, amyotrophic lateral sclerosis; BERT, Bidirectional Encoder Representations from Transformers; SVMs, support vector machines.

**www.nature.com/nrneurol**

Reviews

Table 1 | **Descriptions of key machine learning methods and examples of their application**

**Method Supervised or unsupervised?**

**Classification, regression or clustering?**

**Definition examples of application**

Deep neural networks

Convolutional neural network

Recurrent

neural

networks

Supervised and unsupervised

Supervised and unsupervised

Supervised and unsupervised

Classification, regression and clustering

Classification, regression and clustering

Classification, regression and clustering

A category of algorithms that are composed of multiple layers of transformation and use artificial neural networks as each layer of processing

Form of a deep neural network designed to mimic neuronal connections and reduce dimensionality of the data

A type of artificial neural network capable of forming internal memory, which makes it optimal for analyses of sequential data

Disease diagnosis from neuroimages, speech recognition, natural

language processing and pattern recognition19,71

Image analysis and

classification, video analysis, natural language processing and acoustic input65,92,93,158

Speech classification, handwriting classification and disease progression classification92,140

Network diffusion algorithm

Unsupervised NA A class of methods that model the spread of information

through networks

Gene regulator analysis and network inference for gene prioritization149

Random forest Primarily supervised

Primarily

classification and regression

An algorithm that averages a large number of weak classifiers (trees) to form a strong classifier using random subsets of data

Disease progress prediction32

Support vector machines

NA, not applicable.

Primarily supervised

Classification, regression and clustering

A method of separating distinct classes within a dataset by applying an optimal hyperplane

Disease

classification54,58–60,82,210

Hyperparameter

A parameter the value of which is set before training; for example, the attributes of the model architecture.

Cross-validation

A training and evaluation procedure that consists of splitting the data into subsets and alternately holding out one subset for evaluation until all subsets have been evaluated.

Metadata

Data about other data; for example, information about an experimental protocol or the time and date of sample collection.

as distinguishing cats from dogs or the identification of cancerous cells39. Many of the problems involved in image classification can be solved by these kinds of algo rithms. Another type of deep neural network, known as a recurrent neural network (RNN)40 (Table 1), can extract information from sequences of data and is particularly useful for analysing clinical records. RNN models such as long short-term memory (LSTM)41 and gated recur

rent units42 form building blocks used in most sequence tasks. These models contain a memory cell that allows the algorithms to learn long-term dependencies and gates that control the exposure of the memory content and the extent of the changes made to the memory content depending on the input.

Some of the key technical risks to mitigate when choosing a machine learning model include insufficient data volume, improper data representations, overfitting, incorrect hyperparameter selection and missing data43,44. Although subject matter expertise can help address issues regarding data volume and data representation, overfitting is a fundamental issue. For example, when a model is trained on a dataset with low bias and high var iance, and the model selected has the best performance on the training data, the model is likely to try to ‘mem

orize’ the training data. This memorization results in overfitting of the model on the training data, leading to poor predictive performance on test data as the model is no longer able to make generalizations to the additional datasets. Overfitting could, for example, lead to some pathological features in neuroimages not being identi

fied. Methods such as cross-validation and regularization

can help minimize this problem45. All machine learning methods and data sources have caveats; therefore, a com bination of multiple data sources and methods followed by confirmatory post-processing steps that use a variety of metadata is the best approach.

Diagnosis and prognosis

In many neurodegenerative diseases, including AD, PD and MND, symptoms do not manifest until a substantial loss of neurons has already occurred46–48, which makes early diagnosis very challenging. Therefore, research into the application of machine learning models to early diagnosis is growing (Table 2). The aim of this research is to use machine learning to detect prognostic signals in data that can be collected relatively easily (for example, electronic health records (EHRs) or MRI data), thus enabling the prospective screening of ageing popu lations. The machine learning-driven automated diag

nosis could then flag individuals for further clinical investigation. Such an approach would require machine learning models that are sensitive enough to detect early disease signals and specific enough not to over

burden health systems with unnecessary follow-up tests. Currently, test results need to be analysed and interpreted by trained staff, which can lead to delays in diagnosis. These delays could be reduced by applying machine learning approaches to the data as they are gathered in the clinic. These same data could be used to predict patient prognosis by comparing disease progression at any given time with historical data from patients shar ing the same endotype or phenotype. Historical health

Nature Reviews **| Neurology**

Reviews

Table 2 | **Key neurodegenerative disease-related data types used by machine learning algorithms**

**Neurodegenerative disease**

**Key data types used**

**For diagnosis For monitoring disease progression**

Dementias, including AD and MCI

Neuroimaging30,55,72,75–77,79,82; cognitive performance tests131,132,135,136; EEG data104; transcriptomic data97,140; biomarker data143,144,146

Neuroimaging77,83,211; cognitive performance tests131,132,158; IADL records125,126

PD Motor performance tests114,115,119; neuroimaging86 Electronic health records165 MS Metabolomic data145; neuroimaging85 Genomic data142

AD, Alzheimer disease; EEG, electroencephalogram; IADL, instrumental activities of daily living; MCI, mild cognitive impairment; MS, multiple sclerosis; PD, Parkinson disease.

records provide a helpful training dataset for prognosis algorithms, as they can cover the entire disease span.

***Neuroimaging.*** Neuroimaging techniques such as CT and MRI are often used in the diagnosis of neurode generative diseases, and radiology was one of the first fields to benefit from the computerization of medicine and the introduction of ‘intelligent machines’49 (Fig. 3). The early 1990s saw the introduction of supervised knowledge-based expert systems50–55, which were capa ble of recognizing pathological events in the brain on the basis of a large amount of data and knowledge col lected by the neuroradiological community56. Initial studies used clinically relevant diagnostic features, such as cortical thickness or morphology of particular brain regions, to classify patients and help radiologists make a diagnosis57–63. This approach is known as computer aided diagnosis, and continued to be developed and improved throughout the early 2000s64–68. As machine learning is purely evidence-based and can analyse prob lems in an unbiased manner, the approach is helpful for making objective diagnoses from medical images and often surpasses the performance of trained professionals

in terms of speed, precision and accuracy30,69. Computer-aided diagnosis systems can be sup plemented with and powered by supervised learning techniques to further improve the interpretation of neuroimaging data and help identify subtle abnormal ities in the images that are not detected by radiologists. For example, CNNs can categorize images by identify ing and mapping a high number of features70 (Table 1). Some studies have used CNNs to predict a diagnosis of AD71,72 and to study cognitive ageing73 from MRI and PET images, sometimes alongside other clinical readouts (for example, biomarker information and assessments of motor or cognitive performance), to increase specificity. SVMs have been used to analyse MRI data, some times combining structural and functional MRI, and cognitive assessment data to improve disease diagnosis74 (Table 1). For example, one study used an SVM to dif ferentiate between structural MR scans from individuals with different severities of AD and cognitively normal elderly individuals, as well as to differentiate between individuals with AD and individuals with frontotempo ral lobar dementia (FTLD)75. In another study of struc tural MRI data, an SVM was able to predict conversion from mild cognitive impairment (MCI) to AD, as well as separate healthy controls, individuals with MCI and individuals with AD better than a combination of sta tistical approaches and expert knowledge76. In the same

year, a study used a combination of linear dynamic system and SVM algorithms to integrate MRI data and cognitive test data to distinguish individuals with AD from healthy controls77. This study is particularly inter

esting because of the use of two approaches to integrate data. SVMs have also been applied to whole-brain ana tomical MRI images to identify new regions of interest that can differentiate between individuals with AD and healthy controls78, and to images of the hippocampus to classify its shape features79. The latter method has been described as more accurate than traditional hippocam pal volumetry80,81. An SVM was also used to compare the utility of different combinations of neuroimaging data (functional MRI or structural MRI) and cognitive performance data for identifying individuals with MCI82. Analysis of MRI data with SVM has also been used in studies of cognitive function83, stroke84, multiple sclerosis (MS)85 and parkinsonism86.

MRI produces images of higher resolution than CT; however, the diagnostic performance of CT can be improved with the use of machine learning algorithms. For example, in one study, a random forest algorithm for automated white matter lesion detection was applied to a set of CT images from individuals with acute ischaemic stroke and performed similarly to the labelling of MR images by radiologists87. The algorithm had a failure rate of 4% and an average processing time of less than 2min utes, thus offering a possibility of similar approaches being extended to diagnosis of neurodegenerative dis

eases. Machine learning can assess images quickly, so it could be used to flag findings from CT images that need urgent review by a radiologist in life-threatening scenarios88–90. Quick image analysis could also be extended to MRI, which could prove particularly use

ful for individuals with MCI91–93, as early prediction of possible conversion of MCI to AD could enable a swifter start of treatment.

In a global effort to improve our understanding of neurodegenerative diseases and their progression, data bases of neuroimaging data from patients are being assembled with the aim of creating a comprehensive picture of the disease course from diagnosis onwards. Resources such as the Parkinson’s Progression Markers Initiative, ADNI, The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER) 94 or European Alzheimer’s Disease Consortium Impact of Cholinergic Treatment Use (EADC-ICTUS) give researchers working with machine learning algorithms access to verified material to be used for algorithm training and validation. Although

**www.nature.com/nrneurol**

Computational pattern

recognition applied to PET images identified subgroups of individuals with AD on the basis of cerebral glucose metabolism54

An automated cross-modality image registration method for

1990 1991 1993

The first computerized 3D brain atlas for the analysis of MRI and PET data was developed51

First use of an

artificial neural

Reviews

Efforts to investigate the interaction between genetics and brain structure have led to the rise of a new field called neuroimaging genomics96. The very first attempts to draw a correlation between genetics and brain func

tion were initiated in the early 2000s and focused on single gene variants97,98. These pioneer studies made several novel observations, including the association

MRI and PET was developed52

A new pattern matching approach used a digital brain atlas and MRI data to measure hippocampal57

1995 1996

network

algorithm to automate the diagnosis of dementia from SPECT data55

First

hierarchical brain mapping algorithm was generated53

of the ε4 allele of the apolipoprotein E gene (*APOE* ε4) with abnormalities in brain activation assessed with functional MRI during memory-activation tasks99. Since these initial studies, the field has generated break

throughs owing to large multicentre collaborations such

and corpus callosum volumes58

Developments in computational techniques helped correlate behavioural data with changes in volume of specific brain structures in individuals with AD60

Automated methods helped measure cerebral cortex

thickness61 and grey matter density62 from MRI images

Early application of SVM to MRI data helps differentiate between males and females65, as well as between healthy controls and individuals with schizophrenia210

SVM applied to MRI data

distinguished between healthy ageing and different types of dementia75 and separated healthy controls from individuals with MCI91

SVM was used to distinguish individuals with stroke from healthy controls on the basis of resting state fMRI and diffusion tensor imaging data84

1999 2000 2001 2003 2004 2005 2008 2009 2013 2014 2015

A machine learning algorithm was used to build an automated system that used MRI data and data on cognitive function to understand neuroanatomical sex differences68

Supervised and unsupervised neural network algorithms for the reconstitution of sparsely sampled fMRI data were introduced68

An automated analysis system based on a digital brain atlas was applied to fMRI data64

SVM applied to MRI data was used to distinguish between individuals with Parkinson

disease and individuals with Parkinson plus syndromes86

A random forest algorithm was applied to a multimodal

dataset to distinguish between individuals with AD and MCI71

A combination of linear dynamic system and SVM algorithms was applied to MRI data and cognitive test data to find features that can distinguish individuals with AD from healthy controls77

as the Enhancing Neuro Imaging Genetics Through Meta-Analysis (ENIGMA) Consortium, which has analysed genome-wide association study (GWAS) and MRI data from 30,717 individuals across 50 cohorts100. The ENIGMA Consortium has now identified com mon genetic variants that influence the volume of sev eral brain structures, thus shedding light on potential complex genetic drivers of development and disease. As ENIGMA expands, specialized working groups ded icated to the study of several disorders, including PD, ataxia and FTLD, have been formed to help elucidate the pathogenesis of these diseases using imaging genomics.

Neuroimaging is one way of investigating brain activity. Other methods of monitoring brain activity, such as electroencephalography (EEG), can also benefit from machine learning-driven data analyses. EEG data have been used to distinguish individuals with AD from healthy controls101 or patients affected by other forms of dementia102, as well as to detect functional changes in dopaminergic neurons to diagnose PD103. In order to do this, changes in electrical activity frequency over time in different areas of the brain are analysed. Algorithms based on artificial neural networks have been generated to differentiate between individuals with AD and indi viduals with MCI using only unprocessed EEG data, with the aim of speeding up diagnosis and simplifying the monitoring of dementia progression104.

***Motor function.*** Many neurodegenerative disorders, such as MND, Huntington disease (HD) and PD, are

Fig. 3 | **Key developments and novel applications of computational and machine learning techniques to neuroimaging since the 1990s.** Interest in machine learning and its use for medical purposes towards the end of the twentieth century led to the automation of previously time-consuming manual analysis and measurement of neuroimaging data. The low variance and high reproducibility of these AI-driven methods made them an attractive tool for use in clinical settings. Moreover, 3D, high-resolution overview of deep brain structures, previously considered to be beyond analysis without surgical interventions, became within reach. AD, Alzheimer disease; fMRI, functional MRI; MCI, mild cognitive impairment; SPECT, single-photon emission computerized tomography; SVM, support vector machine.

the training data are becoming widely available, apply

ing algorithms to it can be challenging. Some pieces

characterized by motor dysfunction, often culminating in loss of movement105,106. An understanding of the pro gression and chronology of motor degeneration is essen tial for supporting patients appropriately at each stage of the disease. Motor performance is measured during routine clinical assessments, which often assess both fine and gross muscle functions (for example, muscle tone when picking up small objects and gait when walking), although the precise battery of tests varies depending on the disorder107–109.

Machine learning can be used to assess the perfor mance of individuals in complex tasks, such as drawing, in a time-efficient manner110. Simple drawing tasks and

Genome-wide association study

(GWAS). An observational method of studying genetic variants across a population in search for associations

between genetic changes and traits such as diseases.

of software are not, or were not for a long time, capable of reading and processing the DICOM file format that is used for medical imaging. However, open source soft ware and libraries with a strong user community, such as Google TensorFlow have now addressed this issue by extending support to additional file formats, including DICOM95.

handwriting analysis are being increasingly used in the early diagnosis of PD111,112, as handwriting is no longer an automated function in individuals with the disease and is performed in a more segmented and sequential manner, with numerous pauses113. Introduction of machine learn

ing techniques into writing task analysis can help clas sify individuals with PD and serve as a diagnostic tool.

Nature Reviews **| Neurology**

Reviews

In one study, classifier-based supervised machine learn

ing algorithms were applied to data from a horizontal

line drawing task in order to differentiate between

normal and abnormal hand movements and identify

irregularities characteristic of PD114. A set of metrics

related to the velocity and spatiotemporal trace of the

pen combined with naive Bayes classification helped to

accurately distinguish between individuals with PD and

healthy controls. Similarly, in another study, a supervised

model was used to automatically analyse Archimedean

spiral tracing performed by individuals with PD115.

Drawings were independently assessed and scored by

trained experts, and the algorithm-scored drawings

showed comparable results. The drawing tests were per

formed on patients’ personal computers and the software

support for such tests can be extended to commercially

available smartphone or tablet set-ups, making it an

easily obtainable and cost-effective approach to motor

function analysis.

Machine learning-based hardware such as the

Parkinson’s KinetiGraph116 and the Kinesia system117

are designed to score motor function, dyskinesia and

bradykinesia in individuals with PD, and are widely

available. The Parkinson’s KinetiGraph can be worn on

a wrist and measures wrist acceleration. The Kinesia

system is worn on either a finger or a wrist and detects

motion with an in-built accelerometer and gyroscope.

Both systems provide automatic scoring of an individ

ual’s motor symptoms, but the output data can also be

analysed further using machine learning algorithms,

such as SVM. For example, a recent study classified PD

tremor severity using the Kinesia system’s recordings118.

Gait analysis algorithms also show a lot of promise in

early diagnosis of motor disorders. For example, an

RNN and LSTM were used to classify a database of gait

analysis recordings, such as measures of stride-to-stride

footfall times, and distinguished healthy controls from

individuals with PD, HD or MND with an accuracy of

95–100%119. Moreover, the two algorithms in combina

tion performed better than an SVM alone120, suggesting

that existing algorithms could be further improved to

optimize performance.

Data on movement can also be useful in the study

of AD. Film footage of patients performing instrumen

tal activities of daily living (IADL) (for example, bath

ing, dressing and eating) can be watched and manually

scored by clinicians121. However, this approach can be

time-consuming and automating this process would

be of major benefit to neurologists investigating IADL.

Deep learning and CNN-based machine learning algo

rithms are capable of recognizing action from video

footage122 and this technology has been applied to action

recognition in IADL recordings123. However, the use of

cameras to monitor IADL has implications for patient

privacy124. New technologies such as the random forest

algorithm-powered SmartFABER solve privacy issues by

collecting data from motion and contact sensors placed

around the home, feeding this data back to software

installed on the individual’s local personal computer,

and analysing user movement and interactions with

objects125. Data from wearable sensors has also been

used for machine learning-based activity recognition.

For example, one study found that LSTM algorithms performed better than standard CNN approaches in activity identification and classification from this kind of data126. In contrast to the Kinesia system or Parkinson’s KinetiGraph, the sensors used to generate the data used in this analysis were worn on multiple parts of the body, which enabled the study of complex movements, such as using a toggle switch, cleaning a table, and opening and closing doors and drawers. Although individuals with PD currently stand to benefit the most from wearable technology, smartwatch-based wearable sensors that use SVM for recognition of agitated behaviour in individuals with dementia show promise127.

***Language features.*** Language features are important indicators of cognitive state, as communication skills and interpersonal behaviour deteriorate in many neurode generative diseases128–130. Machine learning approaches have been used to extract language features from audio recording transcripts in order to distinguish between individuals with AD and healthy individuals. For exam ple, the authors of one study analysed the way patients compose verbal utterances and used the resulting lex ical features to differentiate between individuals with AD, MCI or vascular dementia and healthy controls131. Several different algorithms were tested, out of which the SVM performed most consistently. In another study, an n-gram language model was used to analyse the vocabulary of participants and the frequencies with which they used specific words together in a sequence132. The algorithm assigned a perplexity score to the utter ances, meaning that the higher the perplexity score, the more unforeseen and convoluted the utterance, and the more likely the individual was to have an AD diag nosis. In individuals with AD, the perplexity scores were found to correlate with mini-mental state examination scores, thus showing the potential of this approach as a diagnostic tool.

In addition to the machine learning-based analysis of transcripts, AI-driven interactive avatars have been used to capture more complex language data. Inspired by early interactive computers133, avatars are animated humans that ask the patient pre-programmed ques

tions and record the resulting conversation134. Avatars most often take the form of software installed on the patient’s personal computer or tablet and no time limit is applied to patients’ responses, which is a scenario that is difficult re-create in face-to-face clinical visits. In one example of the use of an avatar135, a set of ques

tions based on clinical examination tools (mini-mental state examination, objective structural mental exami nation and the Wechsler memory scale-revise test) was used to assess participants for symptoms of dementia. In this study, SVM and logistic regression were used to assign a diagnostic category to patients on the basis of speech features and audiovisual cues (for example, smil ing, eye contact and consistent delays in answering the questions) extracted from the video recordings. This approach was able to distinguish between healthy indi viduals and individuals with dementia with an accuracy of 93%135. This study highlights a key benefit of using avatars over the traditional analysis of transcripts, as

**www.nature.com/nrneurol**

Next-generation sequencing High-throughput, deep

sequencing of DNA and RNA; this technique utilizes

sequencing technologies that are capable of processing multiple DNA or RNA

sequences in parallel.

Metabolomics

A study of metabolites, that is, the small molecule substrates, intermediates and products of cellular metabolism, and their interactions within living organisms.

allowing patients to speak freely and naturally enables the recognition of new speech features, such as pitch and tone changes or breathiness. Transcripts of conversation would not take such features into account and are lim

ited by patient–doctor contact time constraints. A simi lar AI system was used as a conversation analysis aid in a memory clinic and identified four out of six partici pants with neurodegenerative dementia and six out of seven participants with functional memory disorders136, which are important in the differential diagnosis of pro dromal dementia137. As the number of individuals with suspected dementia is rising17, installation of avatar soft ware on an individual’s personal computer could enable the evaluation of speech features at home, which would help speed up diagnosis, save medical personnel time, and reduce the fatigue and potential distress caused by the need to travel to a memory clinic136. Speech data are complex and contain a large number of different fea tures. Therefore, deep neural networks19 (Table 1) are frequently used to perform pattern recognition in this kind of multilayered data.

***Molecular and genetic data.*** Improving our understand ing of the molecular foundations of neurodegenerative disorders is key for the development of new therapies and for diagnosis and prognosis. Next-generation sequencing techniques have increased the speed of DNA sequencing, enabling large volumes of data to be acquired relatively quickly. The volume of genomic data produced, espe cially in GWAS and other large cohort studies, requires a well-refined analysis approach and machine learning techniques are proving useful in this area. Multiple AD-associated genes have been identified138, but the idiopathic nature of the disease, along with its high heritability, suggests that further genetic risk factors or complex genetic interactions might play important roles in disease onset or progression139. GWAS aim to unravel some of these complex relationships. For example, in one study a supervised SVM-based algorithm was used to interrogate brain-specific gene expression data with the aim of identifying novel AD-associated genes140. The authors used a training dataset of 335 AD-associated genes identified through previous GWAS and other genetic studies and 335 non-AD-associated genes, and integrated brain-specific gene expression data to train the classifier to identify AD-associated pathways. The authors then used the trained algorithm to identify genes that interacted closely with the known AD-associated genes in the brain-specific network and ranked these new candidate genes by AD association probability. The top candidates corresponded to genes previously identified as AD-associated by GWAS, but the authors also identified a number of genes involved in cellular processes, such as enzyme binding, that had not been identified before.

A different approach was taken in another study, in which gene expression profiles were predicted from ADNI GWAS data and a range of different machine learn ing algorithms used to identify associations between AD diagnosis and gene expression profiles across dif ferent tissues141. In this study, the RNN was the most accurate algorithm for distinguishing individuals with

Reviews

AD from healthy individuals. The authors also tested the performance of the algorithm in predicting disease phenotype, but the results were inconclusive, probably owing to the complexity of the interaction between genetics and environment. Using machine learning to cluster patients based on their genomic similarity is also helping to develop stratification tools for MS142, the hope is that this stratification will help triage patients after diagnosis and predict their individual disease trajectory.

Applying machine learning to study protein sig natures in samples from patients can aid biomarker discovery, which in turn is likely to improve disease diagnosis. In a study by Ray et al.143, published in 2007, a classification algorithm called predictive analysis of microarrays was used to identify plasma proteins that could discriminate between individuals with AD and healthy individuals when given a cohort of blinded sam

ples. Starting from a pool of 120 proteins assessed using an enzyme-linked immunosorbent assay, the authors identified 18 signalling proteins, the blood expression level of which distinguished between samples from indi

viduals with AD and healthy controls with close to 90% accuracy. These 18 proteins were also used to identify patients who had MCI that progressed to AD within 2–6 years of sample collection. Several years later, in a study by Agarwal et al.144, an unsupervised artificial neural network algorithm for both feature selection and classification was applied to the same dataset as used by Ray et al.143. The artificial neural network identified a smaller set of 9 proteins, as opposed to 18, that distin guished individuals with AD from healthy controls with accuracy similar to that found by Ray et al., resulting in significant economic savings. Of these nine proteins, seven were common between the two studies, whereas two were new findings. In addition, Agarwal et al.144 identi fied a cluster of 29 proteins that identified individuals with MCI that would progress to AD, individuals with MCI that would progress towards other dementias, and individuals with AD. This prediction accuracy could not be achieved with either the 9-protein or the 18-protein clusters identified in the two studies. Comparative studies of this kind highlight how advances in machine learning approaches can refine and improve disease classification and prediction accuracy to benefit patient health, as well as reduce economic costs.

Similarly, in recent studies machine learning has been applied to metabolomics data from individuals with MS145 or AD146 to identify new biomarkers for these diseases. In the first study145, 400 plasma metabolites were asses sed in a small cohort of 12 individuals with MS and 13 healthy controls. Supervised random forests, with 5,000 trees and 100 randomly selected metabolites to determine classification at each node in a tree, identified six metabolites, an increase in the expression of which predicted a diagnosis of MS with a probability of 80%. In another study146, a similar random forests approach was used to analyse the prediction power of a combina tion of clinical and biochemical data, including data on metabolomic, genetic, functional health, lifestyle, cog nitive and bio-demographic risk markers. The analysis showed that different combinations of the six risk mark ers resulted in statistically significant discrimination

Nature Reviews **| Neurology**

Reviews

between individuals with AD, individuals with MCI and

healthy controls.

***Clinical records.*** In addition to the applications dis

cussed above, machine learning can be used to mine rou

tinely collected health-care data for new insights. EHRs

are compiled by health-care providers and contain the

medical history of individuals under their care, which

can include information on immunizations, prescribed

medications, test results and vital signs. EHRs are being

increasingly implemented worldwide147,148 and the col

lection of data in this way requires no additional input

from patients149. However, EHR data were intended to

be read by humans and often consist of unstructured

notes written by health professionals. Therefore, EHRs

need to be converted into computer-readable formats

before being analysed with machine learning techniques.

Machine learning-enabled natural language processing

techniques focus on methods to process and interpret

human language150–156 (Table 1), and can be used to access

the information contained within EHRs157,158. The cur

rent state-of-the-art approach153 for natural language

processing uses deep neural networks pre-trained on

billions of words to perform tasks such as predicting

missing words in sentences or identifying whether two

sentences follow each other. After training, these deep

neural networks can be fine-tuned for a variety of tasks,

including question answering or inference, in an exam

ple of transfer learning. One example of the application

of natural language processing to health-care data is the

Comprehend Medical initiative by Amazon, which is a

service that helps extract meaningful information, such

as the presence of a gene mutation, date of symptom

onset or identification of pathologies, from unstructured

data such as EHRs.

Machine learning can be used to perform time series

analyses on longitudinal EHR data. In these analyses,

an algorithm learns prognostic signatures from histor

ical data and looks for these signatures in new datasets

to create personalized health forecasts for patients. For

example, in one study, data from cognitive tests per

formed regularly at a memory clinic were used to plot

the pattern of change associated with cognitive decline.

This information, in combination with other clinical

information, was applied to data from individuals who

were in the early stages of the disease to identify early

signs of worsening of dementia159. Such an approach

might prove particularly useful in diseases character

ized by aggressive decline and poor survival such as

MND, for which the average survival after diagnosis

is 3–5 years160. Existing statistical models of MND use

data from routine clinical assessments to predict patient

prognosis for up to 1 year161, and recent improvements

to this approach can make personalized predictions for

up to 120 months depending on the disease severity162.

These models, however, do not take into consideration

an individual’s previous medical history, which could be

informative, and missing data can result in biases.

In a study of PD progression, a Bayesian multivari

ate predictive inference platform was applied to clinical

information, including analysis of motor progression

assessments as well as complete genetic and molecular

data, collected over a 2 year period from a cohort of 117 healthy controls and 312 individuals with PD163. A total of 17,499 features were included in the model with the aim of identifying novel predictors of motor progres

sion in the early stages of PD. The progression modelling confirmed some known factors for faster motor decline, such as higher baseline motor score, male sex and older age, but it also identified new predictors, such as genetic variation and cerebrospinal fluid biomarkers.

Deep learning methods rely on the input of large quantities of data and are suited to the analysis of EHRs, which, in some cases, contain information on the major ity of a national population. RNN models have been effectively applied to EHRs to predict clinical events and improve diagnosis. For example, one study applied LSTM RNNs to data provided by the US National Alzheimer’s Coordinating Center, which includes 12 years of heterogeneous medical information on 5,432 individuals with probable AD. The study aimed to predict the AD progression stage of the next hospital visit by a patient only on the basis of the information of the patient’s historical visits. By integrating clinical data, including global staging Clinical Dementia Rating scores and the Functional Activities Questionnaire results, the algorithm could predict a patient’s AD pro gression on the next visit with over 99% accuracy, sig

nificantly outperforming classic time series forecasting methods164. Similarly, an RNN-based method was used to compare longitudinal health records from different individuals with PD and group these records according to similarity165. This kind of approach could be used to identify disease subtypes within a patient population. In another study LSTM was applied to EHR data to pre dict the length of time that patients affected by differ

ent pathologies would stay in hospital166. The algorithm performed better than traditional clinical predictive methods across different hospital wards, including neurological units, which is indicative of the versatility of machine learning approaches and shows promise for the application of this method to neurodegeneration.

Ensuring that the private data of individual patients is protected while allowing access to health records for research purposes is an ongoing challenge. In an attempt to address this challenge, machine learning was used to anonymize EHRs from a mental health-care provider and was able to mask 98.8–100% of patient identifiers that appeared in the text — the only errors resulted from misspelling of words in the original record167.

Therapy development

Effective treatments for many neurodegenerative dis eases are lacking, but the high failure rate of clinical trials for these diseases has led to the withdrawal of investment by large pharmaceutical companies168–171. For example, more than 400 clinical trials of potential treatments for AD were performed between 2002 and 2012, but only one drug, memantine, was approved172. Similarly, in the past 20 years, 50 clinical trials of drugs for MND have failed to show positive results. Riluzole and edaravone are the only drugs approved to treat MND and both have demonstrated only a modest improvement in patient survival and functional ability173. These unfortunate

**www.nature.com/nrneurol**

Bayesian inference

A method of statistical inference that uses Bayes’ theorem to calculate the probability of a hypothesis being true on the basis of observed data and prior information.

failures highlight the complexity of developing therapies for brain conditions and create opportunities for new approaches to drug development.

***Target identification.*** Neurodegenerative disorders involve a vast array of mechanisms that all contribute to disease pathology. For example, in MND, multiple processes, such as RNA metabolism, axonal transport, mitochondrial function and autophagy, are implicated in the degeneration and death of motor neurons174. The ability to explore the data related to these pathways in a thorough, holistic and efficient manner is key to under standing disease, but can be challenging for individual scientists. Machine learning can help make sense of this complexity and even predict drug targets.

One machine learning approach to drug target iden tification is relational inference on a knowledge graph, which links entities such as genes, diseases and drugs. Knowledge graphs are typically built from the integra tion of multiple data types; for example, data extracted from full text articles on PubMed, and from databases such as KEGG, OmniPath, Ensembl and ChEMBL175–178. Knowledge graph approaches can learn non-obvious links between diseases and biological drug targets (for example, identifying a new therapeutic protein target on the basis of its interaction with a protein known to be mutated in a particular disease), and are attractive because a single algorithm can be used to make predic tions for multiple diseases. One downside of using these approaches on their own is that they can lack granularity in their biological relationships (for example, context of different brain regions), which can lead to predictions with low specificity. This can be a particular problem in neuroscience, where differences in gene interaction networks between different brain regions might be important to understanding the disease pathophysiology and treatment potential179. Several relational inference methods have been published that performed well on a benchmark dataset for a wide range of disorders, includ

ing neurodegenerative diseases180. However, to date, new hypotheses generated using these approaches have not been scientifically validated.

Machine learning can also be used to perform large-scale text mining to suggest proteins that might be related to a disease of interest. In contrast to knowl edge graphs, which only take into account relationships between entities, this approach uses the entire text as substrate, thus enabling a more detailed specification of biological context. In one study, an automatic method was used to extract text features from the published lit

erature and create a model of the RNA-binding proteins (RBPs) previously associated with MND181. This model was then applied to a candidate list of other RBPs and used a network diffusion algorithm to identify those most similar to the known MND-associated RBPs. Of the ten best candidate RBPs identified by the machine learning analysis, five showed significantly altered levels in individuals with MND when compared with controls, indicating that the model’s predictions were accurate.

Machine learning-based analysis of biological sam ples (for example, post-mortem CNS tissue) might also provide useful information for target identification.

Reviews

Gene expression data from individuals with disease and healthy controls can be used to build molecular networks that visualize the biological processes that are altered in the disease state. For example, a combination of co-regulation, clustering and Bayesian inference was used to analyse transcriptomic data from brain tissue samples from individuals with late-onset AD and con trols, and identified groups of genes that were altered in the diseased tissue182. A group of immune-related and microglial-specific genes were more highly expressed in individuals with late-onset AD than in control individ

uals, and the microglial protein TYROBP was identified as a key regulator of this group. Deficiency of this pro tein was subsequently found to be neuroprotective in a mouse model of AD183,184, suggesting TYROBP as a new therapeutic target.

***Patient stratification.*** Heterogeneity in clinical man ifestation, disease progression and genetic predisposi tion often exists within groups of individuals diagnosed with the same neurodegenerative disease185. This hetero geneity makes it difficult to understand disease mecha nisms from studying the diagnostic group as a whole, as different mechanisms could be responsible for the disease in different individuals, and makes identifying effective therapies more challenging. Therefore, strat ifying study participants according to more detailed criteria than a diagnostic class is becoming more com mon. The use of machine learning techniques for this purpose is becoming increasingly popular, as the entirety of an individual’s clinical history and additional data, including, transcriptomic, neuroimaging or biomarker expression data, can be fed into the algorithm186. One approach to patient stratification using deep data might be to use unsupervised machine learning methods to reduce dimensionality in high-dimensional labelled data and derive classifiers of patient outcomes. This approach can identify patients with different subtypes or endotypes of the disease, which would otherwise not have been obvious187, for further study of disease mecha nisms or development of endotype-specific therapeutic strategies (Fig. 4).

Heterogeneity in patient populations is also a prob lem for clinical trial design. Natural heterogeneity in the outcome variable is an unhelpful source of noise that can mask the effects of a therapeutic intervention. A lack of biomarkers means that clinicians often rely on subjec tive self-reported clinical measurements for diagnosis and detecting a response to therapeutic intervention188. Therefore, using machine learning models to stratify patients and identify biomarkers of treatment response from clinical and molecular data could improve the effi cacy of clinical trials. Indeed, patient stratification and biomarker identification are major objectives of large publicly funded databases such as the ADNI. For exam ple, an approach was developed to combine multiple machine learning models that used clinical, cognitive and genetic data collected in an international, multicentre effort, to predict survival in patients with MND162. The aim of this model was to provide information that could be used to stratify patients for clinical trials. Another study used a random forest algorithm to predict disease

Nature Reviews **| Neurology**

Reviews

Electronic health records

Omics data Neuroimaging Clinical assessment **Patient stratification**

Personalized drug regimen

Then

and

Drug repurposing

+ Chemical and literature

database search

Clinical

assessment aids

Telehealth

services and

Neuroimaging based diagnosis



Biomarkers New disease

Drug for disease X avatars AD MCI

biomarkers

Fig. 4 | **Application of machine learning approaches to development and implementation of treatments for neurodegeneration.** Patient stratification, a process through which a heterogeneous patient population is separated into specific endotypes, helps identify different aetiologies of disease and the targets or hallmarks that characterize them. Unsupervised machine learning algorithms could be fed all existing patient information at the point of diagnosis, as well as new information that is gathered as the disease progresses, in order to better understand the underlying disease aetiology. During the course of therapy, any significant event such as adverse reactions, lack of response or cross-reactivity with other medication, could be fed into machine learning algorithms to help build a better understanding of both the disease and the drug, as well as the best therapeutic course of action in each patient. As machine learning algorithms continue to harness the knowledge contained within relevant literature and databases (including chemical and genetic databases), they could help build a comprehensive picture of molecular changes happening in the disease, which could consequently aid identification of new disease targets. This in turn could initiate a new drug discovery cascade, if the target is not yet druggable. Outside the realm of pharmacological treatments, machine learning algorithms could provide therapeutic avatars, for individuals with dementias or could be an indirect point of contact for patients with decreased mobility. Moreover, the field of neurochemical artificial intelligence is focusing on development of novel drugs as well as modifying and/or repurposing existing ones. AD, Alzheimer disease; MCI, mild cognitive impairment. Brain scan image credit: M. Andritoiu/Alamy Stock Photos.

progression in individuals with MND on the basis of 3 months of clinical examination data32. Although this model has not yet been used to stratify patients in a clinical trial, seeing patient data-driven models that are predictive of disease outcomes is encouraging.

Conclusions and future challenges

Machine learning algorithms can recognize patterns and make novel inferences from large amounts of multidi mensional data in a way that humans cannot. However, the use of machine learning to aid diagnosis, prognosis

**www.nature.com/nrneurol**

and therapeutic development is still in its infancy. In

the future, machine learning technologies might enable

more precise, earlier diagnosis of neurodegenerative

diseases on the basis of medical history, molecular pro

files and imaging information, and through the identi

fication of more specific diagnostic biomarkers. More

precise diagnosis could be followed by a personalized

treatment regimen that will take into consideration the

patient’s endotype. Machine learning could also reduce

the time and cost involved in performing clinical tri

als, and increase the likelihood of success, by enabling

efficient patient stratification and identifying accurate

biomarkers of treatment response. Recent advances in

machine learning technology have been made possible

by the increased availability of large, multidimensional

datasets that are curated by multicentre initiatives,

the democratization of machine learning algorithms

through open-source code and libraries, and the

increased affordability of high-performance computing

infrastructures.

Despite the potential of machine learning, creating

and applying machine learning algorithms to neurode

generative disease data remains difficult. One challenge

relates to the data itself — machine learning models

are only as powerful as the data they rely on. The lack

of large datasets, especially multidimensional patient

data, for many diseases is a barrier to the application

of machine learning. Patient datasets typically consist of

only tens or hundreds of patients and tend to be noisy

because of measurement inconsistency, error or partici

pant drop-out; these factors all make statistical analyses

more prone to errors. Metadata analyses are often neces

sary because the overlap between results from different

datasets can be small. Conducting metadata analyses

helps detect, analyse and understand inconsistencies

across datasets and increases the statistical power of

the data, as the combined dataset now includes more

individual cases189. Additionally, data are often biased

towards certain demographic populations, which limits

the generalization of the learning and results in disparity

in health care190.

Data limitations mean that, in practice, most machine

learning pipelines start with careful curation of the

data, which requires time and expert input. However,

new machine learning approaches are being developed

to address the problem of small datasets. For example,

active learning strategies enable quality inferences from

fewer samples than other machine learning methods

and remove the need to label large datasets, which

can be costly191. Active learning methods evaluate and

provide feedback on the data as the model is being

built, simultaneously making predictions for current

datasets and identifying shortcomings in the model

by suggesting additional data points to be included.

Although active learning has been used in drug design

and optimization192, and could be applied to medical

image classification or even patient response profiling,

the approach remains largely unexplored with regard to

neurodegenerative disease research. In addition, transfer

learning is a new method that uses data from one learn

ing task to help in learning another task193. For exam

ple, a transfer learning algorithm trained to diagnose

Nature Reviews **| Neurology**

Reviews

AD from T1-weighted MRI images, was applied to T2-weighted FLAIR MRI images without a drop in performance194. Transfer learning relies on the analy sis of a selected subset of data features to reduce data dimensionality, thus avoiding overfitting and making the approach suitable for use on small datasets. Transfer learning has also been used to transfer information on biomarker expression across different neurodegenera

tive diseases195, as well as for biomarker discovery from genomic data196 and diagnosis and classification of indi viduals with AD from neuroimaging data197–199. Special types of generative models, called general adversarial neural networks200, have been used to generate more training data for learning on biomedical images, for example, by generating MRI images of the brain from images obtained via different imaging platforms such as PET73. General adversarial neural networks enable such image manipulation and creation by learning the use

ful features of the training dataset and generating a new dataset based on the learnt features201,202.

Robust assessment of machine learning model per formance needs to be carried out to select the best model for the task, and to ensure that clinicians can have con fidence in the model’s output. For well-defined tasks, supervised models can be trained on labelled benchmark data (that is, sources of truth), and the performance of these models can be evaluated by comparing the model output with the benchmark data. However, many other tasks in neurology, such as patient stratification, require unsupervised models that involve no benchmark data, which means that assessing the performance of the model presents a significant challenge. Therefore, eval uation of the performance of unsupervised models uses feedback from experts to establish whether the model output is rational203, or correlation of the output with other known features such as clinical markers204. Even for well-defined tasks, benchmark data are often sparse, meaning that the performance of the model on bench mark data does not necessarily represent the perfor mance of the model on new datasets. This sparsity of benchmark data is especially concerning because some machine learning models are known to be prone to over fitting, which means they become specific to the bench mark data and do not perform well on new datasets. If a machine learning model cannot reliably generalize to new, unseen scenarios, the practical applications are limited. Poor methods of performance evaluation can also lead to over-interpretation of findings and incorrect assumptions about causality.

Another limitation of many machine learning algo rithms is that they are ‘black boxes’, that is, they cannot be used to understand the problem they address or the outputs they produce. Although algorithms are trained using medical knowledge and expertise, explaining exactly why the algorithm performed in a given way is not possible205,206. For example, understanding why a deep learning algorithm labelled certain retinal images as showing hallmarks of retinopathy is impossible, even though the predictions might later prove accu

rate on review by experts207. This lack of transparency can severely limit the usefulness of machine learning outputs and therefore the willingness of researchers to

Reviews

adopt these approaches. Fortunately, explainable AI,

which aims to build models that can be interpreted and

explained, is a growing field208. In explainable AI, algo

rithms trace or rationalize their decision-making in a

way that can be understood by humans.

Resolving the challenges involved in applying

machine learning to neurodegenerative disease data will

require collaboration between experts in biomedicine

and machine learning. For example, selecting the right

datasets for training and validation, and knowing how

to deal with missing data require a deep understand

ing of data collection procedures. In order to respond

to the pressing demands of developing machine learn

ing systems in a highly complex and often ambiguous

space, more cross-disciplinary training programmes are

needed. In addition, given the caveats of using and eval

uating machine learning technologies, it would be wise

to create industry-wide AI assessment and certification

tests to ensure that only robust, well-validated technol

ogy can impact research or patient care. The widespread

integration of machine learning into health-care settings

would also pose several practical challenges. For exam

ple, implementation of new systems into clinics needs

to take into account concerns around job security and

career progression of existing health-care personnel, and training needs to be conducted in a way that allows personnel to adapt to the new technology. In a survey of physician perspectives on AI implementation in clinical practice, the commercialization of medical AI systems and the legal, as well as ethical, responsibility of vendors of these systems were amongst the main concerns raised, which highlights the need for appropriate regulatory bodies to be introduced209.

In conclusion, the integration of machine learning into diagnostic and prognostic neurology practice, as well as the design of future therapeutics, is likely to be achieved with national and international efforts to establish multidisciplinary groups of experts to tackle some of the main challenges discussed in this Review article. The potential of these collaborative efforts is immense when we consider the health challenges our society will have to face in the next 50 years as a result of an increasing ageing population, and the benefits that would result from the introduction of faster and more accurate health-care tools for disease diagnosis and therapy implementation.

Published online xx xx xxxx

1. McCarthy, J. Basic questions. *What is Artificial Intelligence?* http://www-formal.stanford.edu/jmc/ whatisai/node1.html (2007).

2. Agatonovic-Kustrin, S. & Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **22**, 717–727 (2000).

3. Yu, K. H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2**, 719–731 (2018).

4. McDougall, R. J. Computer knows best? The need for value-flexibility in medical AI. *J. Med. Ethics* **45**, 156–160 (2019).

5. McDougall, R. J. No we shouldn’t be afraid of medical AI; it involves risks and opportunities. *J. Med. Ethics* **45**, 559 (2019).

6. Vellido, A. Societal issues concerning the application of artificial intelligence in medicine. *Kidney Dis.* **5**, 11–17 (2019).

7. Di Nucci, E. Should we be afraid of medical AI? *J. Med. Ethics* **45**, 556–558 (2019).

8. de Saint Laurent, C. In defence of machine learning: debunking the myths of artificial intelligence. *Eur. J. Psychol.* **14**, 734–747 (2018).

9. Buch, V. H., Ahmed, I. & Maruthappu, M. Artificial intelligence in medicine: current trends and future possibilities. *Br. J. Gen. Parctice* **68**, 143–144 (2018).

10. Denaxas, S. C. & Morley, K. I. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur. Heart. J. Qual. Care Clin. Outcomes* **1**, 9–16 (2015).

11. Weber, G., Mandl, K. & Kohane, I. Finding the missing link for big biomedical data. *JAMA* **311**, 2479–2480 (2014).

12. Van Horn, J. & Toga, A. Human neuroimaging as a “big data” science. *Brain Imaging Behav.* **8**, 323–331 (2014).

13. Zhou, L. & Verstreken, P. Reprogramming neurodegeneration in the big data era. *Curr. Opin. Neurobiol.* **48**, 167–173 (2018).

14. Vallejos, C. A., Richardson, S. & Marioni, J. C. Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol.* **17**, 1–14 (2016).

15. Ritchie, M. D. et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* **69**, 138–147 (2001).

16. Xu, J., Zhang, Y., Qiu, C. & Cheng, F. Global and regional economic costs of dementia: a systematic review [abstract]. *Lancet* **390**, S47 (2017).

17. Prince, M., Prina, M. & Guerchet, M. *World Alzheimer’s Report 2013. The Journey of Caring:*

*An Analysis of Long-Term Care for Dementia*

(Alzheimer’s Disease International, 2013).

18. Bishop, C. *Pattern Recognition and Machine Learning* (Springer, 2006).

19. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

**This extensive review provides an elegant summary of deep learning methods and their application to images, video footage, speech recordings and written text**.

20. Van Der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

21. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat*. *Biotechnol*. **37**, 38–44 (2018).

22. Oyelade, J. et al. Clustering algorithms: their application to gene expression data. *Bioinform. Biol. Insights* **10**, 237–253 (2016).

23. Chapelle, O., Schölkopf, B. & Zien, A. (eds) *Semi-Supervised Learning* (MIT Press, 2006). 24. Vapnik, V. *Statistical Learning Theory* (Wiley-Interscience, 1998).

25. Joachmis, T. in *ICML ’99: Proceedings of the Sixteenth International Conference on Machine Learning* (eds Bratko, I. & Dzeroski, S.) 200–209 (Morgan Kaufmann, 1999).

26. Watkins, C. J. C. H. *Learning with Delayed Rewards*. Thesis, King’s College, Cambridge (1989).

27. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).

28. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de-novo drug design. *Sci. Adv.* **4**, 1–14 (2017).

29. Raudys, Š. *Statistical and Neural Classifiers: An Integrated Approach to Design* (Springer, 2001). 30. Summers, M. J. et al. Deep machine learning application to the detection of preclinical

neurodegenerative diseases of aging. *Sci. J. Digit. Cult.* **2**, 9–24 (2017).

31. Ho, T. K. Random decision forests perceptron training. in *ICDAR ’95: Proceedings of the Third International Conference on Document Analysis and Recognition* 278–282 (IEEE Computer Society, 1995).

32. Hothorn, T. & Jung, H. H. RandomForest4Life: a random forest for predicting ALS disease

progression. *Amyotroph. Lateral Scler. Front. Degener.* **15**, 444–452 (2014).

33. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **297**, 273–297 (1995).

34. Rosenblatt, F. *The Perceptron – A Perceiving and Rocognizing Automation* (Cornell Aeronautical Laboratory, 1957).

35. McCulloch, W. S. & Pitts, W. A logical calculus of the idea immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).

**This paper describes the first steps towards mathematical modelling of neuronal function, which eventually resulted in the development of artificial neural networks**.

36. Fukushima, K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).

37. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).

38. LeCun, Y., Hafner, P., Bottou, L. & Bengio, Y. in *Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science* Vol 1681 (eds Forsyth, D. A., Mundy, J. L., di Gesú, V. & Cipolla, R.) 319–345 (Springer, 1999).

39. Burt, J. R. et al. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. *Br. J. Radiol.* **91**, 2–11 (2018).

40. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).

41. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997). 42. Cho, K. et al. in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language*

*Processing (EMNLP)* 1724–1734 (Association for Computational Linguistics, 2014).

43. Cawley, G. C. & Talbot, N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11**, 2079–2107 (2010).

44. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, 1–17 (2017).

45. Neumaier, A. Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Rev.* **40**, 636–666 (1998).

46. Michel, P. P., Hirsch, E. C. & Hunot, S. Understanding dopaminergic cell death pathways in Parkinson disease. *Neuron* **90**, 675–691 (2016).

47. Donev, R., Kolev, M., Millet, B. & Thome, J. Neuronal death in Alzheimer’s disease and therapeutic opportunities. *J. Cell. Mol. Med.* **13**, 4329–4348 (2009).

48. Fischer, L. R. et al. Amyotrophic lateral sclerosis is a distal axonopathy: evidence in mice and man. *Exp. Neurol.* **182**, 232–240 (2004).

49. Hainc, N. et al. The bright, artificial intelligence augmented future of neuroimaging reading. *Front. Neurol.* **8**, 10–12 (2017).

**www.nature.com/nrneurol**

50. Grenander, U., Chow, Y. & Keenan, D. *HANDS: a Pattern Theoretic Study of Biological Shapes* (Springer, 1990).

51. Evans, A. C., Marrett, S., Torrescorzo, J., Ku, S. & Collins, L. MRI-PET correlation in three dimensions using a volume-of-interest (VOI) atlas. *J. Cereb. Blood Flow. Metab.* **11**, A69–A78 (1991).

52. Woods, R. P., Mazziotta, J. C. & Cherry, S. R. MRI-PET registration with automated algorithm. *J. Comput. Assist. Tomogr.* **17**, 536–546 (1993).

53. Joshi, S. C. et al. Hierarchical brain mapping via a generalized dirichlet solution for mapping brain manifolds. in *Proceedings of the SPIE’s 1995*

*international symposium on optical science*,

*engineering, and instrumentation. Vision geometry IV* Vol. 2573 (eds Melter, R. A., Wu, A. Y., Bookstein, F. L. & Green, W. D. K.) 278–289 (SPIE, 1995).

54. Grady, C. L. et al. Subgroups in dementia of the Alzheimer type identified using positron emission tomography. *J. Neuropsychiatry Clin. Neurosci.* **2**, 373–384 (1990).

55. DeFigueiredo, R. J. P. et al. Neural-network-based classification of cognitively normal, demented, Alzheimer disease and vascular dementia from single photon emission with computed tomography image data from brain. *Proc. Natl Acad. Sci. USA* **92**, 5530–5534 (1995).

**This study is one of the first to have used an artificial neural network algorithm to automate the identification of normal ageing, AD and vascular dementia from SPECT data**.

56. Wang, S. et al. Pathological brain detection by artificial intelligence in magnetic resonance imaging scanning. *Prog. Electromagn. Res.* **156**, 105–133 (2016).

57. Haller, J. W. et al. Hippocampal MR imaging morphometry by means of general pattern matching. *Radiology* **199**, 787–791 (1996).

58. Davatzikos, C. et al. A computerized approach for morphological analysis of the corpus callosum. *J. Comput. Assist. Tomogr.* **20**, 88–97 (1996).

59. Gur, R. C. et al. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *J. Neurosci.* **19**, 4065–4072 (1999).

60. Mega, M. S. et al. Cerebral correlates of psychotic symptoms in Alzheimer’s disease. *J. Neurol.*

*Neurosurg. Psychiatry* **69**, 167–171 (2000).

61. Fischl, B. & Dale, A. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl Acad. Sci. USA* **97**, 11050–11055 (2000).

62. Ashburner, J. & Friston, K. Voxel-based morphometry– the methods. *Neuroimage* **11**, 805–821 (2000). 63. Haller, J. W. et al. Three-dimensional hippocampal volumetry by high dimensional transformation of a neuroanatomical atlas. *Radiology* **202**, 504–510 (1997).

64. Maldijan, J. A., Laurienti, P. J., Kraft, R. A. & Burdette, J. H. An automated method for

neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* **19**, 1233–1239 (2003).

**The paper presents the first automated analysis system based on a digital brain atlas to show robust application to fMRI data, without the need for pre-definied region of interest masks**.

65. Lao, Z. et al. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage* **21**, 46–57 (2004). **This paper presents an early application of SVM to MR image analysis and highlights the importance of analysing all voxels simultaneously, rather than focusing on a pre-defined region of interest.**

66. Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H. & Stetter, M. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *Neuroimage* **28**, 980–995 (2005).

**This study is an early demonstration of the superior performance of SVM over traditional statistical methods for MRI analysis and highlights the ability of SVM to select the brain regions from which the most accurate classification can be drawn**.

67. Mitchell, T. M. et al. Learning to decode cognitive states from brain images. *Mach. Learn.* **57**, 145–175 (2004).

**In this study multiple machine learning algorithms, including SVM, are used on functional MR images to assess the feasibility of detecting patients’ transient cognitive states during a single time interval**.

Nature Reviews **| Neurology**

68. Reczko, M., Karras, D. A., Mertzios, B. G., Graveron Demilly, D. & Van Ormondt, D. Improved MR image reconstruction from sparsely sampled scans based on neural networks. *Pattern Recognit. Lett.* **22**, 35–46 (2001).

69. Zhu, G. et al. Applications of deep learning to neuro-imaging techniques. *Front. Neurol.* **10**, 1–13 (2019).

70. He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)* 770–778 (IEEE, 2016).

71. Gray, K. R. et al. Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. *Neuroimage* **65**, 167–175 (2013).

72. Korolev, S., Safiullin, A., Belyaev, M. & Dodonova, Y. in *Proceedings of the 14th International Symposium on Biomedical Imaging* 835–838 (IEEE, 2017).

73. Choi, H., Kang, H. & Lee, D. S. Predicting aging of brain metabolic topography using variational

autoencoder. *Front. Aging Neurosci.* **10**, 212 (2018). 74. Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **29**, 102–127 (2019).

75. Klöppel, S. et al. Automatic classification of MR scans in Alzheimer’s disease. *Brain* **131**, 681–689 (2008). **This study shows that an SVM can use MR scans to successfully distinguish between individuals with AD and individuals with FTLD as well as between individuals with AD and healthy individuals**.

76. Bron, E. E., Smits, M., Niessen, W. J. & Klein, S. Feature selection based on the SVM weight vector for classification of dementia. *IEEE J. Biomed. Heal. Inform.* **19**, 1617–1626 (2015).

77. Moradi, E., Pepe, A., Gaser, C., Huttunen, H. & Tohka, J. Machine learning framework for early MRI based Alzheimer’s conversion prediction in MCI subjects. *Neuroimage* **104**, 398–412 (2015).

78. Magnin, B. et al. Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI. *Neuroradiology* **51**, 78–83 (2009).

79. Gerardin, E. et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer’s disease and mild cognitive impairment from normal aging. *Neuroimage* **47**, 1476–1486 (2009).

80. Li, S. et al. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. *Am. J. Neuroradiol.* **28**, 1339–1345 (2007).

81. Amoroso, N. et al. Alzheimer’s disease diagnosis based on the hippocampal unified multi-atlas network (HUMAN) algorithm. *Biomed. Eng. Online* **17**, 1–16 (2018).

82. De Marco, M., Beltrachini, L., Biancardi, A., Frangi, A. F. & Venneri, A. Machine-learning support to individual diagnosis of mild cognitive impairment using

multimodal MRI and cognitive assessments. *Alzheimer Dis. Assoc. Disord.* **31**, 278–286 (2017).

83. Ahn, W., Krawitz, A. & Kim, W. A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *J. Neurosci. Psychol. Econ.* **4**, 95–110 (2011).

84. Rehme, A. K. et al. Identifying neuroimaging markers of motor disability in acute stroke by machine learning techniques. *Cereb. Cortex* **25**, 3046–3056 (2014).

85. Weygandt, M. et al. MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS One* **6**, e21138 (2011).

86. Duchesne, S., Rolland, Y. & Vérin, M. Automated computer differential classification in Parkinsonian syndromes via pattern analysis on MRI. *Acad. Radiol.* **16**, 61–70 (2009).

87. Chen, L. et al. Rapid automated quantification of cerebral leukoaraiosis on CT images: a multicenter validation study. *Radiology* **288**, 573–581 (2018).

88. Prevedello, L. M., Little, K. J., Qian, S. & White, R. D. Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology* **285**, 923–931 (2017).

89. Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat. Med.* **24**, 1337–1341 (2018).

90. Chilamkurthy, S. et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* **392**, 2388–2396 (2018).

91. Davatzikos, C., Fan, Y., Wu, X., Shen, D. & Resnick, S. M. Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging* **29**, 514–523 (2008).

92. Fan, Y., Resnick, S. M., Wu, X. & Davatzikos, C. Structural and functional biomarkers of prodromal Alzheimer’s disease. *Neuroimage* **41**, 277–285 (2008).

Reviews

93. Fan, Y., Batmanghelich, N. K., Clark, C. M. & Davatzikos, C., Alzeimer’s Disease Neuroimaging Initiative. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* **39**, 1731–1743 (2008).

94. Kivipelto, M. et al. The Finnish geriatric intervention study to prevent cognitive impairment and disability (FINGER): study design and progress. *Alzheimer’s Dement.* **9**, 657–665 (2013).

95. Zhang, Y. C. & Kagen, A. C. Machine learning interface for medical image analysis. *J. Digit. Imaging* **30**, 615–621 (2017).

96. Mufford, M. S. et al. Neuroimaging genomics in psychiatry–a translational approach. *Genome Med.* **9**, 1–12 (2017).

97. Bookheimer, S. Y. et al. Patterns of brain activation in people at risk for Alzheimer’s disease. *N. Engl. J. Med.* **343**, 450–456 (2000).

98. Heinz, A. et al. Genotype influences in vivo dopamine transporter availability in human striatum.

*Neuropsychopharmacology* **22**, 133–139 (2000). 99. Liang, Z. & Lauterbur, P. *Principles of Magnetic Resonance Imaging: a Signal Processing Approach* (IEEE, 2000).

100. Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).

101. Czigler, B. et al. Quantitative EEG in early Alzheimer’s disease patients – power spectrum and complexity features. *Int. J. Psychophysiol.* **68**, 75–80 (2008).

102. Lee, H., Brekelmans, G. J. F. & Roks, G. The EEG as a diagnostic tool in distinguishing between dementia with Lewy bodies and Alzheimer’s disease. *Clin. Neurophysiol.* **126**, 1735–1739 (2015).

103. Barcelon, E. A. et al. Grand total EEG score can differentiate Parkinson’s disease from Parkinson related disorders. *Front. Neurol.* **10**, 1–11 (2019).

104. Buscema, M. et al. An improved I-FAST system for the diagnosis of Alzheimer’s disease from unprocessed electroencephalograms by using robust invariant features. *Artif. Intell. Med.* **64**, 59–74 (2015).

105. Bosco, D. A., LaVoie, M. J., Petsko, G. A. & Ringe, D. Proteostasis and movement disorders: Parkinson’s disease and amyotrophic lateral sclerosis. *Cold Spring Harb. Perspect. Biol.* **3**, 1–24 (2011).

106. Ross, C. A. & Tabrizi, S. J. Huntington’s disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.* **10**, 83–98 (2011).

107. [No authors listed.]. The amyotrophic lateral sclerosis functional rating scale: assessment of activities of daily living in patients with amyotrophic lateral sclerosis. *Arch. Neurol.* **53**, 141–147 (1996).

108. [No authors listed.]. Unified Huntington’s disease rating scale: reliability and consistency. *Mov. Disord.* **11**, 136–142 (1996).

109. Fahn, S., Elton, R. & Members of the UPDRS Development Committee. in *Recent Developments in Parkinson’s Disease* Vol. 2 (eds. Fahn, S., Marsden, C. D., Calne, D. B. & Goldstein, M.) 153–163, 293–304 (Macmillan Health Care Information, 1987).

110. Davenport, T. & Kalakota, R. The potential for artificial intelligence in healthcare. *Futur. Healthc. J.* **6**, 94–98 (2019).

111. Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I. & Schlesinger, I. Handwriting as an objective tool for Parkinson’s disease diagnosis. *J. Neurol.* **260**, 2357–2361 (2013).

112. Alty, J., Cosgrove, J., Thorpe, D. & Kempster, P. How to use pen and paper tasks to aid tremor diagnosis in the clinic. *Pract. Neurol.* **17**, 456–463 (2017).

113. McLennan, J., Nakano, K., Tyler, H. & Schwab, R. Micrographia in Parkinson’s disease. *J. Neurol. Sci.* **15**, 141–152 (1972).

114. Kotsavasiloglou, C., Kostikis, N., Hristu-Varsakelis, D. & Arnaoutoglou, M. Machine learning-based

classification of simple drawing movements in Parkinson’s disease. *Biomed. Signal. Process. Control.* **31**, 174–180 (2017).

**This is the first study to have used a combination of simple line drawings and machine learning algorithms to aid PD diagnosis**.

115. Westin, J. et al. A new computer method for assessing drawing impairment in Parkinson’s disease. *J. Neurosci. Methods* **190**, 143–148 (2010).

116. Griffiths, R. I., Kotschet, K., Arfon, S., Ming, Z. & Johnson, W. Automated assessment of bradykinesia and dyskinesia in Parkinson’s disease. *J. Parkinson’s Dis.* **2**, 47–55 (2012).

117. Giuffrida, J. P., Riley, D. E., Maddux, B. N. & Heldman, D. A. Clinically deployable KinesiaTM

Reviews

technology for automated tremor assessment. *Mov. Disord.* **24**, 723–730 (2009).

118. Jeon, H., Lee, W. & Park, H. High-accuracy automatic classification of parkinsonian tremor severity using machine learning method. *Physiol. Meas.* **38**, 1980–1999 (2017).

119. Zhao, A., Qi, L., Dong, J. & Yu, H. Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases. *Knowl. Syst.* **145**, 91–97 (2018).

120. Pushparani, M. & Athisakthi, A. Detection of movement disorders using multi SVM. *Glob. J. Comput. Sci. Technol.* **13**, 23–25 (2013).

121. Sacco, G. et al. Detection of activities of daily living impairment in Alzheimer’s disease and mild cognitive impairment using information and communication technology. *Clin. Interv. Ageing* **7**, 539–549 (2012).

122. Ji, S., Xu, W., Yang, M. & Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 221–231 (2013).

123. Raya, Z. et al. in *Proceedings of SPIE: Applications of Machine Learning* Vol. 11139 (eds Zelinski, M. E., Taha, T. M., Howe, J., Awwal, A. A. S. & Iftekharuddin, K. M.) 1113909 (SPIE, 2019).

124. Brand, D., DiGennaro Reed, F. D., Morley, M. D., Erath, T. G. & Novak, M. D. A survey assessing privacy concerns of smart-home services provided to individuals with disabilities. *Behav. Anal. Pract.* **13**, 11–21 (2020).

125. Riboni, D., Bettini, C., Civitarese, G., Janjua, Z. H. & Helaoui, R. SmartFABER: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artif. Intell. Med.* **67**, 57–74 (2016).

126. Ordóñez, F. J. & Roggen, D. Deep convolutional and LSTM recurrent activity recognition. *Sensors* **16**, 115–140 (2016).

127. Alam, R., Homdee, N., Wolfe, S., Hayes, J. & Lach, J. In *IoTDI 2019: Proceedings of the International Conference on Internet of Things Design and*

*Implementation* 281–282 (Association for Computing Machinery, 2019).

128. Rankin, K. P., Baldwin, E., Pace-Savitsky, C., Kramer, J. H. & BL, M. Self awareness and personality change in dementia. *J. Neurol. Neurosurg. Psychiatry* **76**, 632–639 (2005).

129. Sollberger, M. et al. Neural basis of interpersonal traits in neurodegenerative diseases. *Neuropsychologia* **47**, 2812–2827 (2009).

130. Christidi, F., Migliaccio, R., Santamaría-García, H., Santangelo, G. & Trojsi, F. Social cognition dysfunctions in neurodegenerative diseases: neuroanatomical correlates and clinical implications. *Behav. Neurol.* **2018**, 18 (2018).

131. Orimaye, S., Wong, J. & Golden, K. in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* 78–87 (Association for Computational Linguistics, 2014).

132. Wankerl, S., Nöth, E. & Evert, S. in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2017* 3162–3166 (Association for Computational

Linguistics, 2017).

133. Weizenbaum, J. ELIZA — a computer program for the study of natural language communication between man and machine. *Commun. ACM* **26**, 23–28 (1983). **This article describes the first question-and-answer computer program, which paved the way for AI-driven avatars as we know them today**.

134. Ireland, D. et al. Hello Harlie: enabling speech monitoring through chat-bot conversations.

*Stud. Health Technol. Inform.* **227**, 55–60 (2016). 135. Tanaka, H. et al. Detecting dementia through interactive computer avatars. *IEEE J. Transl. Eng. Heal. Med.* **5**, 1–11 (2017).

136. Blackburn, D. et al. An avatar aid in memory clinic [abstract PO029]. *J. Neurol. Neurosurg. Psychiatry* **88**, A19–A20 (2017).

137. Schmidtke, K., Pohlmann, S. & Metternich, B. The syndrome of functional memory disorder: definition, etiology, and natural course. *Am. J. Geriatr. Psychiatry* **16**, 981–988 (2008).

138. Mahley, R. W., Weisgraber, K. H. & Huang, Y. Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer’s disease. *Proc. Natl Acad. Sci. USA* **103**, 5644–5651 (2006).

139. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med.* **18**, 421–430 (2016).

140. Huang, X. et al. Revealing Alzheimer’s disease genes spectrum in the whole-genome by machine learning. *BMC Neurol.* **18**, 1–8 (2018).

141. Maj, C. et al. Integration of machine learning methods to dissect genetically imputed transcriptomic profiles in Alzheimer’s disease. *Front. Genet.* **10**, 1–16 (2019).

142. Lopez, C., Tucker, S., Salameh, T. & Tucker, C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *J. Biomed. Inform.* **85**, 30–39 (2018).

143. Ray, S. et al. Classification and prediction of clinical Alzheimer’s diagnosis based on plasma signaling proteins. *Nat. Med.* **13**, 1359–1362 (2007).

144. Agarwal, S., Ghanty, P. & Pal, N. R. Identification of a small set of plasma signalling proteins using neural network for prediction of Alzheimer’s disease. *Bioinformatics* **31**, 2505–2513 (2015).

145. Andersen, S. L. et al. Metabolome-based signature of disease pathology in MS. *Mult. Scler. Relat. Disord.* **31**, 12–21 (2019).

146. Sapkota, S. et al. Alzheimer’s biomarkers from multiple modalities selectively discriminate clinical status: relative importance of salivary metabolomics panels, genetic, lifestyle, cognitive, functional health and demographic risk markers. *Front. Aging Neurosci.* **10**, 1–13 (2018).

147. Tavares, J. & Oliveira, T. Electronic health record portal adoption: a cross country analysis. *BMC Med. Inform. Decis. Mak.* **17**, 1–17 (2017).

148. Stone, C. P. A glimpse at EHR implementation around the world: the lessons the US can learn. *e-healthpolicy. org* https://www.e-healthpolicy.org/sites/e-healthpolicy. org/files/A\_Glimpse\_at\_EHR\_Implementation\_ Around\_the\_World1\_ChrisStone.pdf (2014).

149. Chen, Y. et al. Applying active learning to high throughput phenotyping algorithms for electronic health records data. *J. Am. Med. Inform. Assoc.* **20**, 253–259 (2013).

150. Schank, R. C. & Tesler, L. in *Proceedings of the 1969 Conference on Computational linguistics* 1–3 (Association for Computational Linguistics, 1969).

151. Winograd, T. Procedures as a representation for data in a computer program for understanding natural language (Massachusetts Institute of Technology, 1971).

152. Schank, R. C. Computer understanding of natural language. *Behav. Res. Methods Instrum.* **10**,

132–138 (1978).

153. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at *arXiv* https:// arxiv.org/abs/1810.04805 (2018).

154. Manning, C. et al. in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* 55–60

(Association for Computational Linguistics, 2014). 155. Honnibal, M. & Johnson, M. in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* 1373–1378 (Association for Computational Linguistics, 2015).

156. Petrov, S. Announcing syntaxnet: the world’s most accurate parser goes open source. *Google AI Blog* https://ai.googleblog.com/2016/05/announcing syntaxnet-worlds-most.html (2016).

157. Ford, E., Carroll, J. A., Smith, H. E., Scott, D. & Cassell, J. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J. Am. Med. Inform. Assoc.* **23**, 1007–1015 (2016).

158. Weissenbacher, D. et al. in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 1198–1207 (Association for Computational Linguistics, 2016).

159. Grassi, M. et al. A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer’s disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* **10**, 1–15 (2019).

160. Gordon, P. H. & Meininger, V. How can we improve clinical trials in amyotrophic lateral sclerosis? *Nat. Rev. Neurol.* **7**, 650–654 (2011).

161. Moura, M. C., Casulari, L. A., Rita, M. & Garbi, C. A predictive model for prognosis in motor neuron disease. *J. Neurol. Disord.* **4**, 4–10 (2016).

162. Westeneng, H.-J. et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol.* **17**, 423–433 (2018).

**This study shows how the application of machine learning to large clinical datasets from various**

**clinical centres enables the prediction of disease prognosis in individuals with amyotrophic lateral sclerosis**.

163. Latourelle, J. C. et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson’s disease: a longitudinal cohort study and validation. *Lancet Neurol.* **16**, 908–916 (2017).

**This study exemplifies how integration of large clinical, molecular and genetic longitudinal datasets can be used to provide information on disease progression in PD**.

164. Wang, T., Qiu, R. G. & Yu, M. Predictive modeling of the progression of Alzheimer’s disease with recurrent neural networks. *Sci. Rep.* **8**, 1–12 (2018).

165. Che, C. et al. in *Proceedings of the 2017 SIAM International Conference on Data Mining* 198–206 (SIAM, 2017).

166. Rajkomar, A. et al. Scalable and accurate deep learning for electronic health records. *NPJ Digit. Med.* **1**, 1–10 (2018).

167. Fernandes, A. C. et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC Med. Inform. Decis. Mak.* **13**, 1–14 (2013).

168. [No authors listed]. Stimulus package. *Nat. Med*. **24**, 247 (2018).

169. Zwierzyna, M., Davies, M., Hingorani, A. D. & Hunter, J. Clinical trial design and dissemination: comprehensive analysis of clinicaltrials.gov and PubMed data since 2005. *BMJ* **361**, 1–11 (2018).

170. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).

171. Cummings, J. Lessons learned from Alzheimer disease: clinical trials with negative outcomes. *Clin. Transl. Sci.* **11**, 147–152 (2018).

172. Cummings, J. L., Morstorf, T. & Zhong, K. Alzheimer’s disease drug-development pipeline: few candidates, frequent failures. *Alzheimers. Res. Ther.* **6**, 1–7 (2014).

173. Mitsumoto, H., Brooks, B. R. & Silani, V. Clinical trials in amyotrophic lateral sclerosis: why so many negative trials and how can trials be improved? *Lancet Neurol.* **13**, 1127–1138 (2014).

174. Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M. & Shaw, P. J. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7**, 616–630 (2011).

175. Neil, D. et al. Interpretable graph convolutional neural networks for inference on noisy knowledge graphs. Preprint at *arXiv* https://arxiv.org/abs/1812.00279 (2018).

176. Zitnik, M., Agrawal, M. & Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **34**, i457–i466 (2018).

177. Duvenaud, D. et al. in *NIPS’15: Proceedings of the 28th International Conference on Neural Information Processing Systems* Vol. 2 2224–2232 (Neural Information Processing Systems Foundation, 2015).

178. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Preprint at *arXiv* https://arxiv.org/abs/1609.02907 (2017).

179. Palop, J. J., Chin, J. & Mucke, L. A network dysfunction perspective on neurodegenerative diseases. *Nature* **443**, 768–773 (2006).

180. Zakeri, P., Simm, J., Arany, A., Elshal, S. & Moreau, Y. Gene prioritization using Bayesian matrix factorization with genomic and phenotypic side information. *Bioinformatics* **34**, i447–i456 (2018).

181. Bakkar, N. et al. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis.

*Acta Neuropathol.* **135**, 227–247 (2018).

182. Zhang, B. et al. Resource integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer’s disease. *Cell* **153**, 707–720 (2013).

**This study exemplifies how machine learning approaches applied to omics data can lead to identification of new therapeutic targets**.

183. Haure-Mirande, J. V. et al. Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early Alzheimer’s pathology. *Acta Neuropathol.* **134**, 769–788 (2017).

184. Haure-Mirande, J. V. et al. Integrative approach to sporadic Alzheimer’s disease: deficiency of TYROBP in cerebral Aβ amyloidosis mouse normalizes clinical phenotype and complement subnetwork molecular

**www.nature.com/nrneurol**

pathology without reducing Aβ burden. *Mol. Psychiatry* **24**, 431–446 (2019).

185. Wauters, E. et al. Neurobiology of aging clinical variability and onset age modifiers in an extended Belgian GRN founder family. *Neurobiol. Aging* **67**, 84–94 (2018).

186. Grollemund, V. et al. Machine learning in amyotrophic lateral sclerosis: achievements, pitfalls, and future directions. *Front. Neurosci.* **13**, 1–28 (2019).

187. Maudsley, S., Devanarayan, V., Martin, B. & Geerts, H. Intelligent and effective informatic deconvolution of ‘big data’ and its future impact on the quantitative nature of neurodegenerative disease therapy. *Alzheimer’s Dement.* **14**, 961–975 (2018).

188. Meyer, S. et al. Optimizing ADAS-Cog worksheets: a survey of clinical trial raters’ perceptions.

*Curr. Alzheimer Res.* **14**, 1008–1016 (2017). 189. McDermott, J. E. et al. Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. *Expert. Opin. Med. Diagn.* **7**, 37–51 (2013).

190. Popejoy, A. & Fullerton, S. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).

191. Cohn, D. A., Ghahramani, Z. & Jordan, M. I. Active learning with statistical models. *J. Artif. Intell. Res.* **4**, 129–145 (1996).

192. Sellwood, M. A., Ahmed, M., Segler, M. H. S. & Brown, N. Artificial intelligence in drug discovery. *Future Med. Chem.* **10**, 2025–2028 (2018).

193. Gupta, A., Ayhan, M. S. & Maida, A. S. Natural image bases to represent neuroimaging data. *PMLR* **28**, 987–994 (2013).

194. Xu, Y., Raj, A. & Victor, J. D. Systematic differences between perceptually relevant image statistics of brain MRI and natural images. *Front. Neuroinform.* **13**, 1–15 (2019).

195. Marinescu, R. V. et al. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science* Vol. 11765 (eds Shen, D. et al.) 860–868 (Springer, 2019).

196. Ganchev, P., Malehorn, D., Bigbee, W. L. & Gopalakrishnan, V. Transfer learning of classification rules for biomarker discovery and verification from molecular profiling studies. *J. Biomed. Inform.* **44**, S17–S23 (2011).

197. Young, J. et al. Accurate multimodal probabilistic prediction of conversion to Alzheimer’s disease in patients with mild cognitive impairment. *NeuroImage Clin.* **19**, 735–745 (2013).

198. Cheng, B. et al. Multi-domain transfer learning for early diagnosis of Alzheimer’s disease. *Neuroinformatics* **15**, 115–132 (2017).

Nature Reviews **| Neurology**

199. Hon, M. & Khan, N. Towards Alzheimer’s disease classification through transfer learning. Preprint at *arXiv* https://arxiv.org/abs/1711.11117 (2017).

200. Goodfellow, I. J. et al. Generative adversarial nets. *Neural Inf. Process. Syst.* **27**, 1–9 (2014).

201. Huang, H., Yu, P. S. & Wang, C. An introduction to image synthesis with generative adversarial nets. Preprint at *arXiv* https://arxiv.org/abs/1803.04469 (2018).

202. Kazuhiro, K. et al. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. *Tomography* **4**, 159–163 (2018).

203. Palacio-Niño, J.-O. & Berzal, F. Evaluation metrics for unsupervised learning algorithms. Preprint at *arXiv* https://arxiv.org/abs/1905.05667 (2019).

204. Lötsch, J., Lerch, F., Djaldetti, R., Tegder, I. & Ultsch, A. Identification of disease-distinct complex biomarker patterns by means of unsupervised machine-learning using an interactive R toolbox (Umatrix). *Big Data Anal.* **3**, 1–17 (2018).

205. Ravi, D. et al. Deep learning for health informatics. *IEEE J. Biomed. Heal. Inform.* **21**, 4–21 (2017). 206. Vial, A. et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Transl. Cancer Res.* **7**, 803–816 (2018).

207. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).

208. Gilpin, L. H. et al. Explaining explanations: an overview of interpretability of machine learning. Preprint at *arXiv* https://arxiv.org/abs/1806.00069 (2018).

209. Sarwar, S. et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *NPJ Digit. Med.* **2**, 28 (2019).

**This article reports the perspective of pathologists towards the integration of artificial intelligence into diagnostic pathology**.

210. Fan, Y., Shen, D. & Davatzikos, C. in *Lecture Notes in Computer Science*, Vol. 3749 (eds Duncan, J. S. & Gerig, G.) 1–8 (Springer, 2005).

211. Shi, B., Chen, Y., Zhang, P., Smith, C. D. & Liu, J. Nonlinear feature transformation and deep fusion for Alzheimer’s disease staging analysis. *Pattern Recognit.* **63**, 487–498 (2017).

**Author contributions**

L.F. researched data for the article, made a substantial con tribution to the discussion of article content, wrote the article, and reviewed and edited the manuscript before submission. M.A.M. researched data for the article, made a substantial contribution to the discussion of article content and wrote the

Reviews

article. P.N.O. and J.D.H. made a substantial contribution to discussion of article content, and reviewed and edited the manuscript before submission. A.M.B.L. and D.N. researched data for the article, and reviewed and edited the manuscript before submission. A.S., R.M. and G.M.H. reviewed and edited the manuscript before submission.

**Competing interests**

M.A.M. is funded by BenevolentAI. P.N.O., A.M.B.L., D.N., A.S. and J.D.H. work for BenevolentAI. R.M. and L.F. have a project in collaboration with BenevolentAI.

**Peer review information**

*Nature Reviews Neurology* thanks S. Baranzini, D. Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher’s note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Review criteria**

A search for original articles published without date limitation was performed in PubMed. Search terms included: ‘artificial intelligence’ AND ‘neuroscience’ OR ‘neurology’ OR ‘neuro degeneration’ OR ‘neuroimaging’; ‘machine learning’ AND ‘neurodegeneration’ OR ‘Alzheimer disease’ OR ‘Parkinson disease’ OR ‘motor neuron disease’ OR ‘multiple sclerosis’ OR ‘dementia’; and ‘machine learning’ AND ‘disease diagnosis’ OR ‘disease prognosis’.

Related links

Allen Brain Atlas: http://portal.brain-map.org/

Alzheimer’s Disease Neuroimaging initiative: http://adni. loni.usc.edu/

Amazon Comprehend Medical initiative: https://aws. amazon.com/comprehend/medical/

enhancing Neuro imaging Genetics through Meta-Analysis (eNiGMA) consortium: http://enigma.ini.usc.edu/

european Alzheimer’s Disease Consortium impact of Cholinergic Treatment Use (eADC-iCTUs): http://www.eadc. info/sito/pagine/d\_01.php?nav=d

Google TensorFlow: https://www.tensorflow.org/

Parkinson’s Progression Markers initiative: http://www. ppmi-info.org/

The Finnish Geriatric intervention study to Prevent Cognitive impairment and Disability (FiNGeR): http://www. alz.org/wwfingers/overview.asp

UK Biobank: http://www.ukbiobank.ac.uk/

© Springer Nature Limited 2020