Machine Learning in Alzheimer's Disease

Statistical Mechanics & Dynamical Complex Systems Laboratory
Institute of Nanoscience and Nanotechnology of NCSR "Demokritos"

Panagiotaras Ilias

TABLE OF CONTENTS

01
Machine Learning in Neuroimaging

“Machine learning is the science of getting computers to act without being explicitly programmed.” – Stanford

“Machine Learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.” – Nvidia

“...what we want is a machine that can learn from experience.” – Alan Turing

Machine Learning Methods

Machine Learning

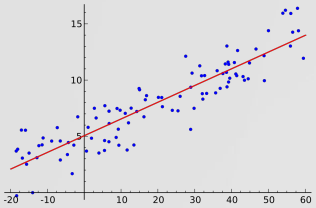
The main focus of machine learning lies in generalization, i.e. finding patterns in data by training a computational model such that it can predict unseen data of the same or similar nature.

Supervised Learning

- Maps input data to labels
- Requires labelled data to train

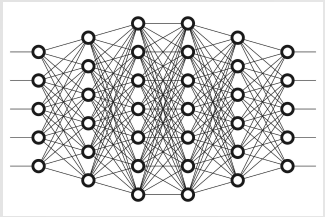
Unsupervised Learning

- Detect patterns in unlabeled data



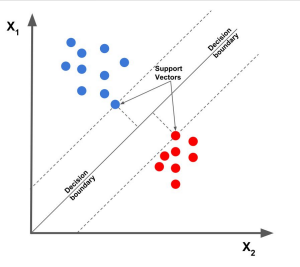
Linear Regression

A predictive technique that uses one (or many) features in order to predict a response as one continuous value



Deep Learning

The foundation of deep learning lies in artificial neural networks (more on that later...)



Support Vector Machine

The aim is to draw a decision boundary between sets of data points, so as to maximize the "separation" between the sets.

Clustering

Grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

Laplacian Eigenmaps

Explore the relationship of voxels in a region of interest based on a feature of interest.

Associative Models

Discover relationships (associations between two sets of multivariate inputs .

...and a lot more

ML in Neuroimaging

Recent "boom" in research

1. More computational power
2. More data
3. Better algorithms

While the overall data for research has increased, their number remains relatively small compared to the necessary algorithmic inputs.

Large data - few samples

Noisy datasets

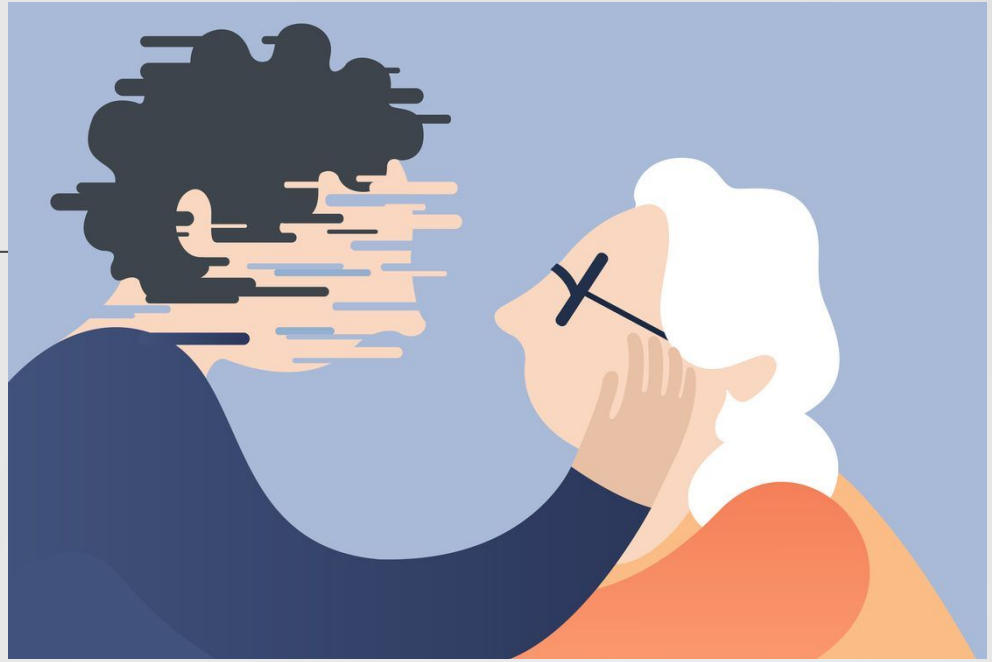
Neuroimaging data contains structured temporal "noise" from a variety of sources, including subject motion, subject physiology, and the necessary equipment.

02

Traditional Machine Learning Techniques in AD Studies

Alzheimer's Disease

Alzheimer’s disease is an irreversible, progressive brain disorder that slowly destroys memory and thinking skills and, eventually, the ability to carry out the simplest tasks. In most people with the disease—those with the late-onset type—symptoms first appear in their mid-60s. Alzheimer’s disease is, also, the most common cause of dementia among older adults.



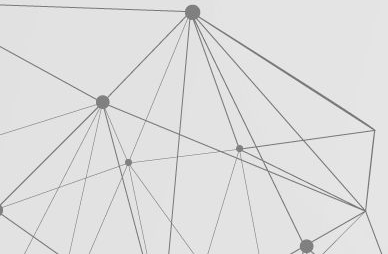
THE NUMBERS

5.8 M — Americans are currently living with AD

14 M — is what the above number is projected to rise to by 2050

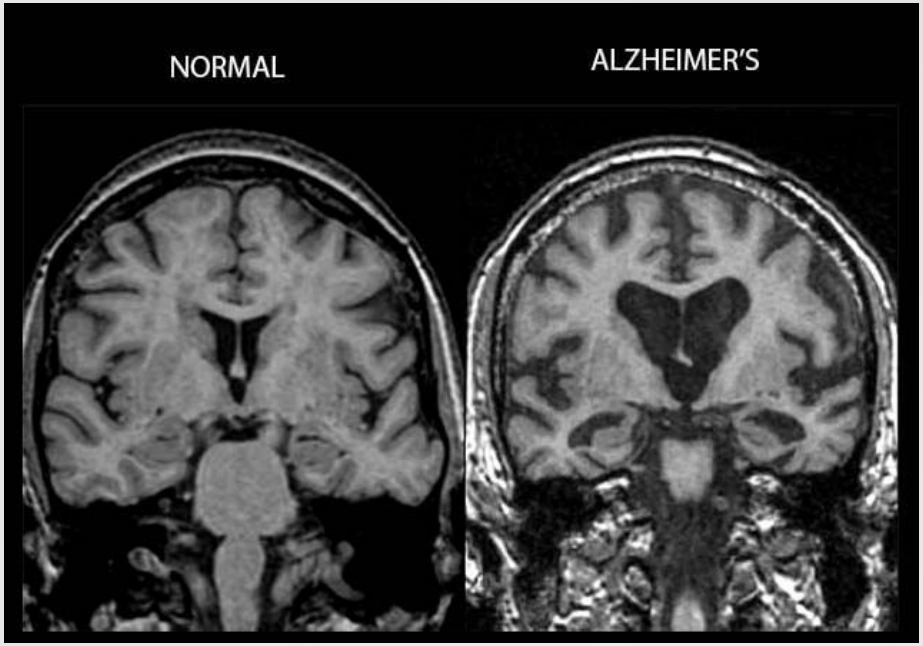
290 B — $ AD and other dementias currently cost

1.1 T — is what the above number is projected to rise to by 2050

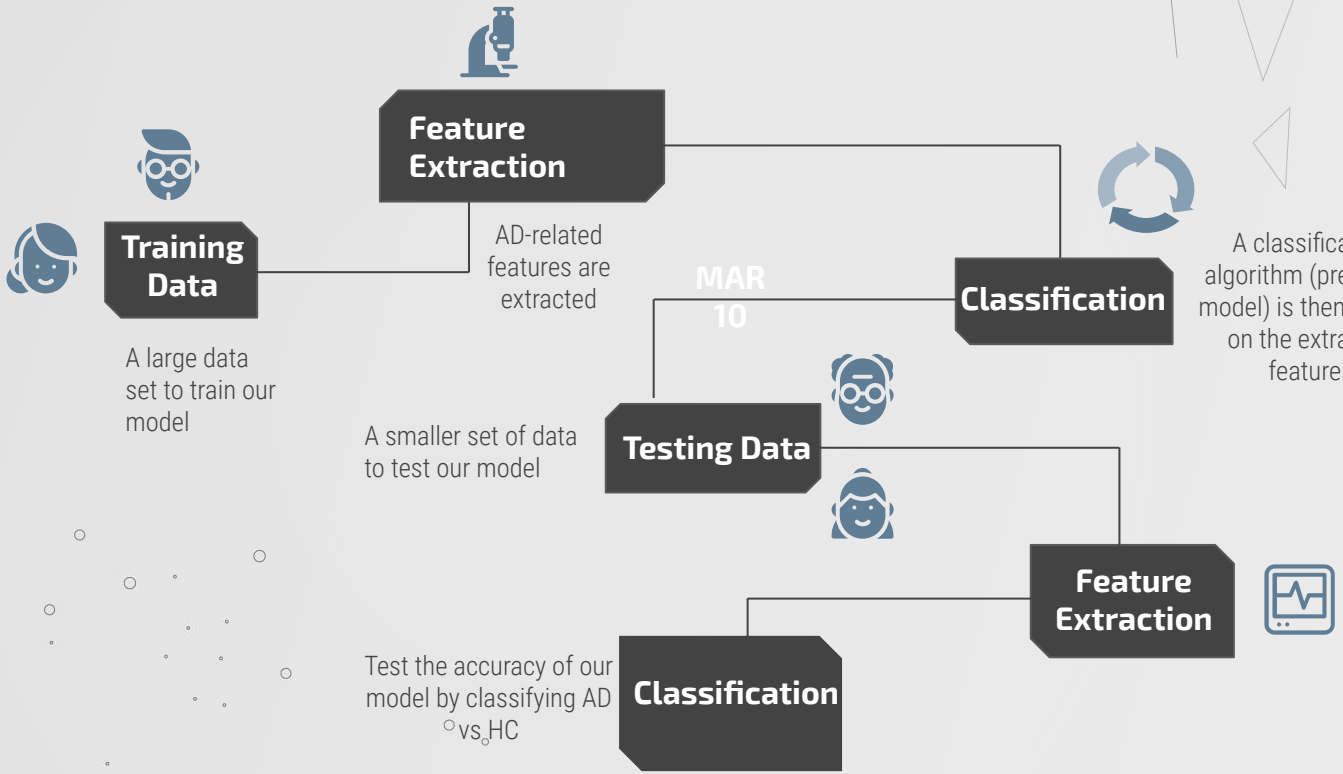


How we can use ML

- Brain changes due to AD occur even before amnestic symptoms appear and occur in a pattern that typically includes the temporal lobe and hippocampus. It has been suggested that this inevitable atrophy can be a valuable marker of neurodegeneration measured with neuroimaging modalities.
- Machine learning offers a systematic approach in developing sophisticated, automatic, and objective classification frameworks for analyzing high-dimensional data and can learn complex and subtle patterns of change across various imaging modalities, such as MRI.



Neuroimaging-based framework for AD classification



Well-known pattern analysis methods, such as:

- Linear discriminant analysis (LDA)
- Linear program boosting method (LPBM)
- Logistic regression (LR)
- Support vector machine (SVM)
- Support vector machine recursive feature elimination (SVM-RFE)

have been used and hold promise for early detection of AD and the prediction of AD progression.

“Traditional” ML algorithms

As we previously saw these machine learning algorithms generally require four steps: feature extraction, feature selection, dimensionality reduction, and feature-based classification algorithm selection, with the middle two being optional (but advisable to use or extra acuracy).

Caveats

Predefined steps

Appropriate architectural design or pre-processing steps must be predefined.

Feature extraction

Requires multiple stages of optimization, which may be computationally more demanding..

Accuracy

As we will see, other machine learning methods (e.g. Neural Networks) are achieving significantly better results.

Feature selection

Mean subcortical volumes, gray matter densities or cortical thickness may be used to derive more informative measures.

Reproducibility

All the pre-processing steps require specialized knowledge in the area of neuroimaging which limits the number of research groups to study AD.

Data Training

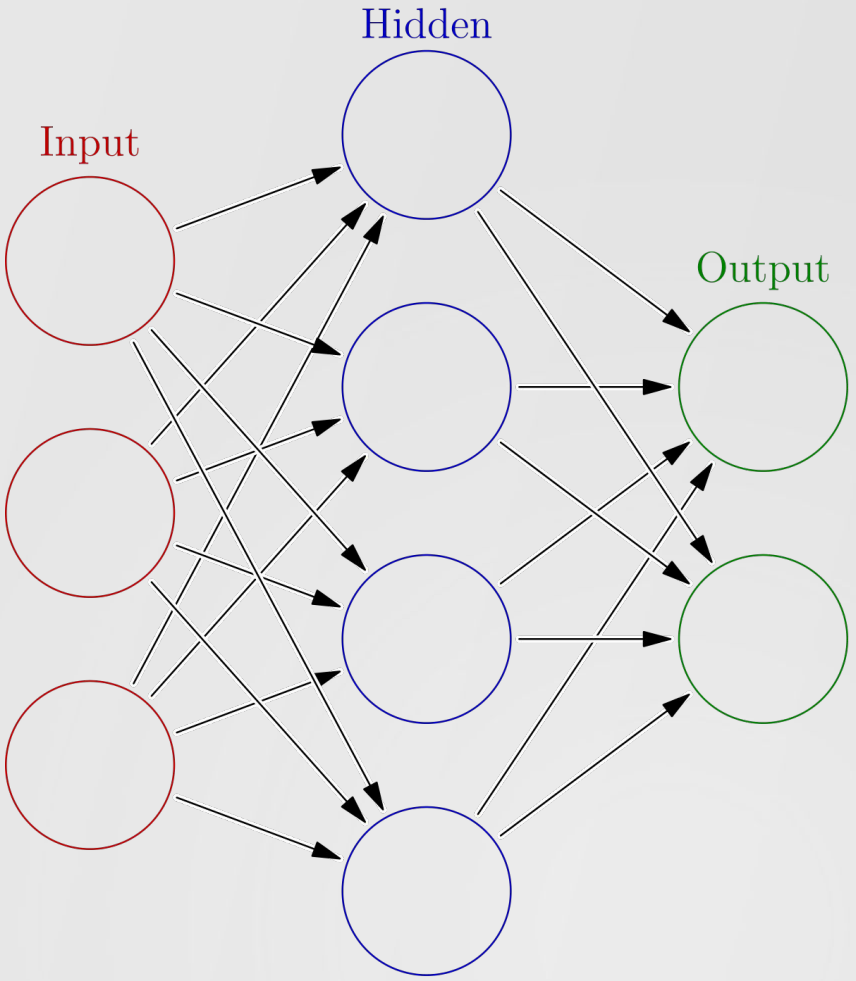
All of the above factors are added up to make the training of the algorithms tedious and time consuming.

03

Deep Learning in AD Studies

Artificial Neural Networks

Artificial neural networks (ANNs) are computing systems that are inspired by, but not identical to, biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules.



Useful concepts

Calculate error between network output and the expected output using gradient descent until it reaches 0.

Back Propagation

Decide whether a neuron should “fire” or not. Popular activation functions include: sigmoid, tanh, ReLU, softplus etc

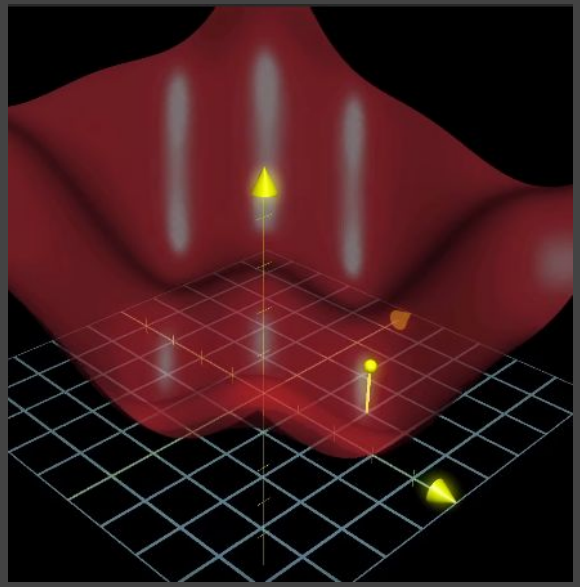
Activation Function

Refers to a model that models our “training” data too well. It learns every detail and noise in that dataset. That impacts the performance on new data.

Overfitting

Learning in DL is accomplished by training the model with new data until it finds the correct combination of weights.

Parameters

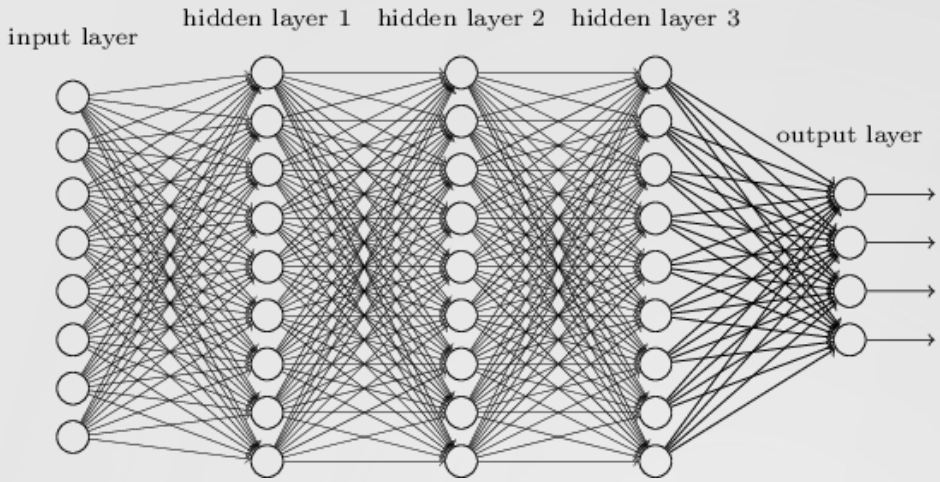


Gradient Descent

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model.

Deep Neural Networks

A deep neural network is a neural network with more than two layers. There's the input, the output and at least two hidden layers. There, also, different kinds of DNNs, such as Deep Boltzman Machines, Deep Belief Networks, Auto-Encoders, Sparse / Stacked Auto - Encoders, Restricted Boltzman Machine etc
By using DNNs we can classify the subjects to AD, HC, MCI and more classes..



DL in Neuroimaging Classification Studies

Feature Extraction

Train the model to extract features of the input (e.g. image). Earlier layers encode low level features such as edges, while later layers extract higher level features.

The last layer of neurons will output the probabilities for each neuron/class. The classification can be binary or have more classes.

Classification

Modalities

The input data that are fed to our model can be in either 3D or 2D format, such as MRI, PET, fMRI, DTI, or combinations of them.

... more useful concepts

While DNNs do feature extraction automatically, preprocessing our raw data can greatly affect the classification performance and avoid data leakage..

Preprocessing

Occurs when the dataset framework is designed incorrectly, resulting in a model that uses inessential additional information for classification.

Data Leakage

Generally the “how” of DNNs is unknown, as the complexity of the models make difficult to determine how the features are selected.

“Black Box”

By using more than one modality our model may be more complex but it will lead to better classification accuracy and results.

Multi-modality

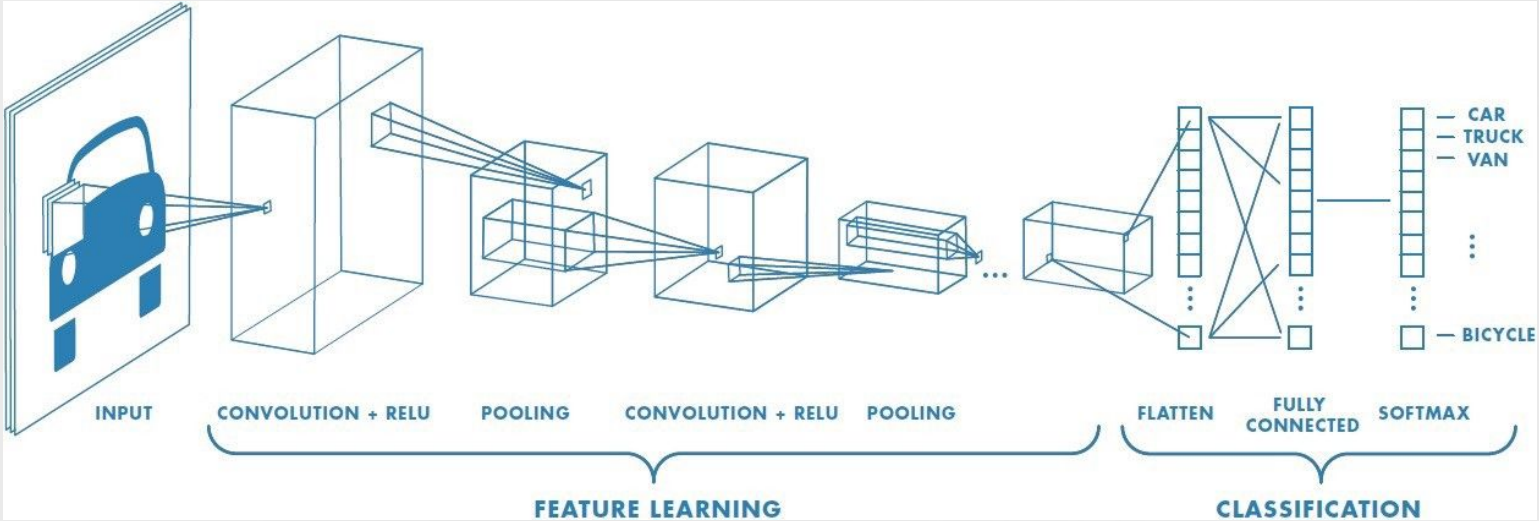
04

Convolutional Neural Networks in AD studies

Convolutional Neural Networks

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.



... even more useful concepts

Rectified Linear Units are used to filter every unit of an array so as to make every negative value equal to zero.

ReLU

The objective of the convolution layer is to extract the high-level features such as edges, from the input image, by computing image*filter.

Convolutional Layer

This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features.

Pooling Layer

The flattened output is fed to a feed-forward neural network and backpropagation is applied to every iteration of training.

Fully-Connected Layer

THANKS

Does anyone have any questions?

iliaspan19@gmail.com
el15746@central.ntua.gr