

---

# BioSim Thesis

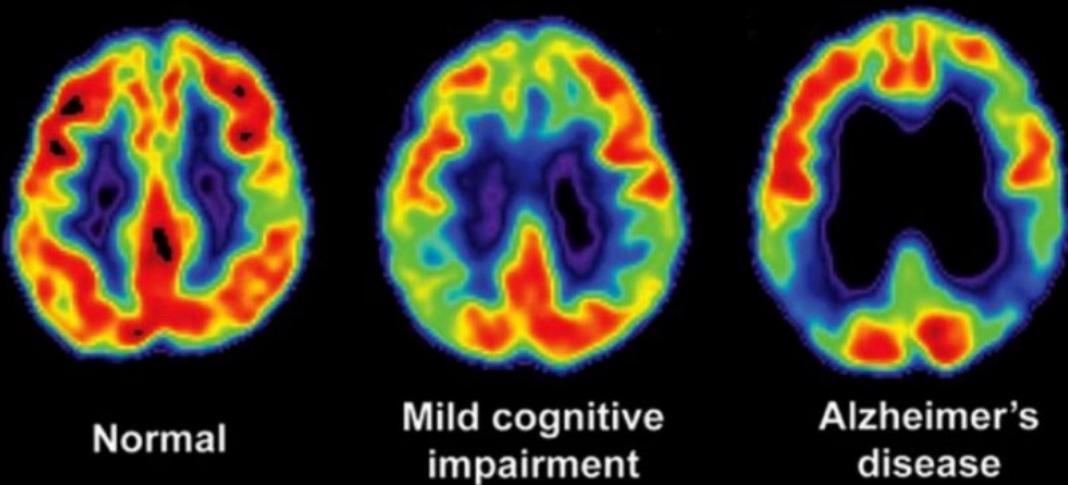
---

Nicolas Pigadas

# The Problem

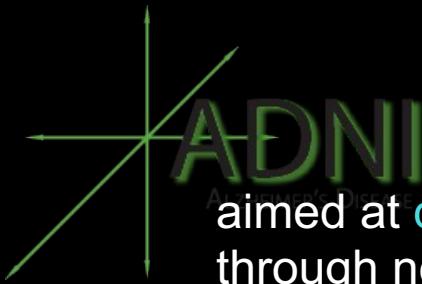
Classification of an elderly group of participants into 3 categories:

- 1) Healthy Individuals (**NC**)
- 2) Individuals with Mild Cognitive Impairment (**MCI**)
- 3) Alzheimer's Disease patients (**AD**)



[2] Photo: <https://www.alzsd.org/resources/mild-cognitive-impairment/>

# ADNI



(Alzheimer's Disease Neuroimaging Initiative) is a **large-scale**, multi-site study aimed at **developing biomarkers** for early detection and tracking of **Alzheimer's disease** through neuroimaging, genetic, cognitive, and biochemical assessments.

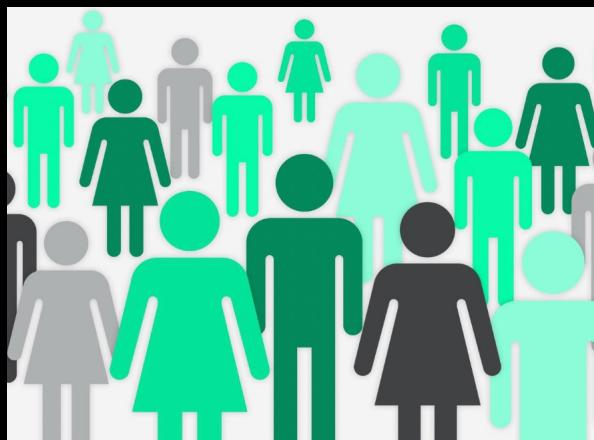
The three overarching **goals** of the ADNI study are:

- To **detect AD at the earliest possible stage** (predementia) and identify ways to track the disease's **progression** with biomarkers.
- To support advances in **AD intervention, prevention, and treatment** through the application of new diagnostic methods at the earliest possible stages (when intervention may be most effective).
- To continually administer ADNI's innovative **data access** policy, which provides all data without embargo to all scientists in the world.

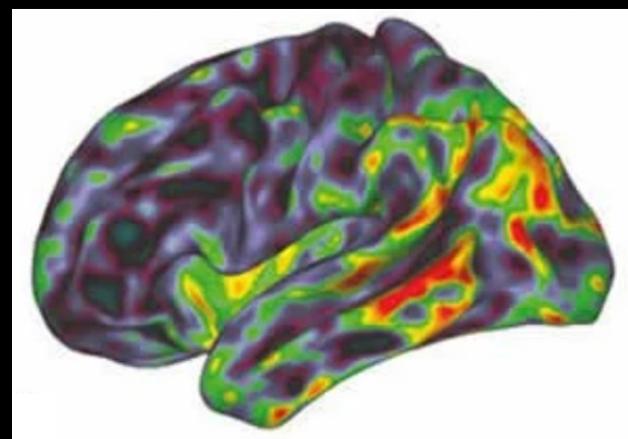
# The Data

The acquired ADNI data are **tabular** and of three categories:

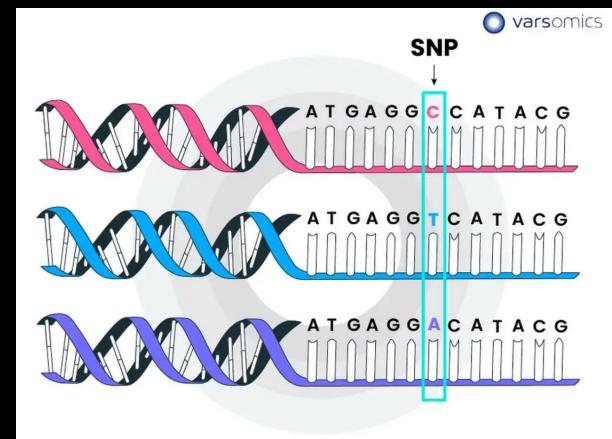
- 1) Demographic Data<sub>binary</sub> (Age, Sex)
- 2) VOIs/Brain Volume ROIs<sub>floats</sub> (Regions Of Interest)
- 3) SNPs<sub>binary</sub> (Single-Nucleotide Polymorphisms: variations in a single nucleotide that occur at a specific position in the genome, distinguishing one individual's DNA from another's.)



1) Demographic Data [3]



2) VOIs [4]



3) SNPs [5]

[3] Photo 1: <https://within3.com/blog/demographic-data-in-clinical-trials>

[4] Photo 2: <https://neurosciencenews.com/musical-memory-alzheimers-2144/>

[5] Photo 3: <https://blog.varsomics.com/en/snp-what-are-single-nucleotide-polymorphisms/>

# The Dataset

The dataset contains 1463 individuals; 449 in the control group (**CN**, 30.69%), 740 affected by mild cognitive impairment (**MCI**, 50.58%) and 274 diagnosed with Dementia (**DEM**, 18.73%). The age of the participants range from 60 to 86 years of age.

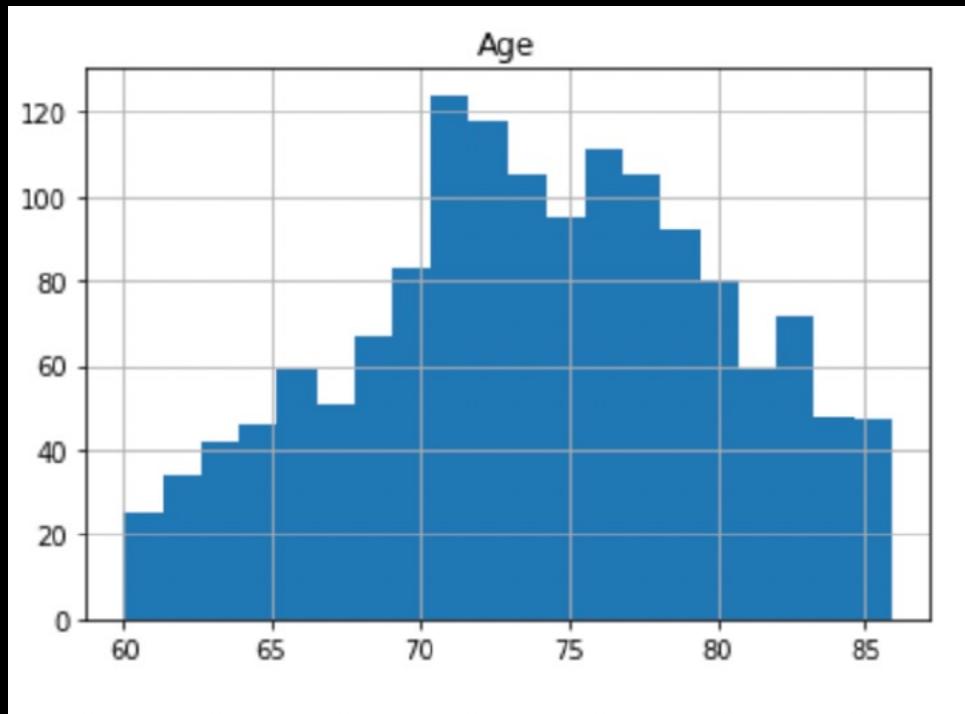


Fig. 1: Histogram: Age of participants.

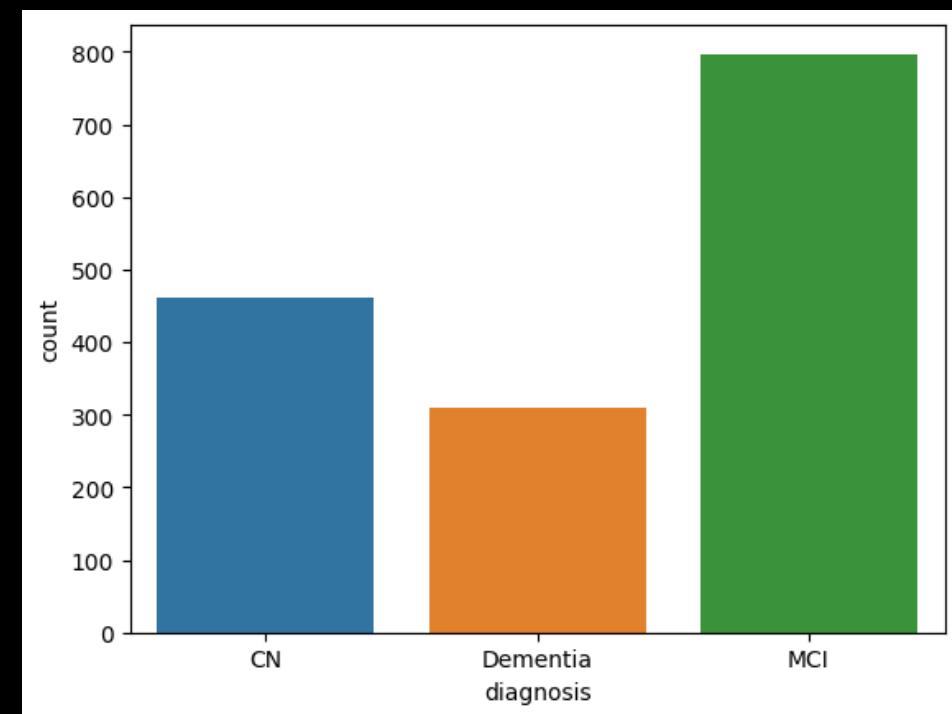


Fig. 2: Dataset's balance.

# Data Pipeline

## Step 1: Data Acquisition

1.5 T and 3T weighted MRI data were acquired from ADNI.

## Step 2: Data Preprocessing

A fully automated pipeline was applied for processing T1 structural MRIs [6]:

- a) T1-weighted scan of each participant is first corrected for intensity inhomogeneities [7].
- b) A multi-atlas skull stripping algorithm was applied for the removal of extra-cranial material [8].
- c) 145 anatomical regions of interest (ROIs) were identified using a multi-atlas label fusion method [9].
- d) Phase-level cross-sectional harmonization was applied on regional volumes of 145 ROIs to remove site effects [10].

## Step 3: Data Preparation

- a) Linear Correction (age, sex) [6]

- b) Data variance-normalization [6]

[6] <https://arxiv.org/pdf/2102.12582.pdf>

[7] <https://pubmed.ncbi.nlm.nih.gov/9617910/>

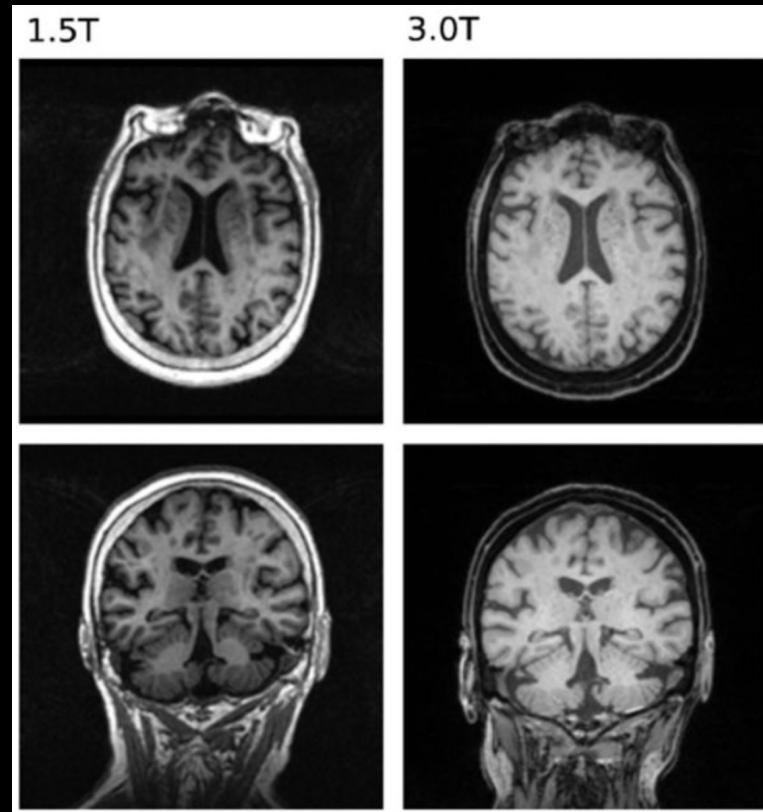
[8] <https://www.sciencedirect.com/science/article/abs/pii/S1076633213004182>

[9] <https://pubmed.ncbi.nlm.nih.gov/26679328/>

[10] <https://pubmed.ncbi.nlm.nih.gov/31821869/>

# Data Acquisition<sub>1</sub>: Raw Data

Step 1: 1.5 T and 3T MRI data were acquired from ADNI.

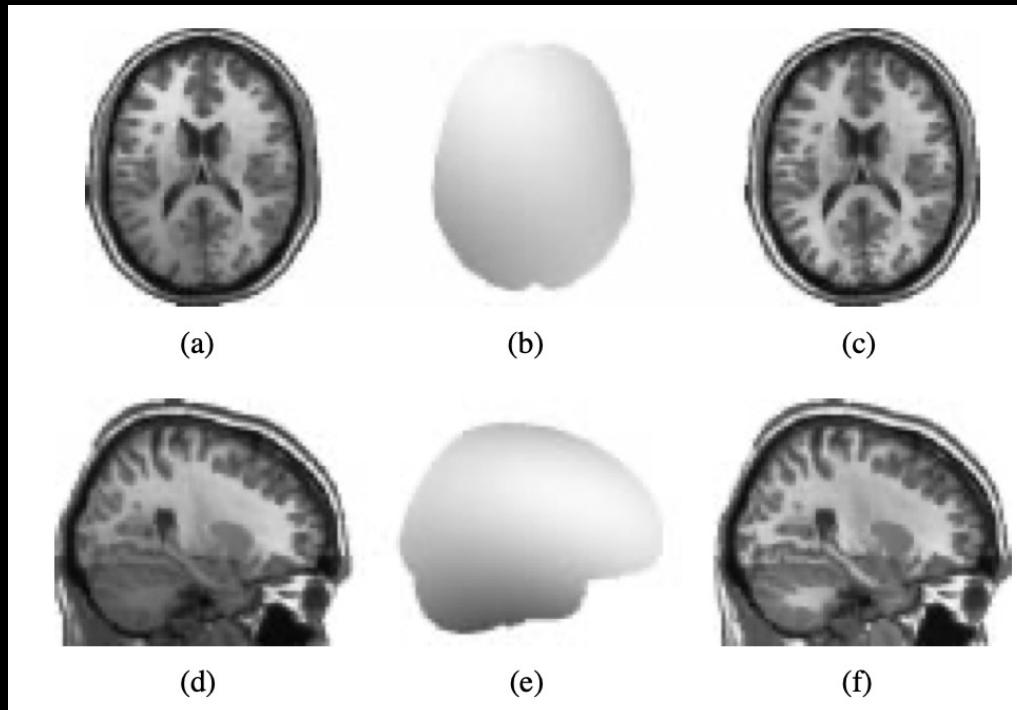


Example of 1.5T and 3T weighted MRI data [6].

# Data Preprocessing<sub>2a</sub>: Intensity Correction

Step 2: A fully automated pipeline was applied for processing T1 structural MRIs:

- a) T1-weighted scan of each participant is first corrected for intensity inhomogeneities [7].

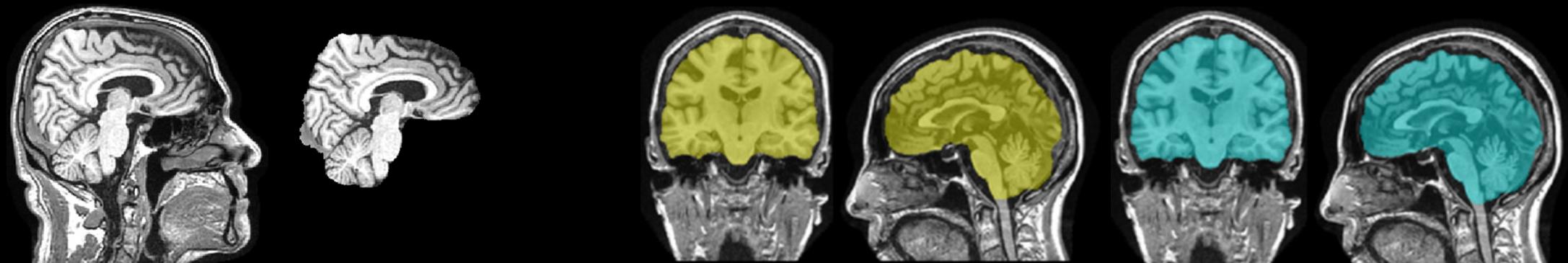


Intensity nonuniformity correction of a T1 weighted 27-scan averaged gradient-echo MR scan: (a) and (d) transaxial and sagittal views of uncorrected data; (b) and (e) nonuniformity field estimated by the N3 method; (c) and (f) corrected data. [7]

# Data Preprocessing<sub>2b</sub>: Skull Stripping

Step 2: A fully automated pipeline was applied for processing T1 structural MRIs:

- b) A multi-atlas skull stripping algorithm was applied for the removal of extra-cranial material [8].



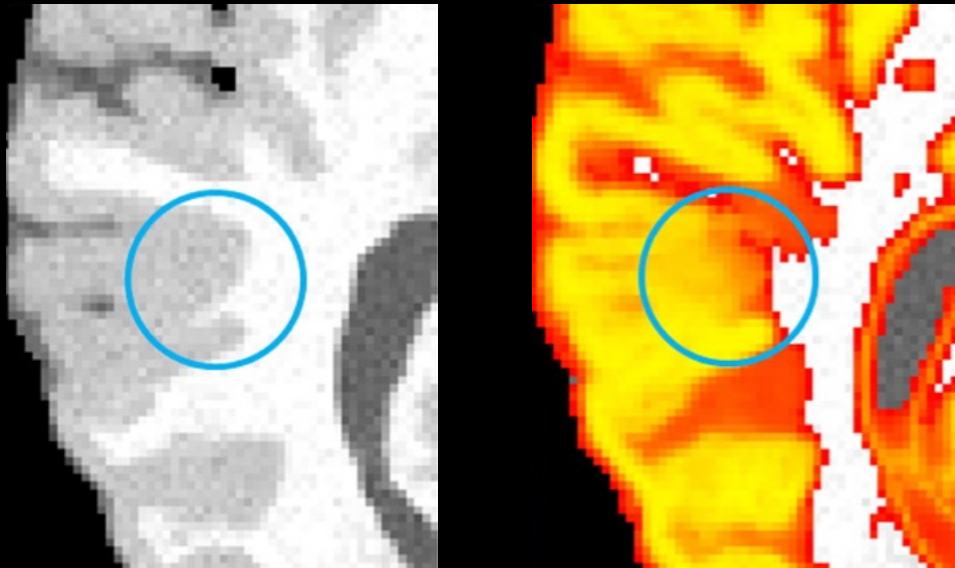
Example of skull stripping and brain extraction [8].

Final brain masks generated by Multi-Atlas Skull-Stripping (MASS) on ADNI data [8].  
Left: The ground truth brain masks are overlaid (in yellow) on the coronal and sagittal views of the brain.  
Right: Final MASS masks are overlaid (in blue) on the same views.

# Data Preprocessing<sub>2c</sub>: ROIs

Step 2: A fully automated pipeline was applied for processing T1 structural MRIs:

c) 145 anatomical regions of interest (ROIs) were identified using a multi-atlas label fusion method [9].



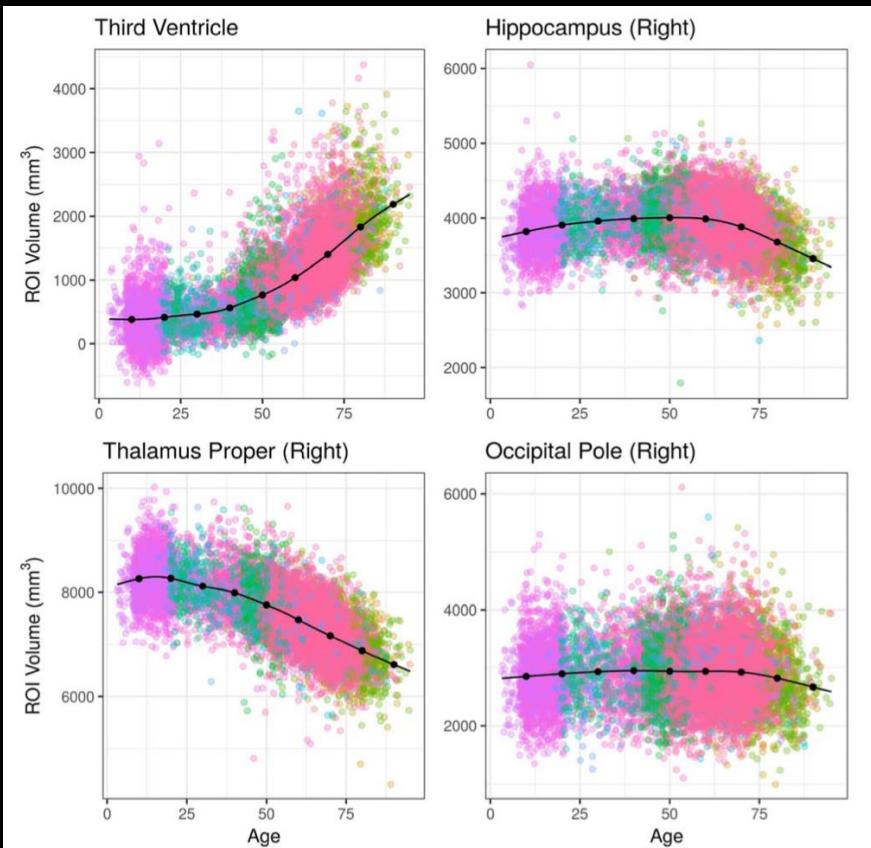
On the left: Original Image and a targeted ROI.

On the right: Probability map of wraps and the proposed label-fusion method [9].

# Data Preprocessing<sub>2d</sub>: Non-Linear Harmonization

Step 2: A fully automated pipeline was applied for processing T1 structural MRIs:

- d) Phase-level cross-sectional harmonization was applied on regional volumes of 145 ROIs to remove site effects [10]. In order to capture non-linearities in age-related volume differences in brain anatomy throughout the individuals' lifespan, a generalized additive model (GAM) with a penalized non-linear term to describe age effects was used. In conjunction with GAMs, ComBat was used in the harmonization pipeline, forming the proposed ComBat-GAM pipeline [10].

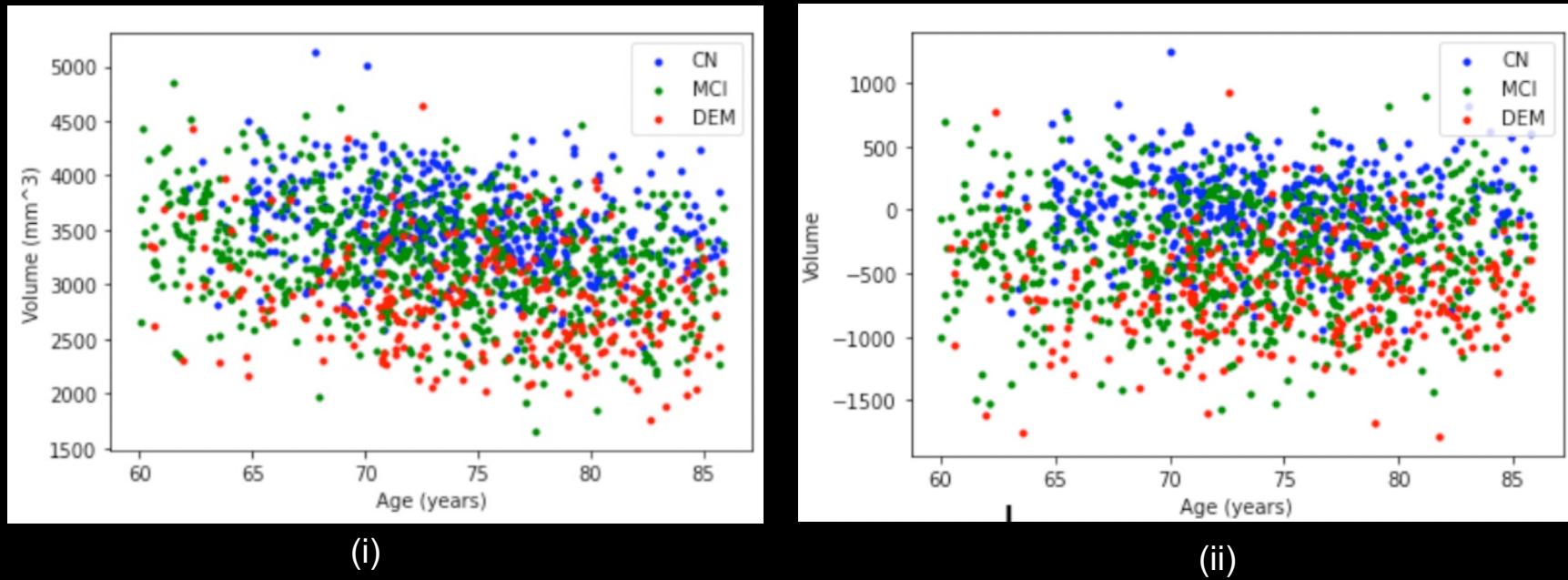


Age trends for selected ROI volumes using a different combined dataset (LIFESPAN) with 18 studies spanning the age range 3 – 96. Data were harmonized using ComBat-GAM. The age trends plotted are for females and assume an average intra-cranial volume (ICV) [10].

# Data Preparation<sub>3a</sub>: Linear Correction (age, sex)

## Step 3: Data Preparation

- a) **Linear Correction [6]:** To correct age and sex effects while keeping disease-associated neuroanatomical variations, we estimated ROIs-specific age and sex associations among the CN participants using a linear regression model. All cross-sectional data were then residualized by age and sex effects, removing the linear trends of age and sex.

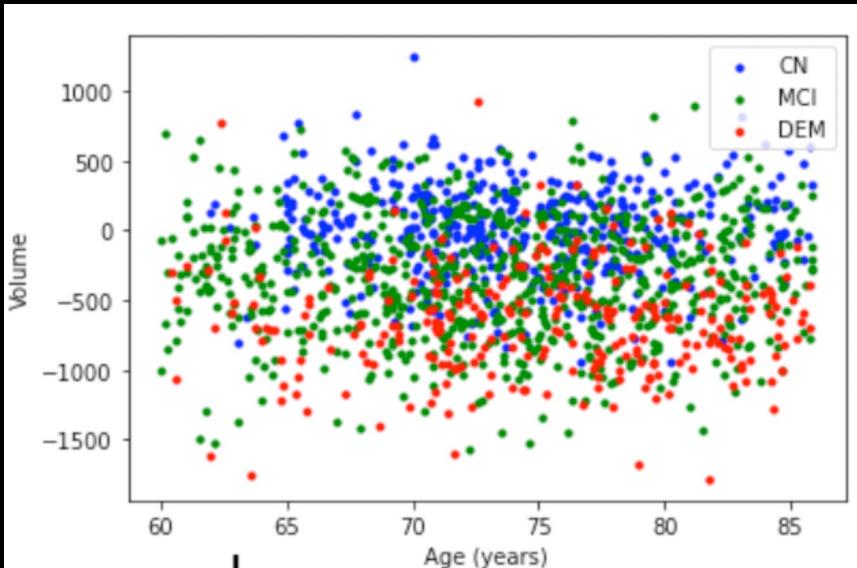


Data before (i) and after (ii) linear trend correction based on age and sex.

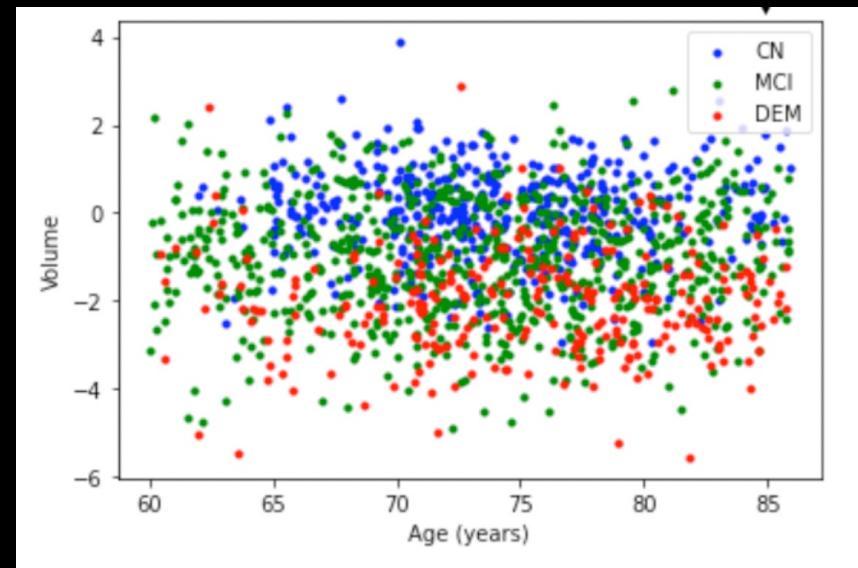
# Data Preparation<sub>3b</sub>: Variance Normalization

## Step 3: Data Preparation

- a) **Linear Correction [6]:** To correct age and sex effects while keeping disease-associated neuroanatomical variations, we estimated ROIs-specific age and sex associations among the CN participants using a linear regression model. All cross-sectional data were then residualized by age and sex effects, removing the linear trends of age and sex.



(ii)



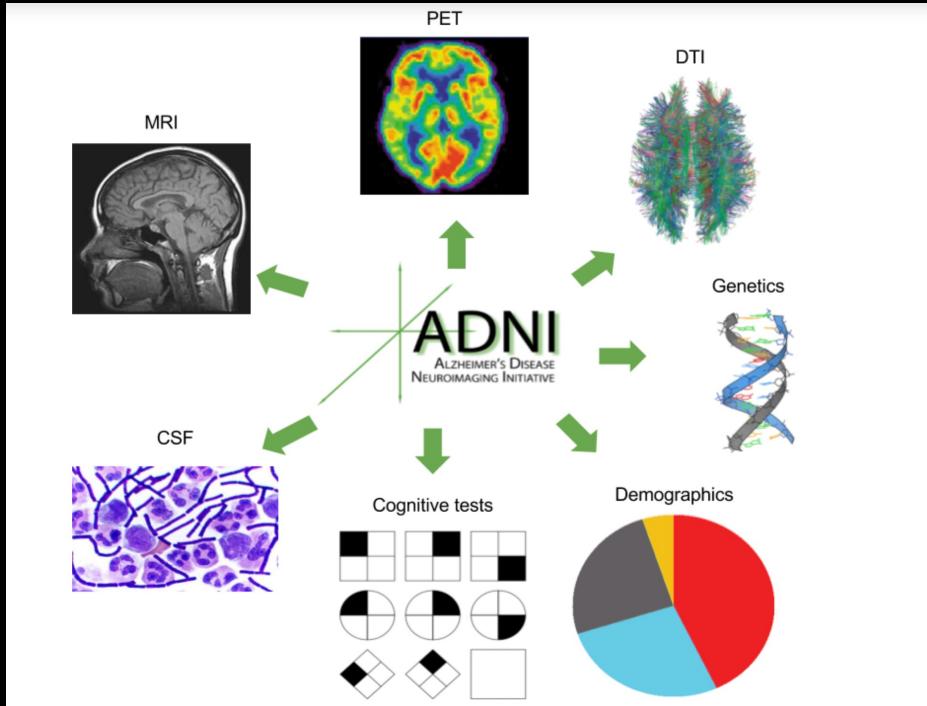
(iii)

Data before (ii) and after (iii) variance (z-score) normalisation after linear trend correction based on age and sex. The distribution of points remains unchanged as expected.

$$z\_score_i = \frac{x_i - \mu_{CN}}{\sigma_{CN}}$$

# Tadpole

There is a subset of ADNI data called Tadpole. It was formed for a grand challenge. Many researchers study this dataset instead of ADNI and it is not considered to be the same as ADNI in the ADNI website. Yet, many research ideas of mine are based on a paper on the Tadpole dataset (Differentiable Graph Module (DGM) for Graph Convolutional Networks).



[18] <https://tadpole.grand-challenge.org/Data/#Data>

# ADNI Analysis

# ADNI Publication Analysis | (as of Jan. 2023)

source: <https://adni.loni.usc.edu/news-publications/publications/?q=>

No of Publications: 3740

Keyword (Suggestion: use one word at a time, use of exact phrase filter)	No of Results	Percentage %
graph	951	25,43
graphs	390	10,43
Davatzikos	796	21,28
gnns	5	0,13
gcn	24	0,64
gat	9 (only cited), real: 0	0,24
mri	3297	88,16
fmri	871	23,29
machine	1491	39,87
machine learning	1282	34,28
deep	1030	27,54
deep learning	665	17,78
learning	2291	61,26
pet	2707	72,38
csf	2288	61,18
tau	1930	51,60
amyloid	2457	65,70
neurofilament	130	3,48
neurogranin	56	1,50
APOE	1977	52,86
SPECT	259	6,93
neuroimaging	3692	98,72
imaging	3665	97,99
biomarkers	2921	78,10
predict	2204	58,93
progression	3172	84,81
treatment	1801	48,16
early	3461	92,54
personalized	18	0,48
drug	2910	77,81
guided	413	11,04
geometric	446	11,92
assume (to find assumptions-weaknesses in literature)	774	20,7
missing	1135	30,34

# ADNI Publication Analysis II (as of Jan. 2023)

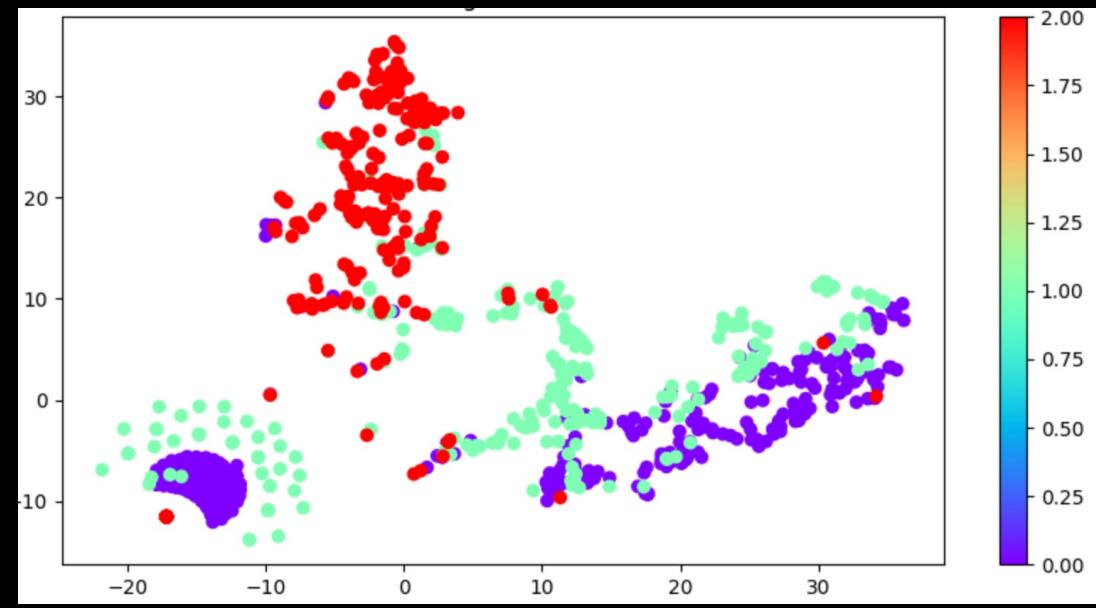
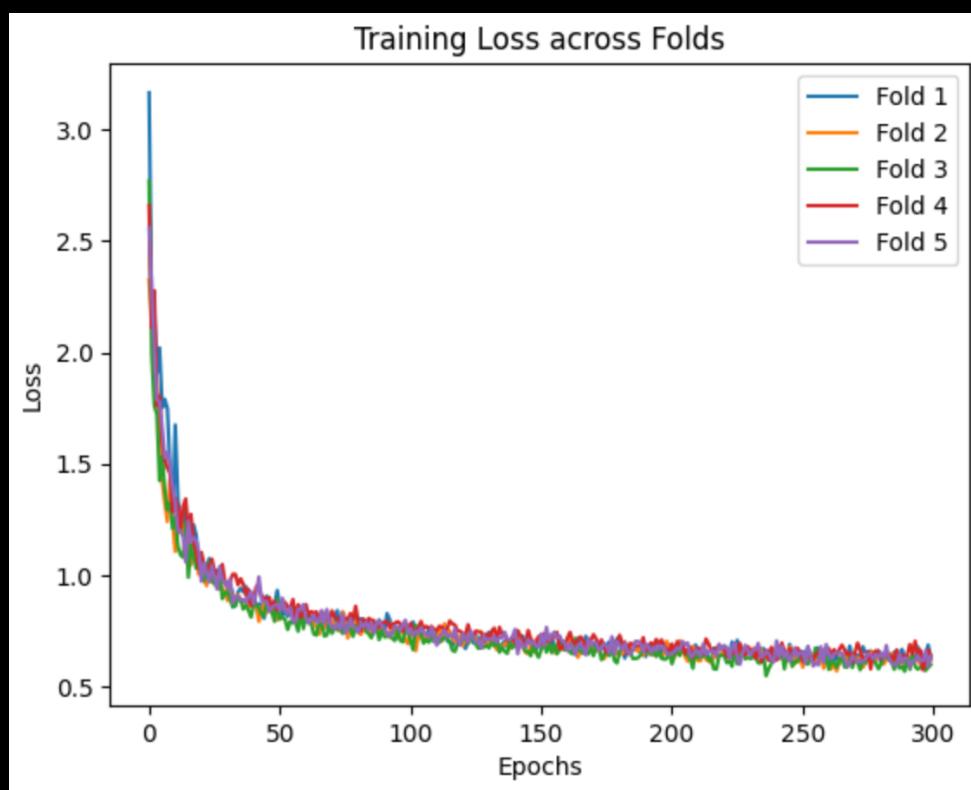
source: <https://adni.loni.usc.edu/news-publications/publications/?q=>

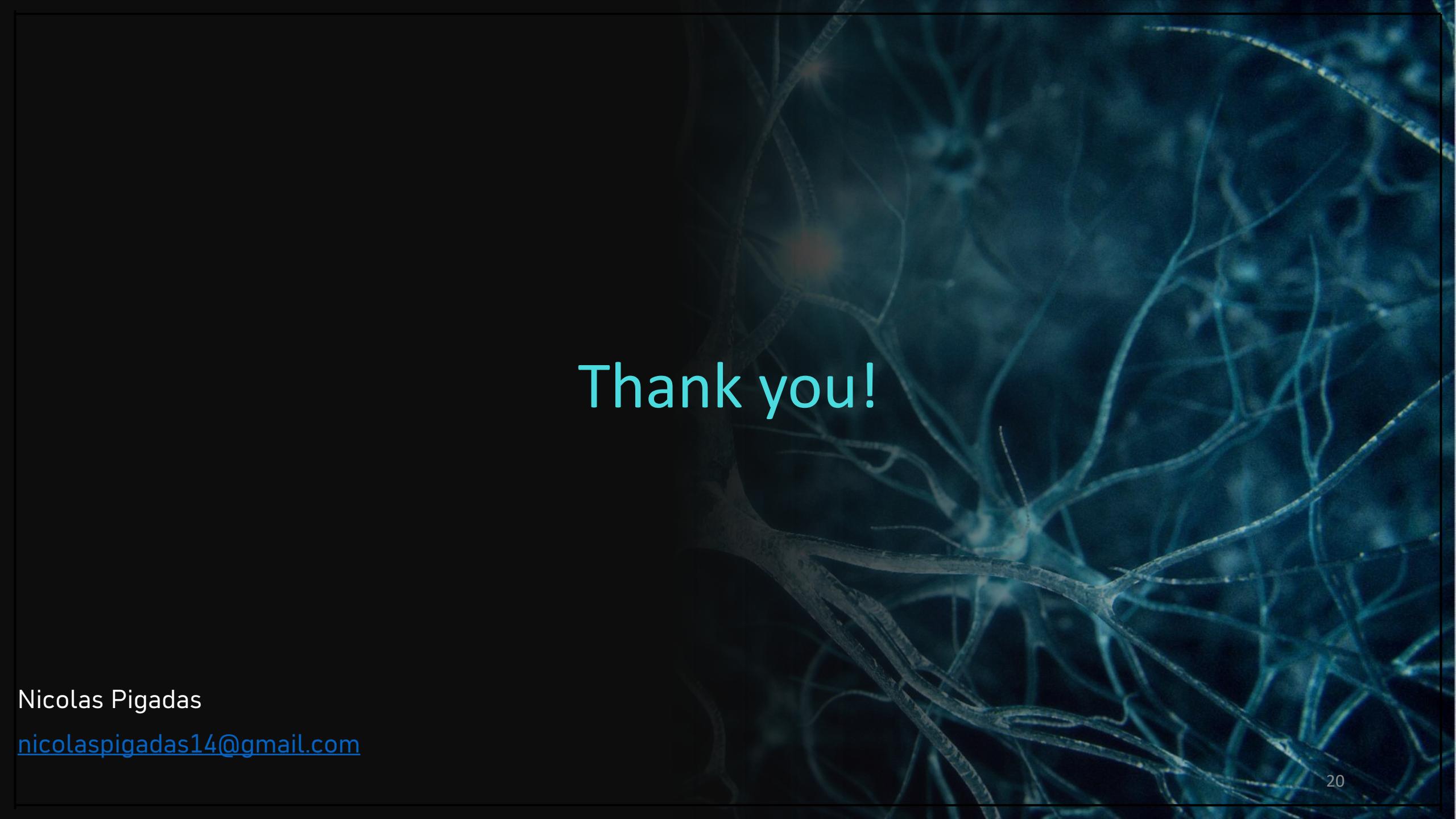
No of Publications: 3740

Keywords, Models (Note: use of exact phrase filter)	No of Results	Percentage %
gnns	5 (results), real: 18 (incl. gcns, gans)	0,13, 0,48
graph neural networks	19 (results), real: 18 (incl. gcns, gans)	0,51, 0,48
fusion	739	19,76
graph fusion	40	1,07
deep fusion	18	0,48
feature fusion	77	2,06
random walk	32	0,86
deepwalk	5	0,13
graph kernel	37	0,99
graph-CNN	7	0,19
graph-RNN	0	0
graphCNN	0 (2 refs.)	0
graphRNN	0	0
graph CNN	7	0,19
graph RNN	0	0
rnnns	33	0,88
recurrent neural networks	95	2,54
cnnns	218	5,82
convolutional neural networks	379	10,13
gcn	24 (results), real: 14 +in 1 review	0,64, 0,37
graph convolutional networks	32 (results), real: 14 +in 1 review	0,86, 0,37
gat	9 (only cited), real: 0	0
graph attention networks	7 (only cited), real: 0	0
gan	160 (only cited), real: a lot	4,28, ~4,5
generative adversarial network	50 (only cited), real: a lot	1,24, ~4,5
ResNet	114	3,05
graph ML	0	0
graph machine learning	0	0
geometric	446	11,92

# Initial Experiments

# Initial Experiments: GAT





Thank you!

Nicolas Pigadas

[nicolaspigadas14@gmail.com](mailto:nicolaspigadas14@gmail.com)