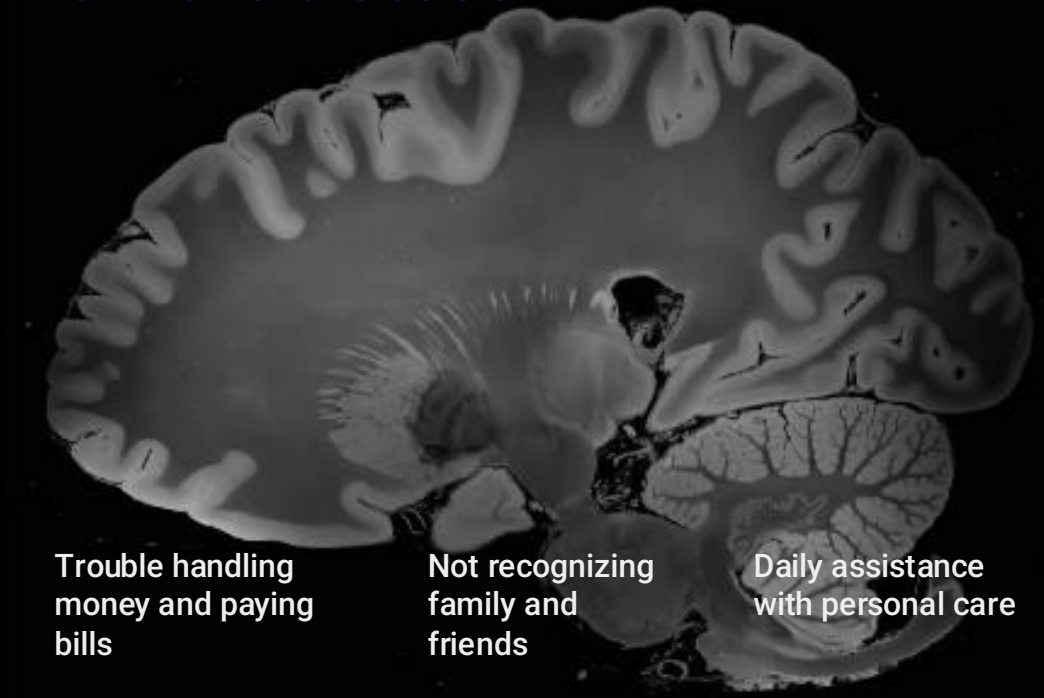
Multimodal Machine Learning algorithm as a tool for Alzheimer's disease clinical trial patient screening

William Jin, Senior Student, Plainsboro High School North,
Graham Zhang, Student, Proof School
Polina Vyniavska, M.Sc. Student, National University of Kyiv-Mohyla Academy,
Jeffrey Zhang, PhD. Principal Biostatistician, Princeton Pharmatech LLC,

What is Alzheimer's disease?



Trouble expressing thoughts

Trouble handling money and paying bills

Not recognizing family and friends

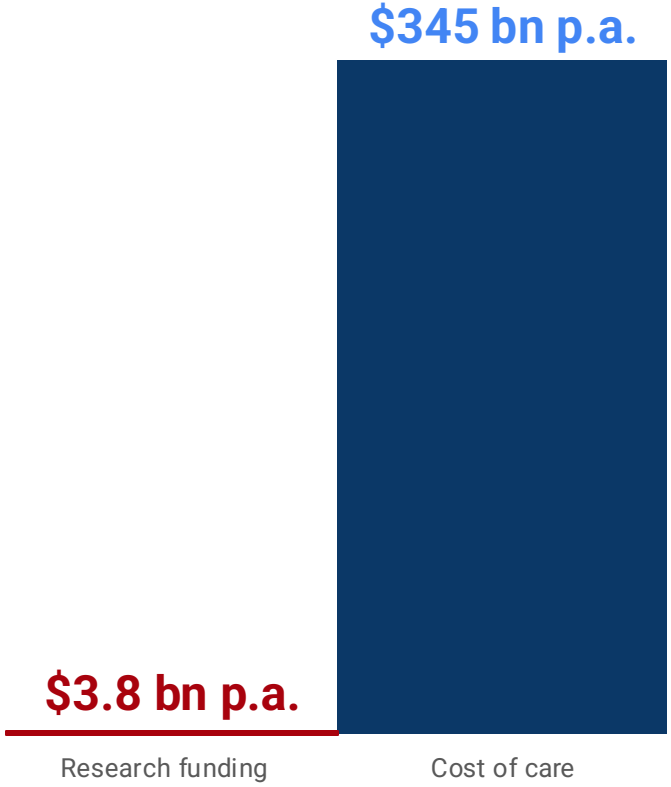
Daily assistance with personal care

Body shuts down

Healthy → Cognitively impaired → Alzheimer's Disease

Economical burden

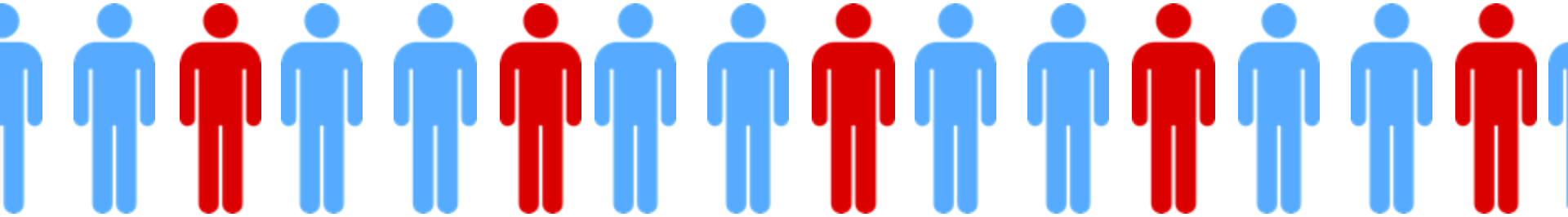
Alzheimer's annual cost of care in 2023 is more than 90 times the cost of Alzheimer's research funding.



Source: Alzheimer's Impact Movement

Future prediction

By 2060, 1 in 3 individuals aged 85 and older will have Alzheimer's dementia.



Source: Alzheimer's Association

PROBLEM

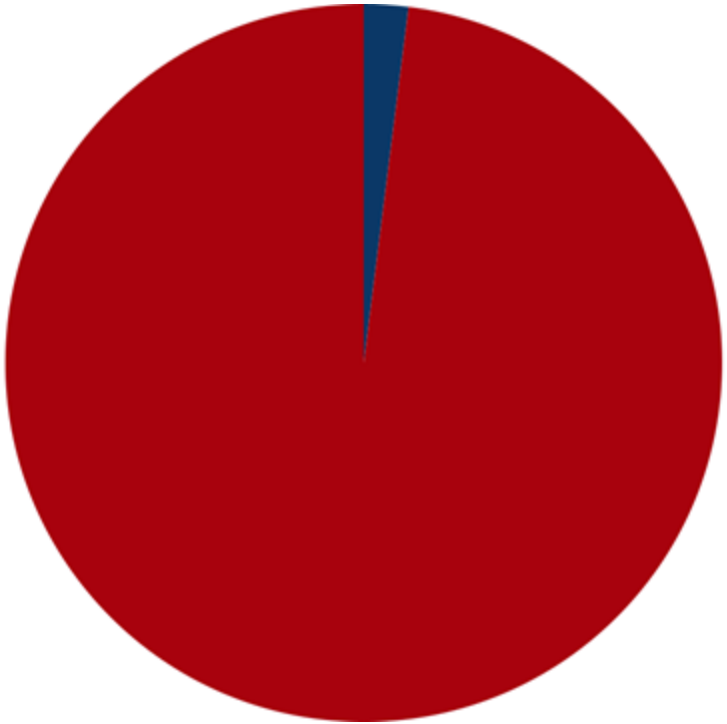
Why is there no cure for Alzheimer’s disease?

PROBLEM

Alzheimer’s clinical trials failure

98%

FAILED



2%

SUCCEEDED

Source: Kim et al., 2022

PROBLEM

A key factor of misdiagnosis

Absence of diagnostic tools

PRACTICE

Physical symptoms

Cognitive tests

POTENTIAL

Genetic test

Lab tests

Medical history

Brain imaging

Cerebrospinal fluid test

Family history

Neuropsychological exam

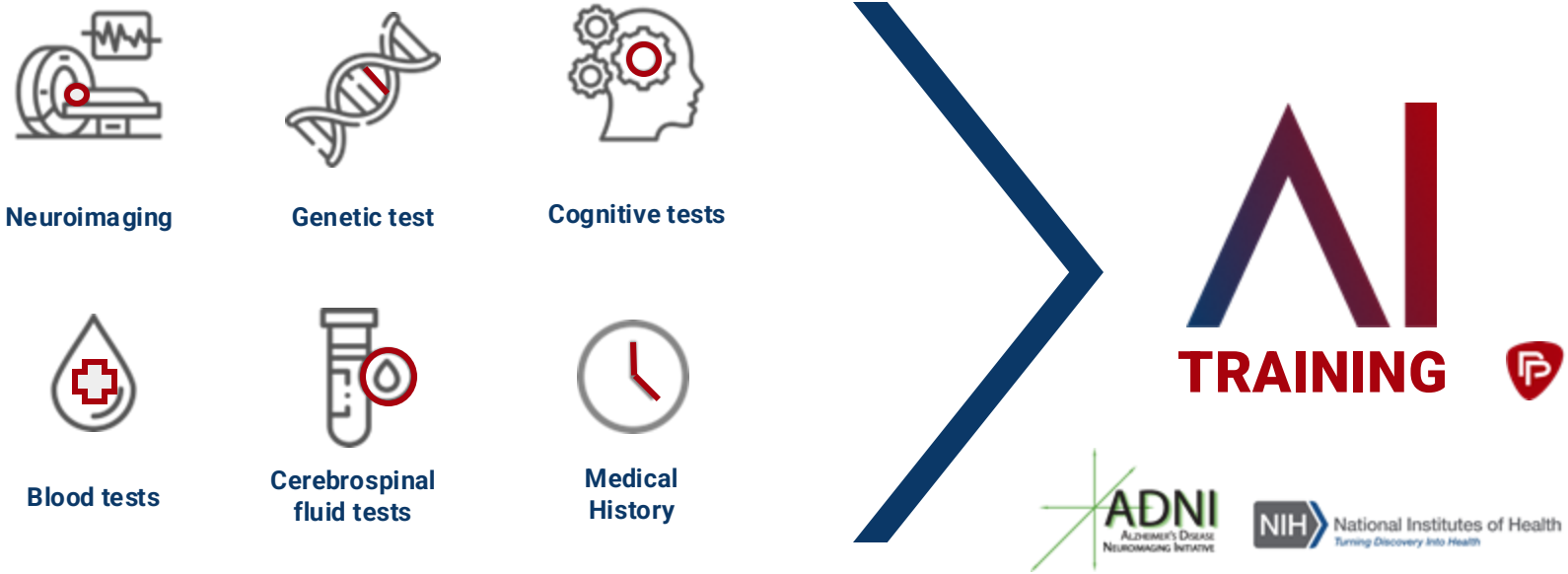
Study partner

RESEARCH PURPOSE

Embrace the rise of AI

Applying machine learning to identify common patterns.

Source of data for training AI model



The Alzheimer's Disease Neuroimaging Initiative (ADNI) provides researchers around the world with open source validated diverse data collected from cognitively healthy, mild cognitive impaired, and Alzheimer's disease patients throughout several years.

Inclusion of ADNI patients in a training sample

	Cognitively Normal (CN)	Mild Cognitive Impairment (MCI)	Alzheimer's Dementia (AD)
No. of subjects	913	1120	423
Males/females	388/507	645/458	233/180
Age range	50-90	54-91	55-90
Total No. of subjects	**2456**		

Inclusion criteria:

- Follow up for at least 2 years
- Not missing values for all 11 features

	CN	MCI	AD
No. of subjects	501	633	166
Total No. of subjects	**1300**		

Misclassification assessment

Rules to identify misdiagnosis:

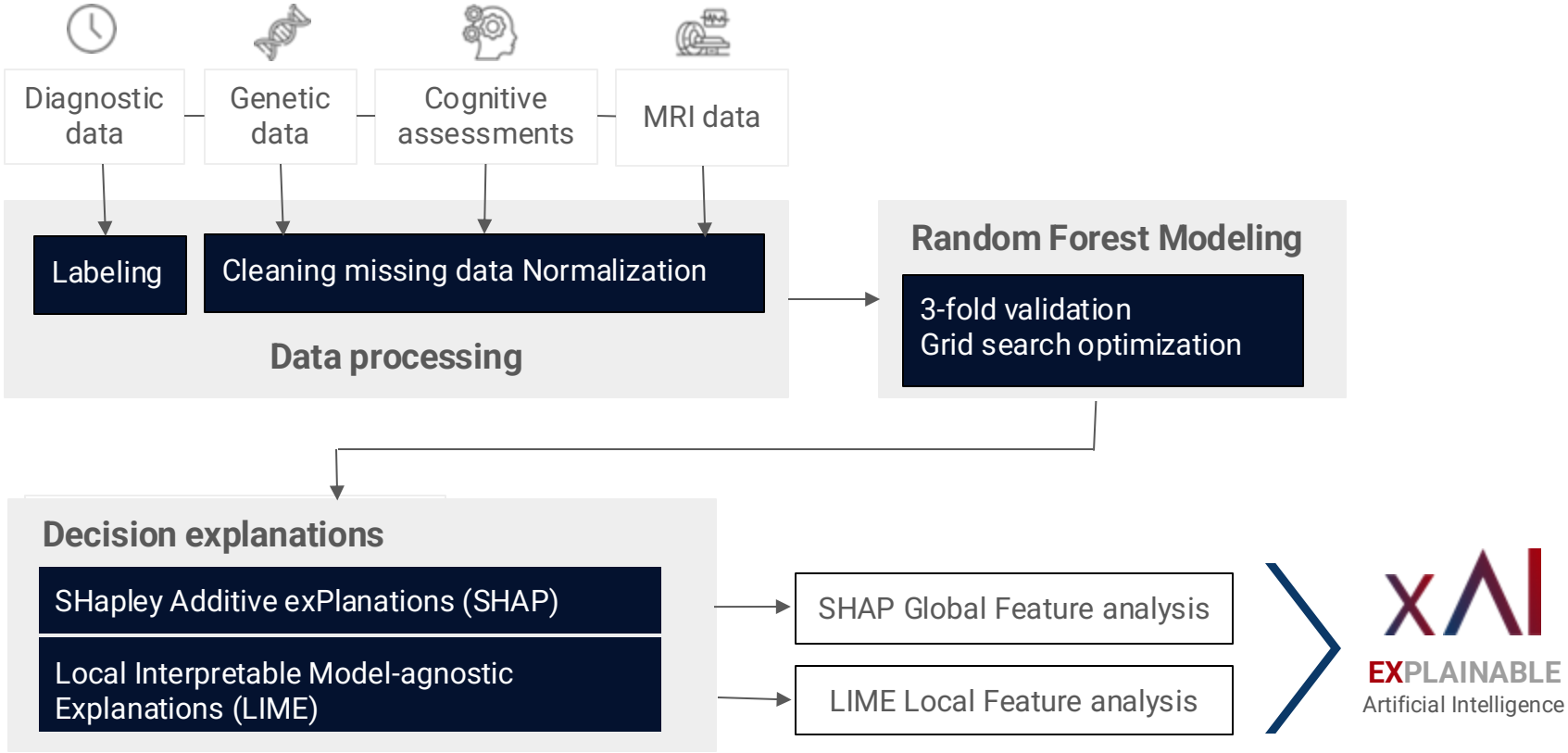
- Patients should follow up for at least 2 years
- Diagnosis was converted within the study
- During the latest visit MCI or AD patients' cause is due to other etiology
- Presence of other symptoms

12% (AD) and 7.3% (MCI) misdiagnosis

	CN	MCI	AD	non-AD	non-AD-MCI
No. of subjects	501	587	146	20	46
Males/females	215/286	354/233	78/68	7/13	27/19
Age range	55-90	55-91	55-90	59-84	56-90
Total No. of subjects	**1300**				

CN - Cognitively Normal; **MCI** - Mild Cognitive Impairment; **AD** - Alzheimer's Disease; **non-AD** - dementia due to other etiology; **non-AD-MCI** - MCI due to other etiology

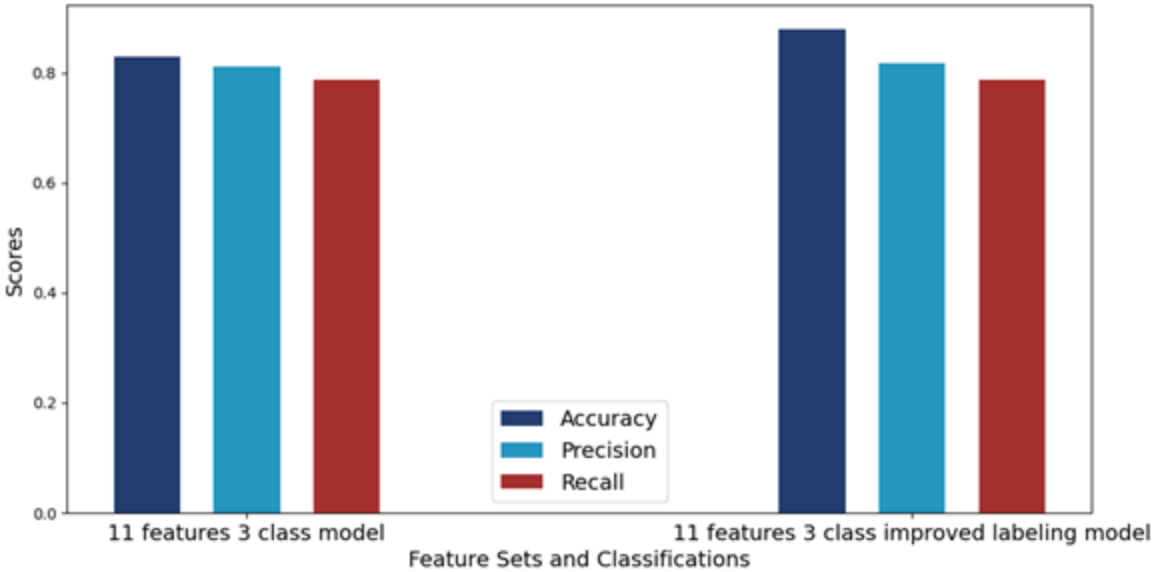
Building the explainable predictive model



RESULTS

Comparison of models' performances:

Original Physician Diagnosed VS. Misclassification Removed Labels



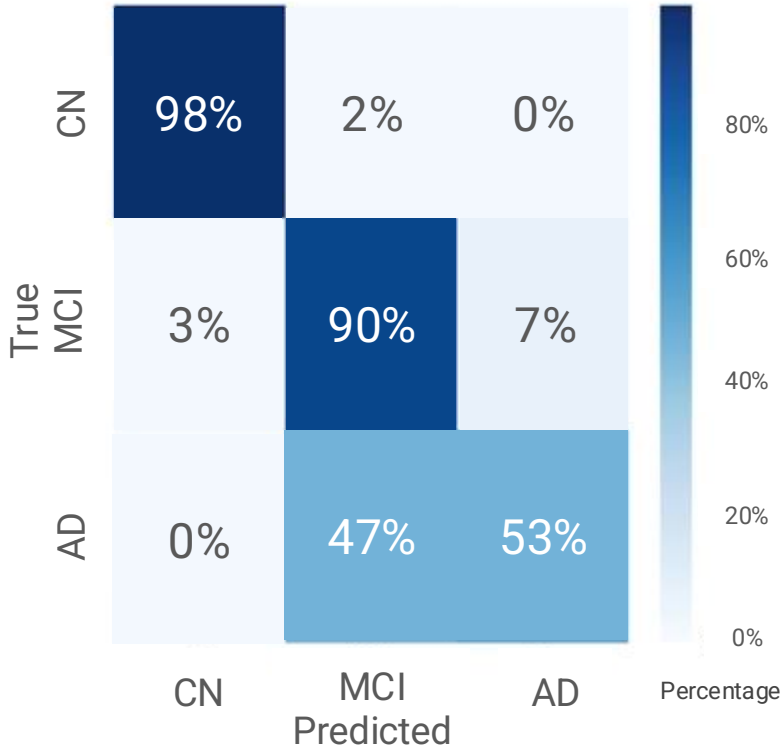
4 modalities with 11 features:

- **Demographics**: gender, ethnicity
- **Genetics**: *APOE4* gene,
- **Clinical assessments**: CDR Global, CDR Sum of Boxes, MMSE, ADAS-Cog 11
- **MRI brain scans**: whole brain, hippocampus, ventricles, intracranial volume

3 class model includes CN, MCI, AD

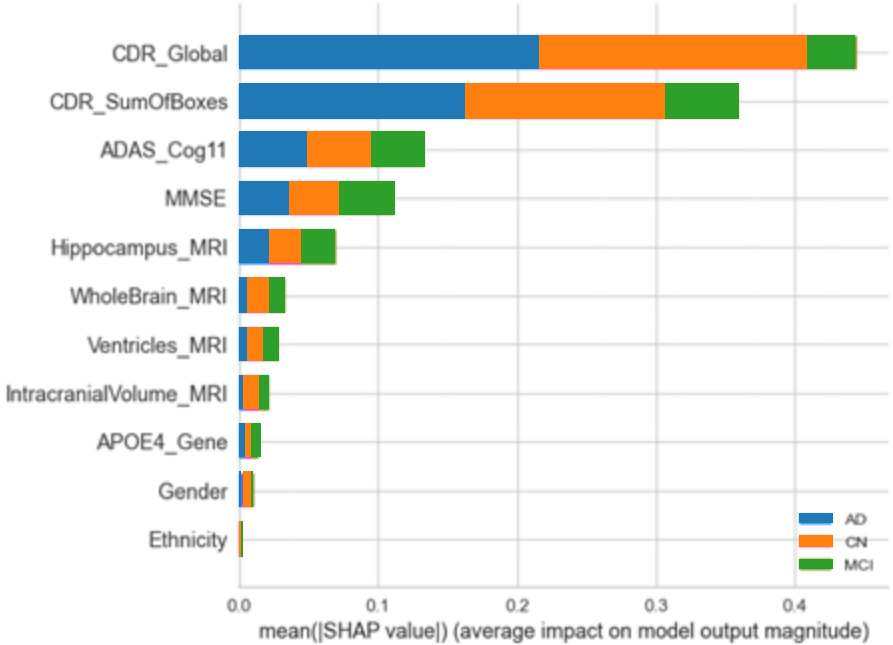
Confusion matrix of model with misclassifications removed:

Proportion of sample predicted among the total true sample within a class.



RESULTS

SHapley Additive exPlanations (SHAP) global feature analysis

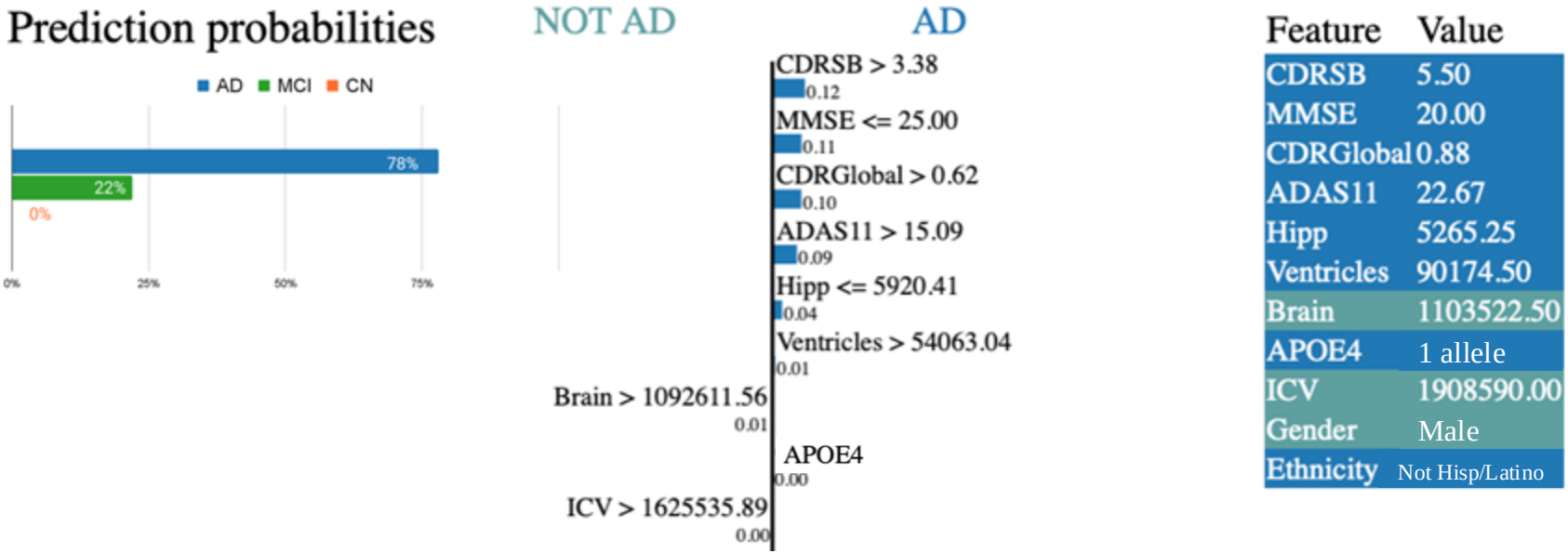


Cognitive scores have the highest impact on all of 3 classes

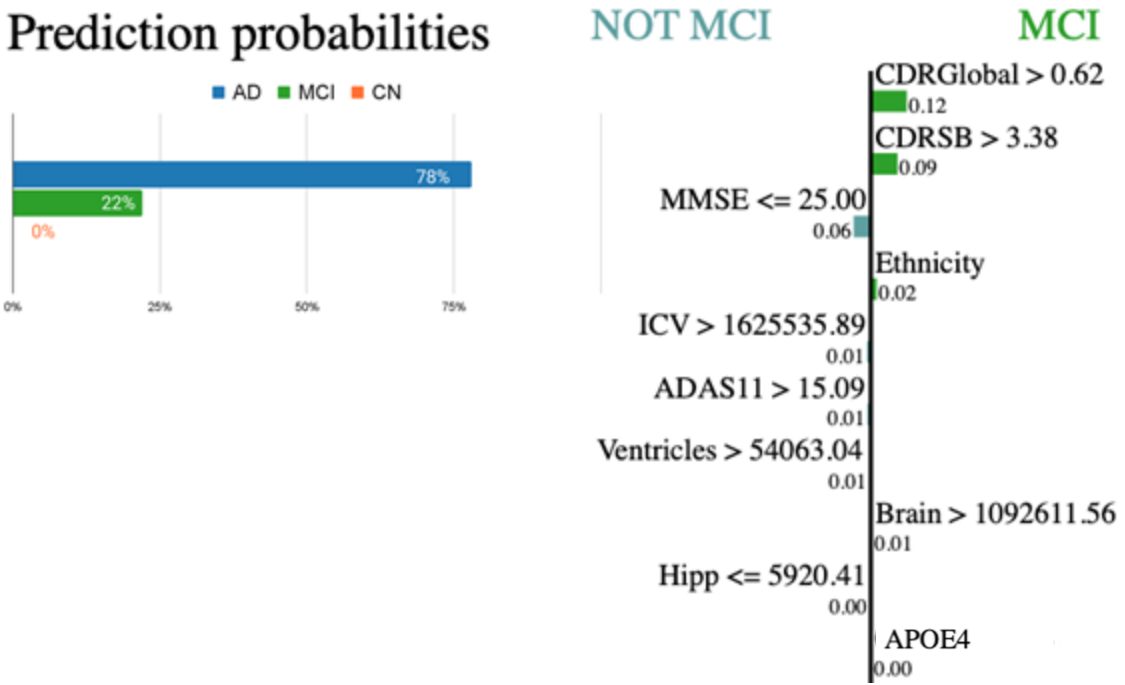
MRI and genetic data are less globally significant for classification

Prediction model works regardless gender or ethnicity

Local Interpretable Model-agnostic Explanations (LIME) feature analysis of a subject predicted as AD

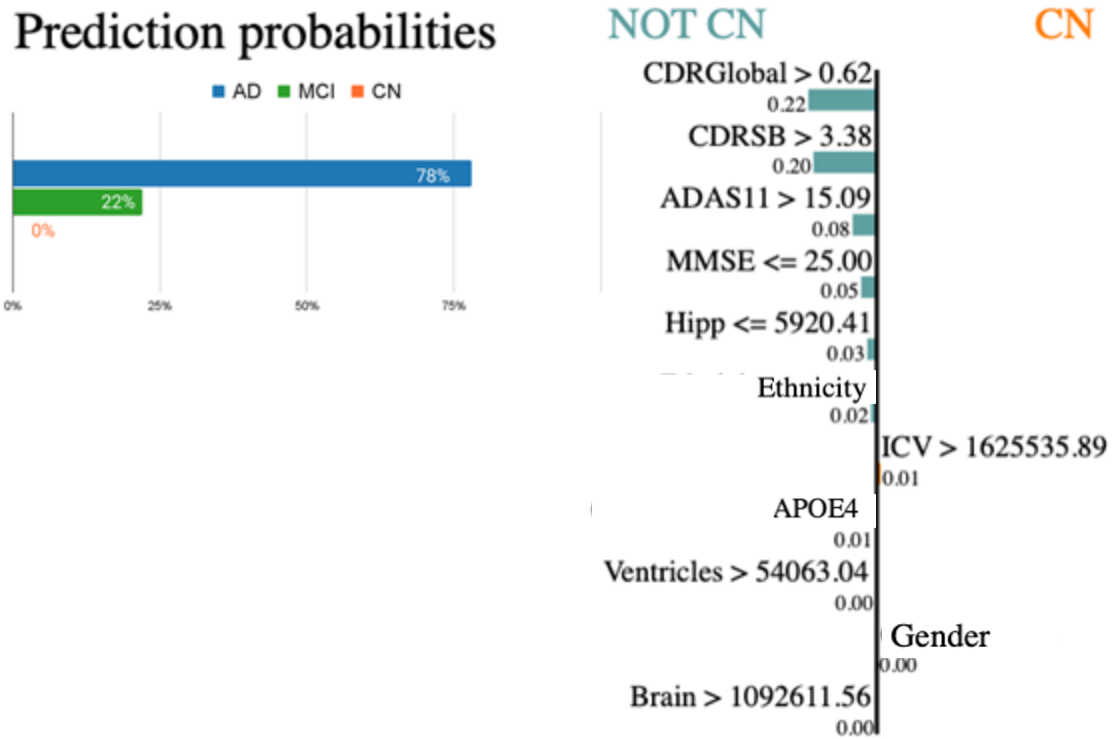


LIME local feature analysis of a subject predicted as AD



Feature	Value
CDRGlobal	0.88
CDRSB	5.50
MMSE	20.00
Ethnicity	Not Hisp/Latino
ICV	1908590.00
ADAS11	22.67
Ventricles	90174.50
Brain	1103522.50
Hipp	5265.25
APOE4	1 allele
Gender	Male

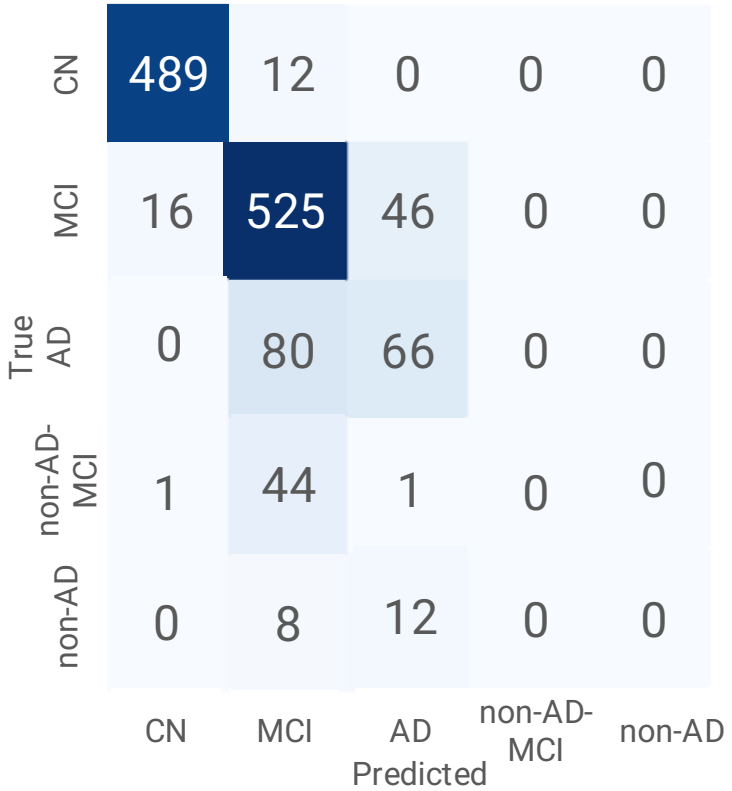
LIME local feature analysis of a subject predicted as AD



Feature	Value
CDRGlobal	0.88
CDRSB	5.50
ADAS11	22.67
MMSE	20.00
Hipp	5265.25
Ethnicity	Not Hisp/Latino
ICV	1908590.00
APOE4	1 allele
Ventricles	90174.50
Gender	Male
Brain	1103522.50

RESULTS

Confusion matrix for 11-features-5-class model



Patients who have dementia or MCI due to non-Alzheimer's etiology are extremely challenging to predict.

Conclusions

1. Multimodal approach improved classifications compared to using only clinical features.
2. Explainable AI approach allows us to get better understanding of model’s decisions.
3. AI model mimics clinician’s decision process of Alzheimer’s disease diagnosis.

Future perspectives

1. Utilize more modalities (e.g. blood tests, medical history) to improve accuracy.
2. Feature selection for non-AD and non-AD-MCI patients to differentiate these 2 classes from others.
3. Test model on real-world clinical data.

Thank you, let’s get in touch.

Dr. Jeffrey Zhang
jeffrey.zhang@princetonpharmatech.com

Slide 2: Content adapted from ComaRecoveryLab/YouTube