
# Machine Learning-based Prediction of disease activity in MS

**Predicting disease activity in Multiple Sclerosis patients – an explainable Machine Learning approach in Mavenclad trials**

Nadia Terranova, PhD
Scientific Director, Head of Quantitative Pharmacology Advanced Data Analytics
Translational Medicine / Quantitative Pharmacology

IQ Workshop on AI/ML for D3
September 15, 2022

MERCK

# Outline

1 **Background and data**

Analysis objectives, MS disease activity definition, clinical trial data

2 **ML methods**

Machine Learning framework

3 **ML-based prediction of MS disease activity**

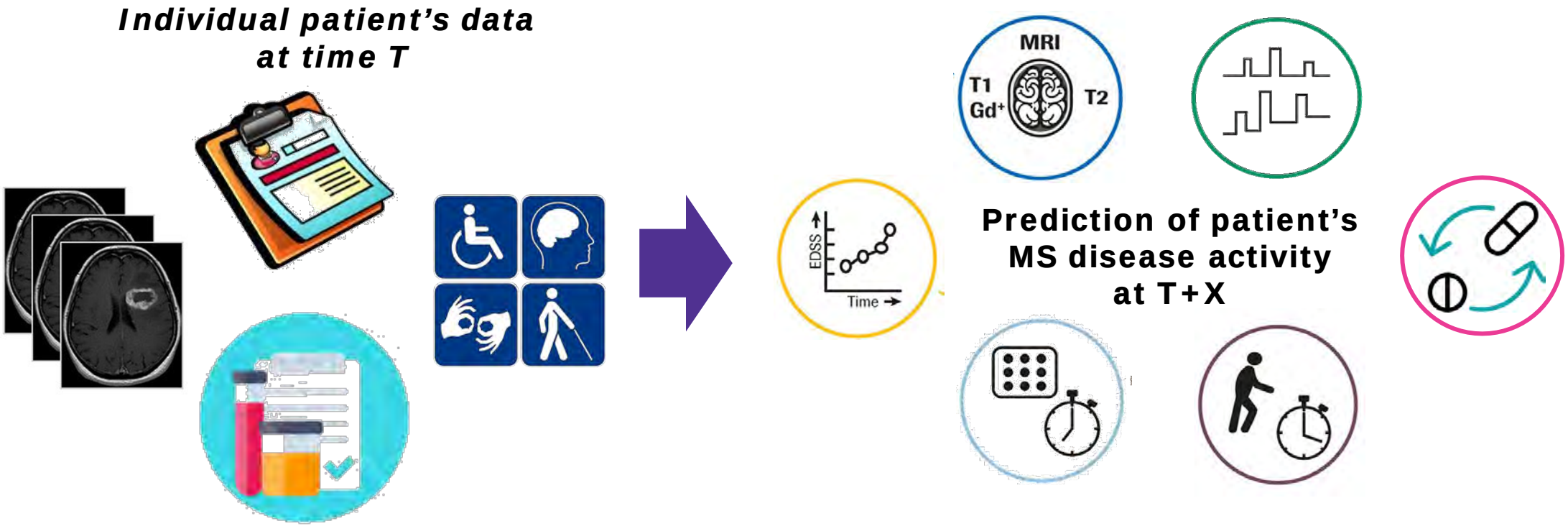
Model performance, covariate importance

4 **Conclusions**

Discussion and ongoing work

# Early identification of patients experiencing the onset of MS disease activity in MAVENCLAD trials

Integrating demographics, response data, MRI and neurological assessments available in cladribine trials to explore which covariates contribute to early identification of MS disease activity by using ML.




Sreetama Basu, Alain Munafo, Ali-Frederic Ben-Amor, Sanjeev Roy, Pascal Girard, Nadia Terranova. *"Predicting disease activity in Multiple Sclerosis patients - an explainable Machine Learning approach in Mavenclad trials"*. CPT:PSP 2022

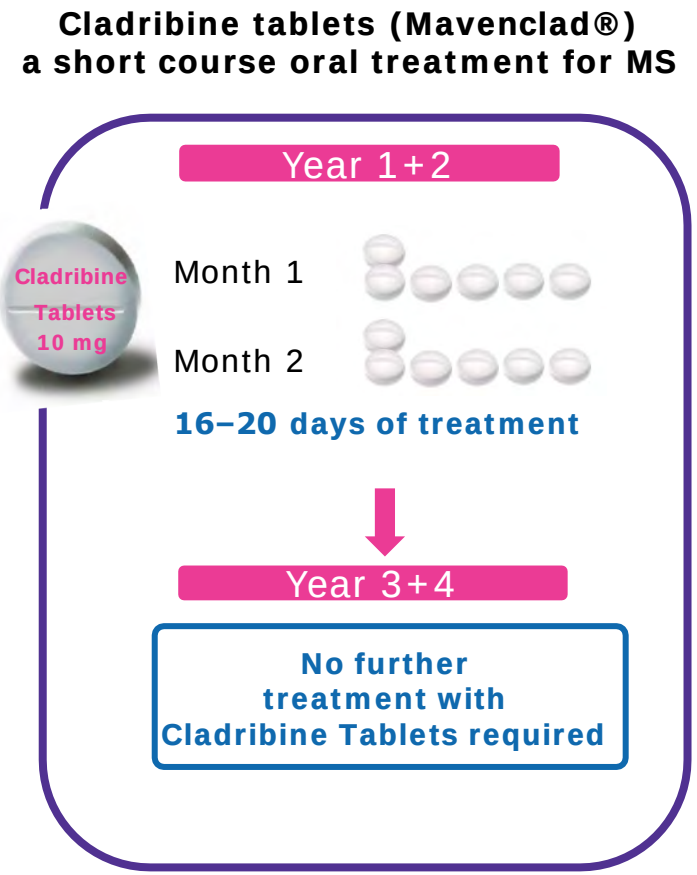


ML: Machine Learning

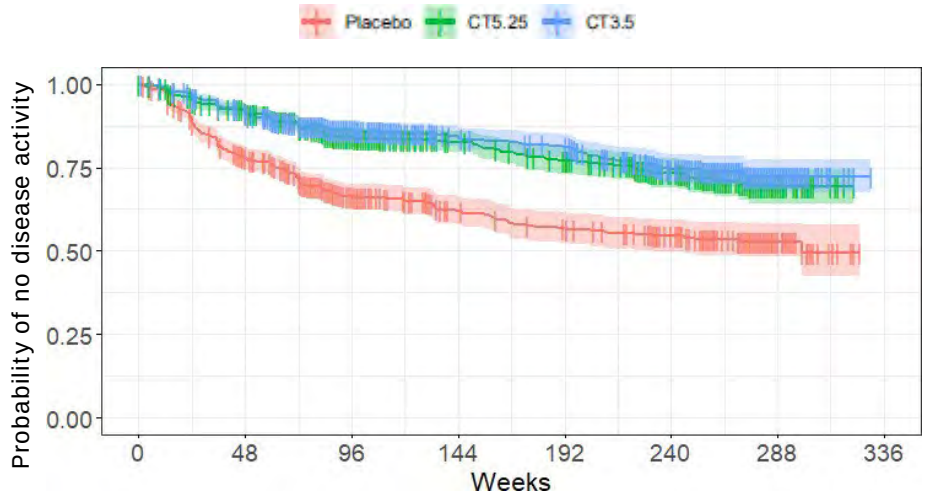
MS: Multiple Sclerosis

# Defining the onset of disease activity to be predicted in cladribine MS dataset

## Application to MAVENCLAD trials




- Pooled Phase III clinical trial data (CLARITY, CLARITY-EXT and ORACLE-MS): 1935 patients, 6+ years of observation, placebo, 3.5mg/kg and 5.25mg/kg cumulative cladribine doses
- Disease activity event for a patient while on cladribine treatment or placebo or observational follow-up, defined as a **composite of endpoints** involving relapses, lesions appearance, EDSS progression and switch to other treatments




DMD: disease-modifying drug (DMD); CT3.5: cumulative cladribine dose of 3.5 mg/kg over 96 weeks; CT5.25: cumulative cladribine dose of 5.25 mg/kg over 96 weeks

# Outline

1 **Background and data**

Analysis objectives, MS disease activity definition, clinical trial data

2 **ML methods**

Machine Learning framework

3 **ML-based prediction of MS disease activity**

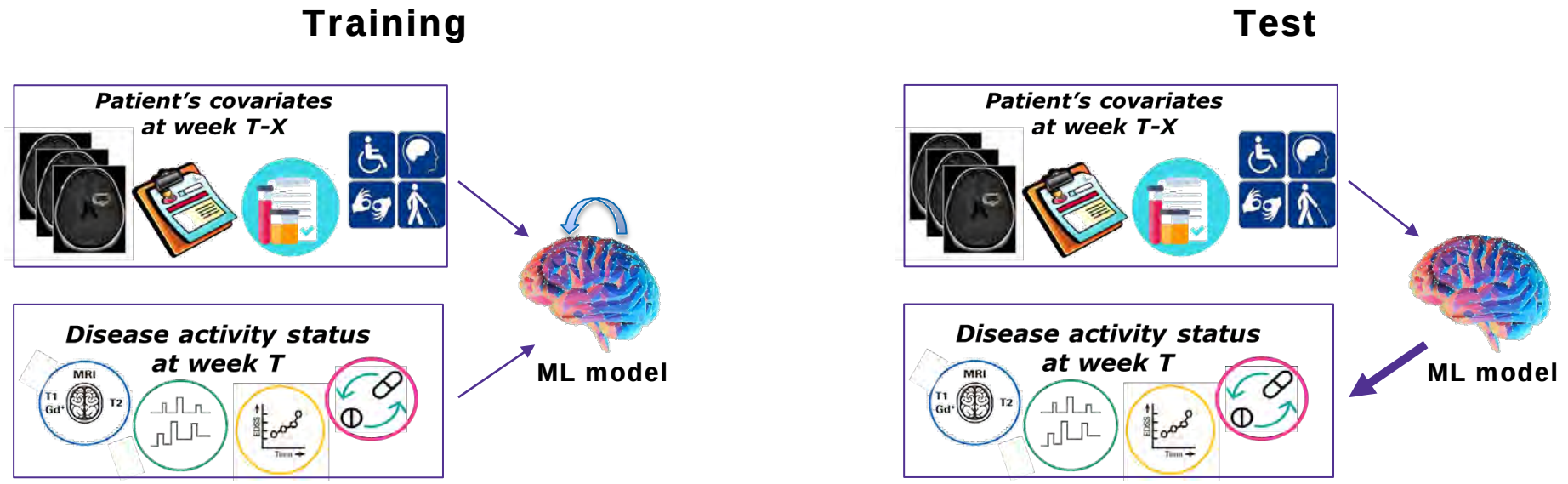
Model performance, covariate importance

4 **Conclusions**

Ongoing work and next steps

# Can patient's MS disease activity be predicted 3 or 6 months in advance?
# **Overview of analysis framework**

We trained and validated 4 models which make predictions of disease activity for patients approximately 3 months (T-12) and 6 months/ (T-24) in advance, based on Phase 3 and Phase 4 covariates.




- 6M/T-24 model : Input covariates 21 -30 weeks, predict disease activity for patient 6 months in advance
- 3M/T-12 model : Input covariates 12-20 weeks, predict disease activity for patient 3 months in advance
- P3 model: Phase 3 covariates
- P4 model: Phase 4 or routine clinical practice covariates

# Phase 3 models are based on a set of 57 independent covariates

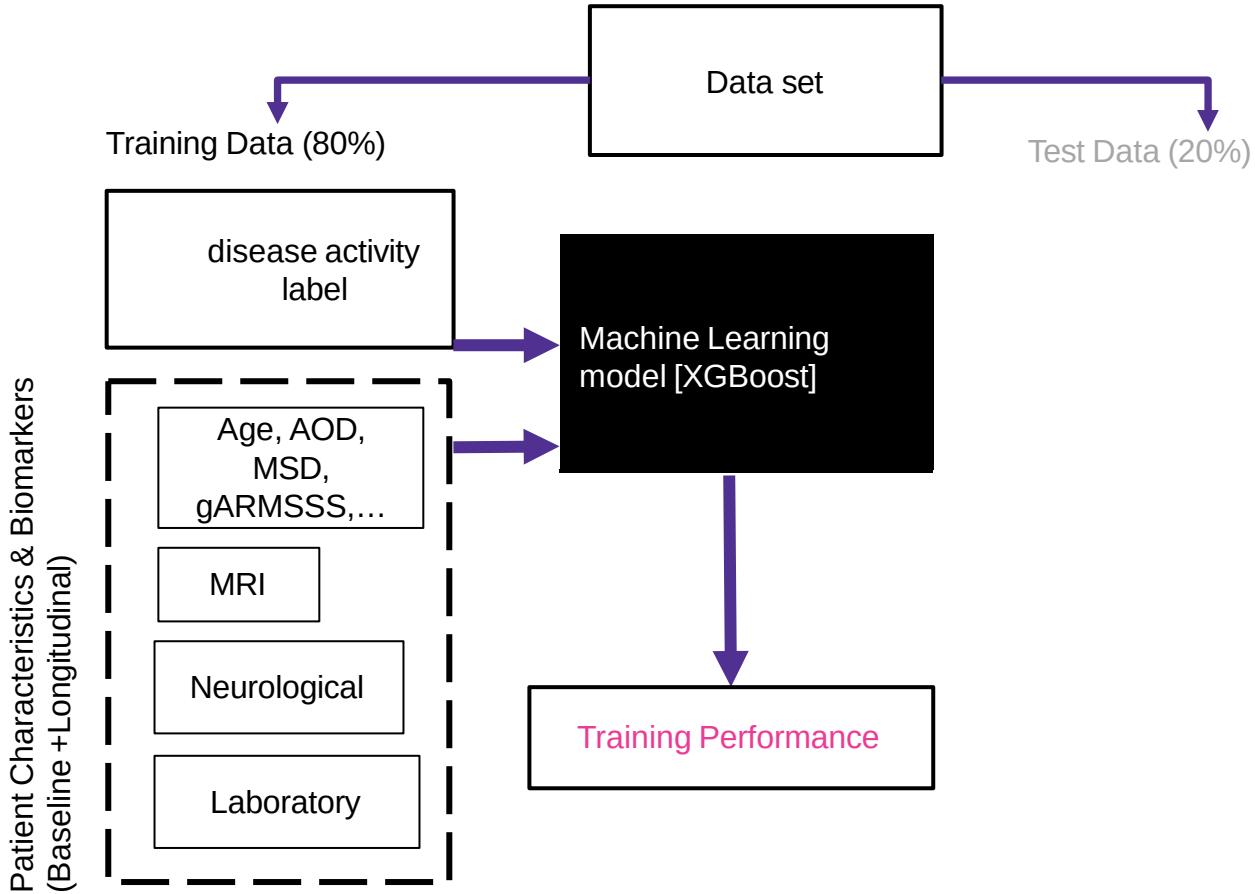
## Independent covariates

| | | | |
|---|---|---|---|
| **Patient characteristics + Baselines** | Age, Sex, Race, **Dose (number of weeks of treatment),** weight, Age of onset of disease, Time since first attack, Lymphocytes_baseline, EDSS_baseline | | |
| **Laboratory** | **Biochemistry**<br><br>Alanine Aminotransferase, Albumin, Alkaline Phosphatase, Aspartate Aminotransferase, Bilirubin, Blood Urea Nitrogen, Calcium, Creatine Kinase, Creatinine, Sodium, Potassium, Urate, Serum Protein, | **Hematology**<br><br>Basophils, Basophils/Leukocytes, Eosinophils, Eosinophils/Leukocytes, Erythrocytes, Hematocrit, Hemoglobin, Leukocytes, Lymphocytes, Lymphocytes/Leukocytes, Monocytes, Monocytes/Leukocytes, Neutrophils, Neutrophils/Leukocytes, Platelets, | **Urinanalysis**<br><br>Urine pH, Glucose, |
| **Neurological assessment** | global age-related multiple sclerosis severity score (**ARMSS**), KFSS1-Bowel and Bladder Functions, KFSS1-Brain Stem Functions, KFSS1-Cerebellar Functions, KFSS1-Cerebral or Mental Functions, KFSS1-Pyramidal Functions, KFSS1-Sensory Functions, KFSS1-Visual or Optic Functions | | |
| **MRI Assessment** | Total Number of T1 Gd+ Lesions, Total T1 Hypointense (Black Holes), Total Number of T2/Flair Lesions, T1 Gd+ (Volume in mm3), T1 Hypointense Lesions (Volume in mm3), T2 Lesions (Volume in mm3), Combined Unique lesion Count, New T1 Hypointense (Black Holes) | | |
| **Removed to avoid target leakage** | Covariates used in the computation of disease activity –<br>Qualified relapse count(RR), New T1 Gd+ lesion count, new& enlarging T2 lesion count, EDSS, and DMT. | | |

# Phase 4 models are based on a subset of 25 covariates routinely available in clinical practice

| | | | |
|---|---|---|---|
| **Patient characteristics + Baselines** | Age, Sex, Race, **Dose (number of weeks of treatment),** weight, Age of onset of disease, Time since first attack, Lymphocytes_baseline, EDSS_baseline | | |
| ~~**Laboratory**~~ | ~~**Biochemistry**~~<br><br>~~Alanine Aminotransferase, Albumin, Alkaline Phosphatase, Aspartate Aminotransferase, Bilirubin, Blood Urea Nitrogen, Calcium, Creatine Kinase, Creatinine, Sodium, Potassium, Urate, Serum Protein,~~ | ~~**Hematology**~~<br><br>~~Basophils, Basophils/Leukocytes, Eosinophils, Eosinophils/Leukocytes, Erythrocytes, Hematocrit, Hemoglobin, Leukocytes, Lymphocytes, Lymphocytes/Leukocytes, Monocytes, Monocytes/Leukocytes, Neutrophils, Neutrophils/Leukocytes, Platelets,~~ | ~~**Urinanalysis**~~<br><br>~~Urine pH, Glucose,~~ |
| **Neurological assessment** | global age-related multiple sclerosis severity score (**ARMSS**), KFSS1-Bowel and Bladder Functions, KFSS1-Brain Stem Functions, KFSS1-Cerebellar Functions, KFSS1-Cerebral or Mental Functions, KFSS1-Pyramidal Functions, KFSS1-Sensory Functions, KFSS1-Visual or Optic Functions | | |
| **MRI Assessment** | Total Number of T1 Gd+ Lesions, Total T1 Hypointense (Black Holes), Total Number of T2/Flair Lesions, T1 Gd+ (Volume in mm3), T1 Hypointense Lesions (Volume in mm3), T2 Lesions (Volume in mm3), Combined Unique lesion Count, New T1 Hypointense (Black Holes) | | |
| **Removed to avoid target leakage** | Covariates used in the computation of disease activity –<br>Qualified relapse count(RR), New T1 Gd+ lesion count, new& enlarging T2 lesion count, EDSS, and DMT. | | |

# Supervised Machine Learning framework
# **XGboost + SHAP**

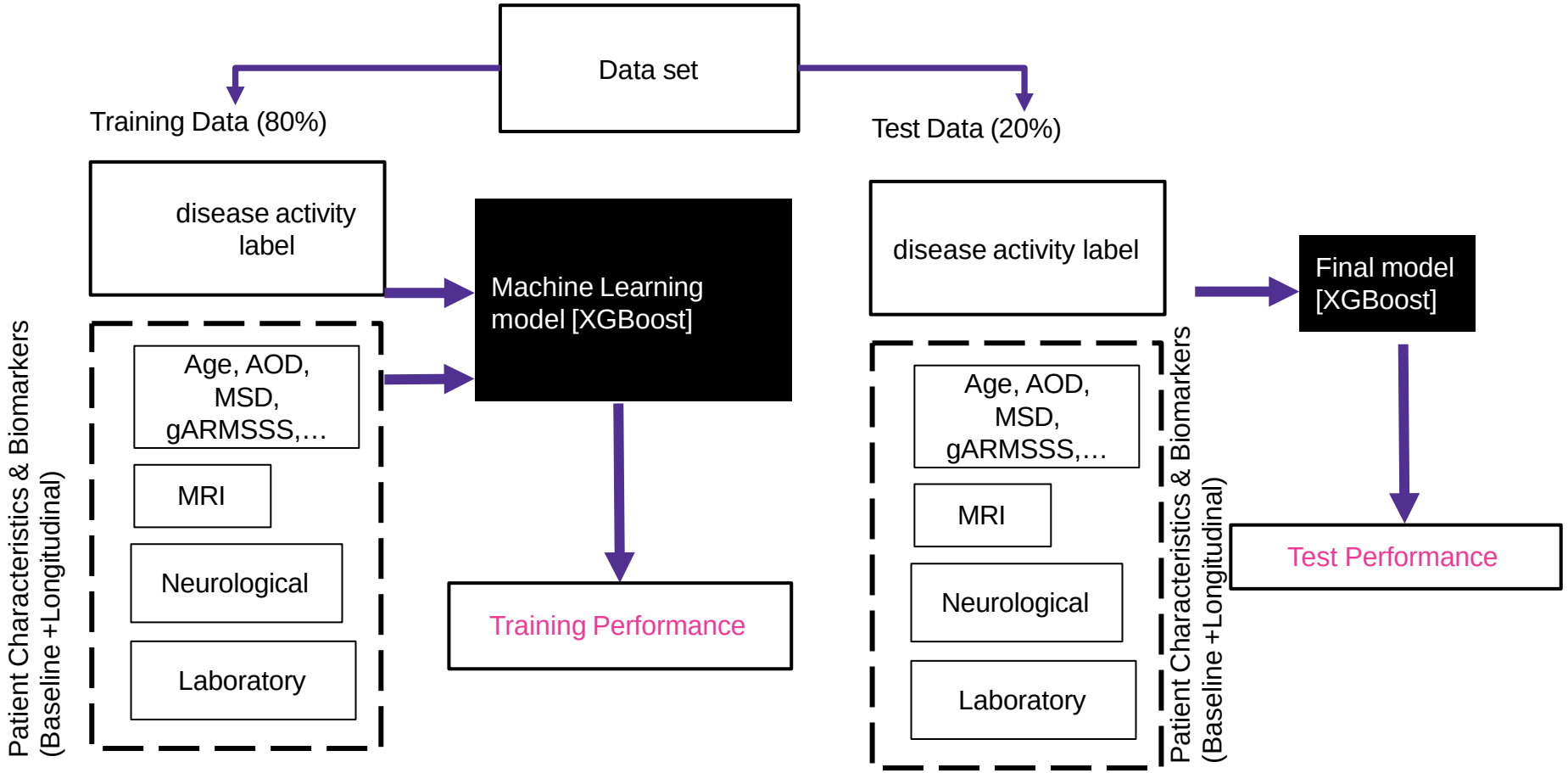



Model parameter selection and optimization through **repeated cross validation (10x10 CV)**



*Basu et. al. CPT:PSP 2022*

# Supervised Machine Learning framework
# **XGboost + SHAP**

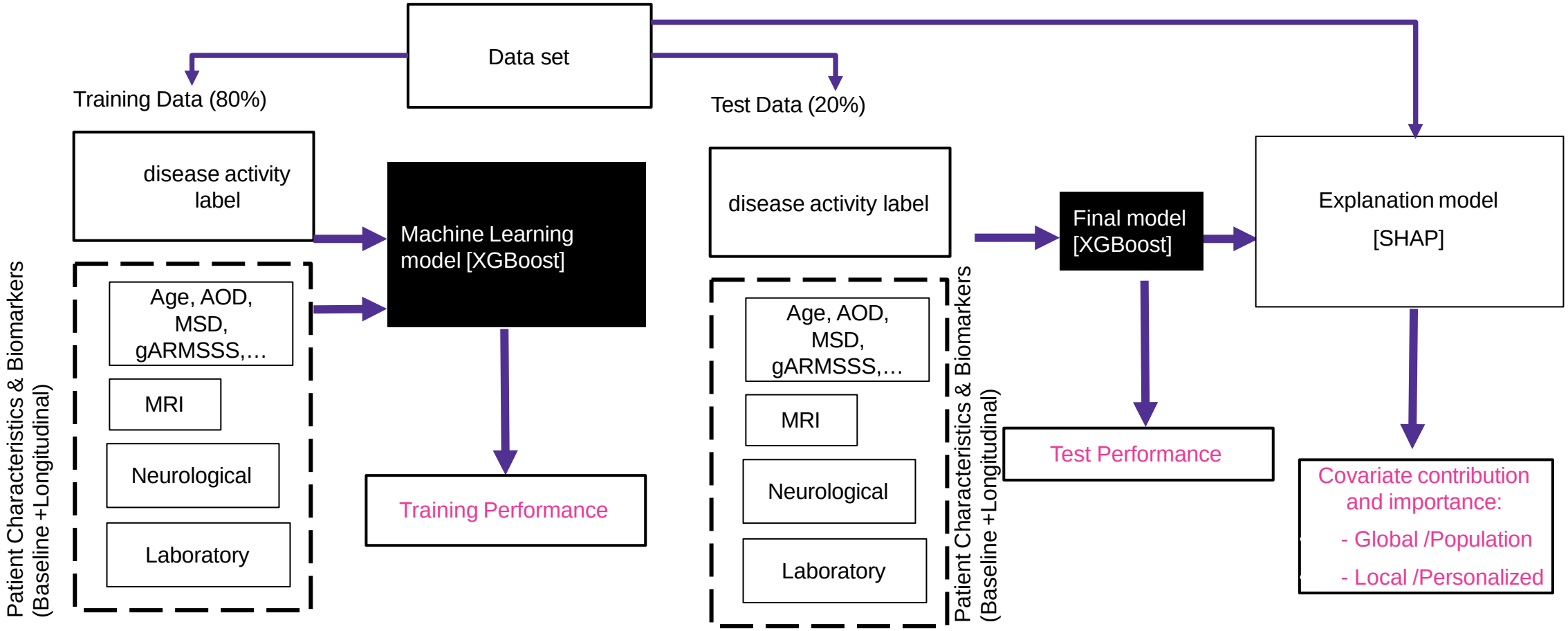



Model parameter selection and optimization through **repeated cross validation (10x10 CV)**



*Basu et. al. CPT:PSP 2022*

# Supervised Machine Learning framework
# **XGboost + SHAP**




Model parameter selection and optimization through **repeated cross validation (10x10 CV)**



*Basu et. al. CPT:PSP 2022*

# Outline

1 **Background and data**

Analysis objectives, MS disease activity definition,

2 **ML methods**

Machine Learning framework

3 **ML-based prediction of MS disease activity**

Model performance, covariate importance

4 **Conclusions**

Ongoing work and next steps

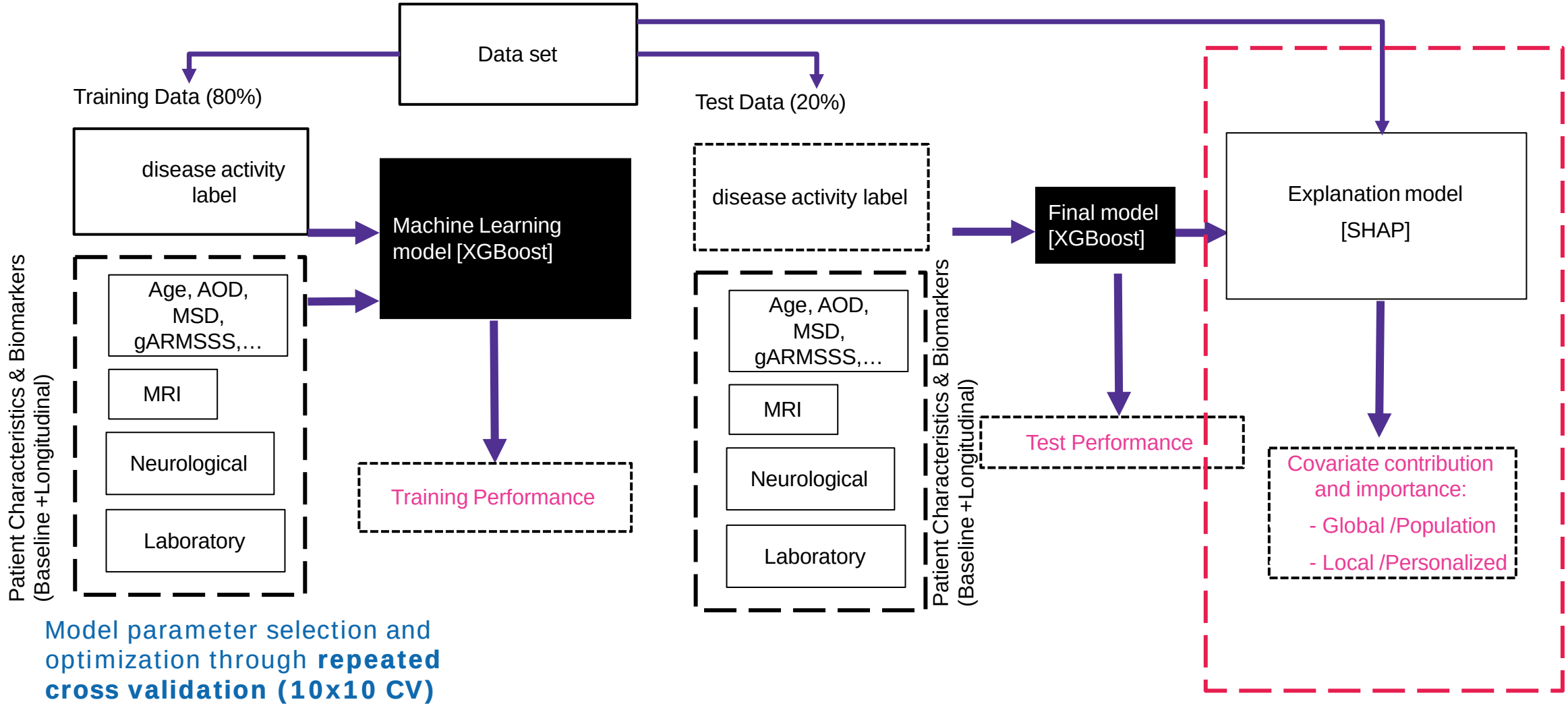
# Good model performance were obtained across all prediction models
## Model results

| | | P3-T-24 | | P4-T-24 | |
|---|---|---|---|---|---|
| | | Train (n=1356) | Test (n=340) | Train (n=1356) | Test (n=340) |
| Specificity | TN/(TN+FP) | 0.76 | 0.76 | 0.77 | 0.78 |
| Sensitivity | TP/(TP+FN) | 0.81 | 0.84 | 0.78 | 0.81 |
| **Balanced Accuracy** | (Sensitivity+Specificity)/2 | 0.79 | **0.8** | 0.78 | **0.8** |
| **Auc-roc** | Area under curve of ROC | 0.79 | **0.8** | 0.78 | **0.8** |

The table lists the model performance on training and test data with several metrics. TN (True Negative), TP (True Positive), FN (False negative), FP (False Positive).

- Training and validation performance are close implying no model overfitting
- Similar results were obtained for the 3-month outcome prediction models (P3-T-12 and P4-T-12)

# Supervised Machine Learning framework
# **XGboost + SHAP**



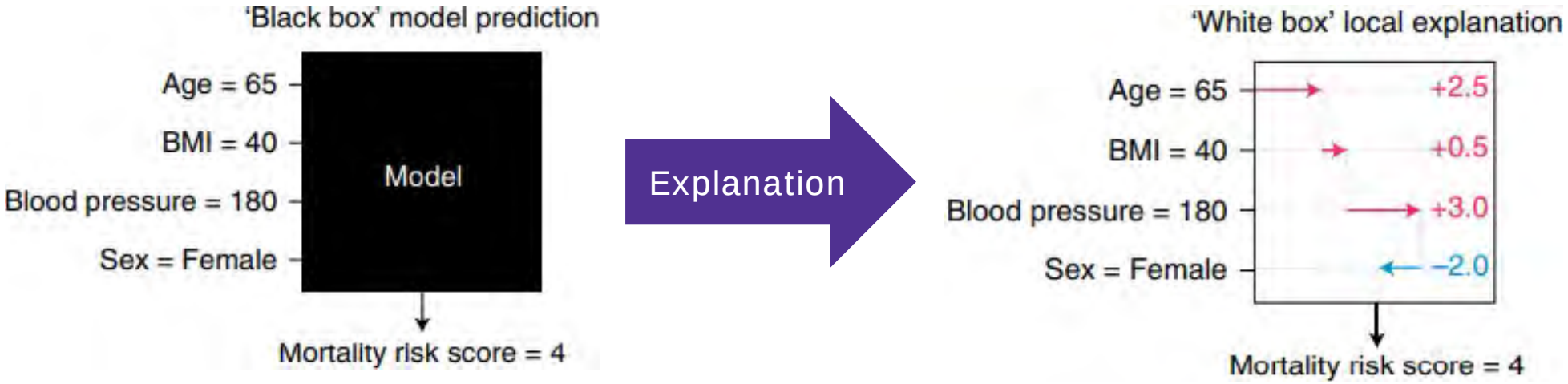

Model parameter selection and optimization through **repeated cross validation (10x10 CV)**



*Basu et. al. CPT:PSP 2022*

# SHapley Additive exPlanations

- **SHAP** is a game theoretic approach to explain the output of any ML model.
- It shows the decomposition of the covariate contributions towards the prediction for a patient.




Lundberg, Scott M., et al. *Nature machine intelligence* 2.1 (2020)

Covariate contribution and importance:

- Local /Personalized
- Global /Population

# SHAP values to explain the predicted DA probabilities of two individuals

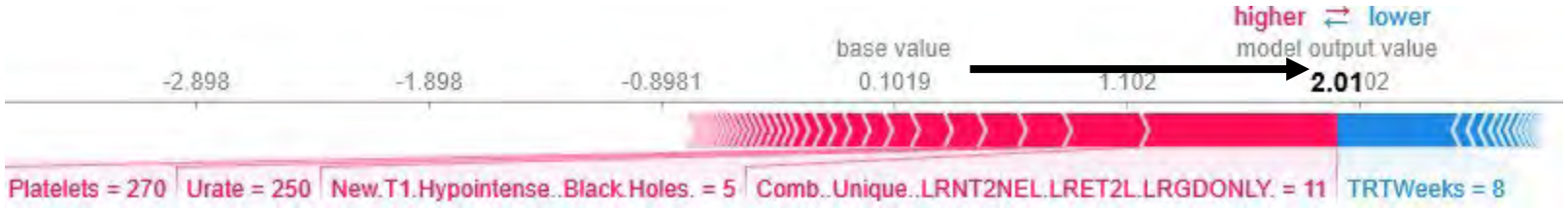
## Examples of P3-T-24 model predictions [Correct]

**No Disease activity**  **Disease activity**

Baseline for SHAP values

### Patient A

| | |
|---|---|
| event | True |
| time | 214 |




### Patient B

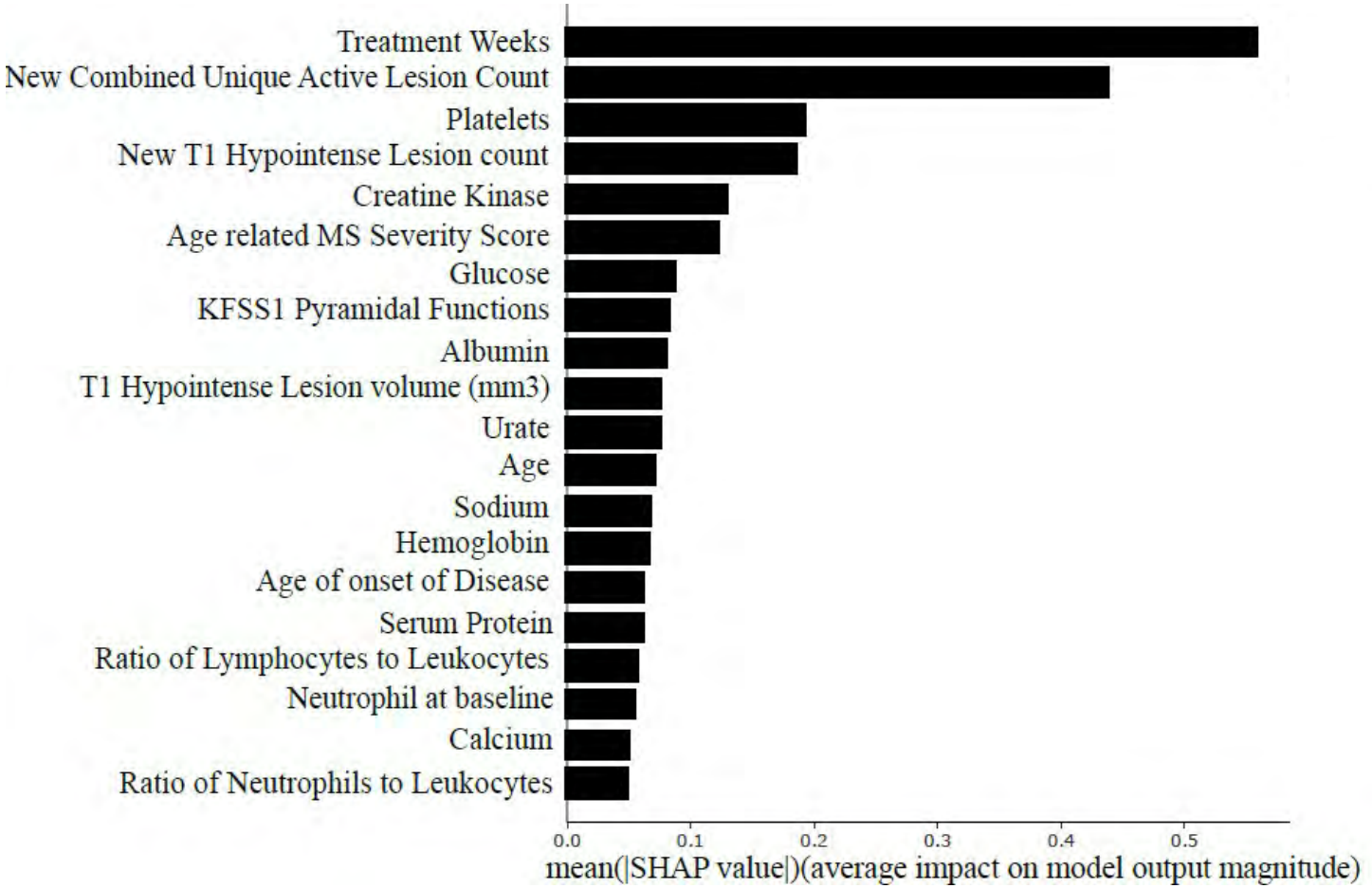
| | |
|---|---|
| event | False |
| time | 231 |


DA: Disease Activity

*Basu et. al. CPT:PSP 2022*

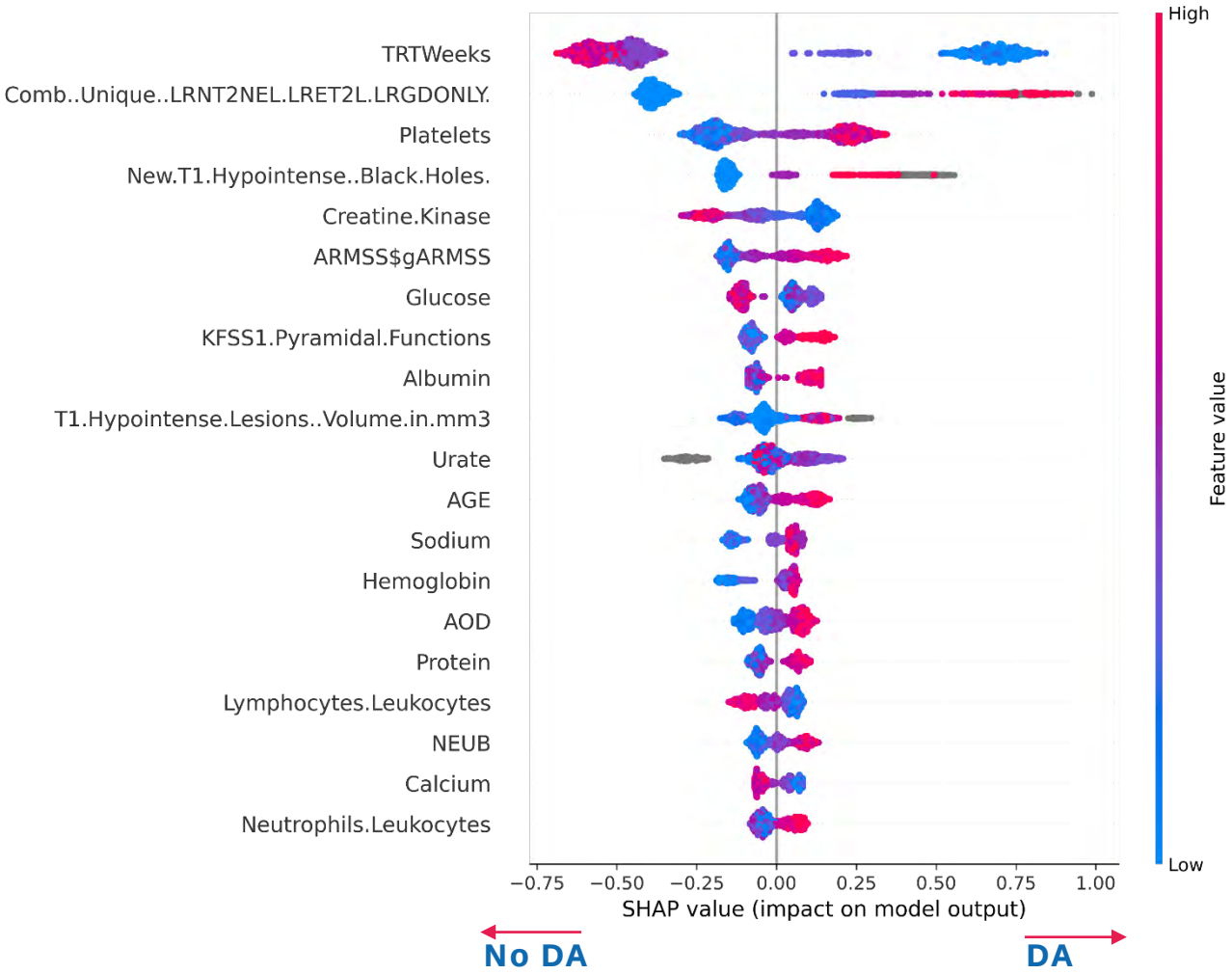
# Treatment weeks, MRI covariates and ARMSS stand out in the global ranking
# **SHAP Feature Importance**

## Top 20 predictive covariates in the P3-T-24 model




*Basu et. al. CPT:PSP 2022*

# Assessing feature effects on DA predictions in a global ranking
## **SHAP summary plot (P3-T-24)**



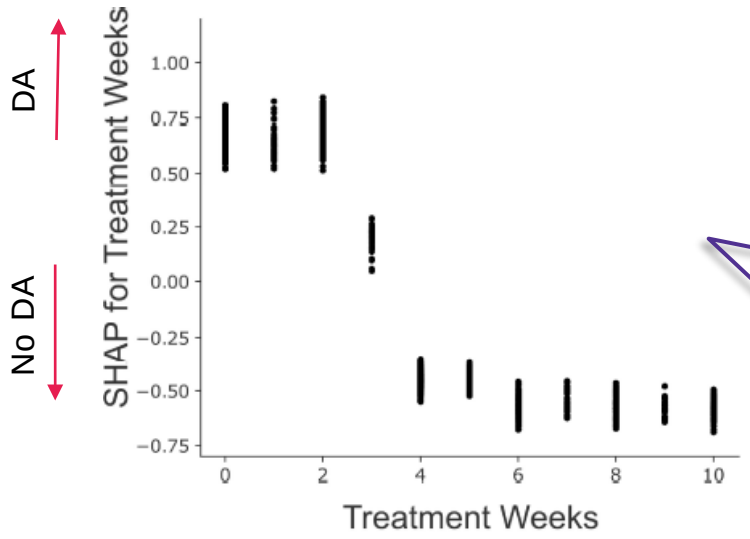

DA: Disease Activity

*Basu et. al. CPT:PSP 2022*

# Global relationships between top predictive covariates and the output variable
# **SHAP dependence plot (P3-T-24)**

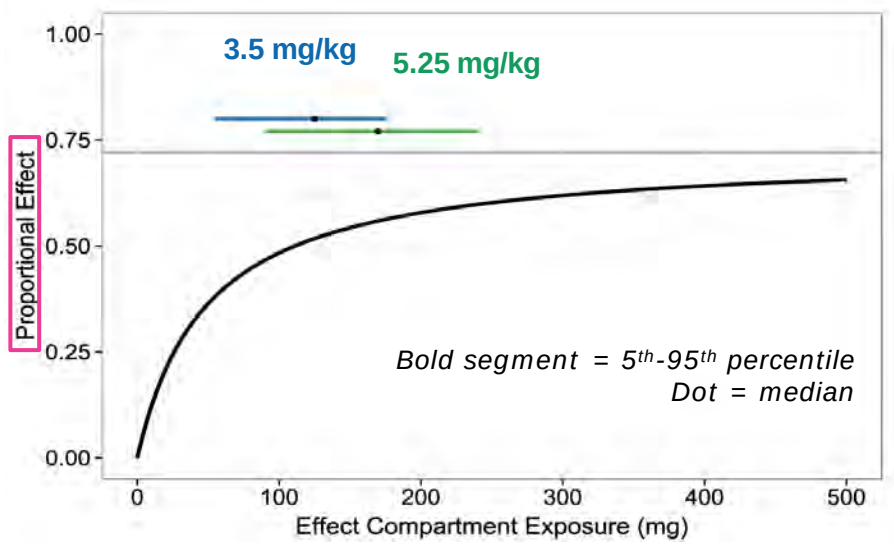
**X:** Covariate ranges

**Y:** DA or No DA




**Population Repeated Time-to-Event model of qualifying relapses**

**Model derived drug-effect relationship**
**with the range of cladribine effect compartment exposure at the end of Year 2**




R. Hermann et al., Clin Pharmacokinet. 2019

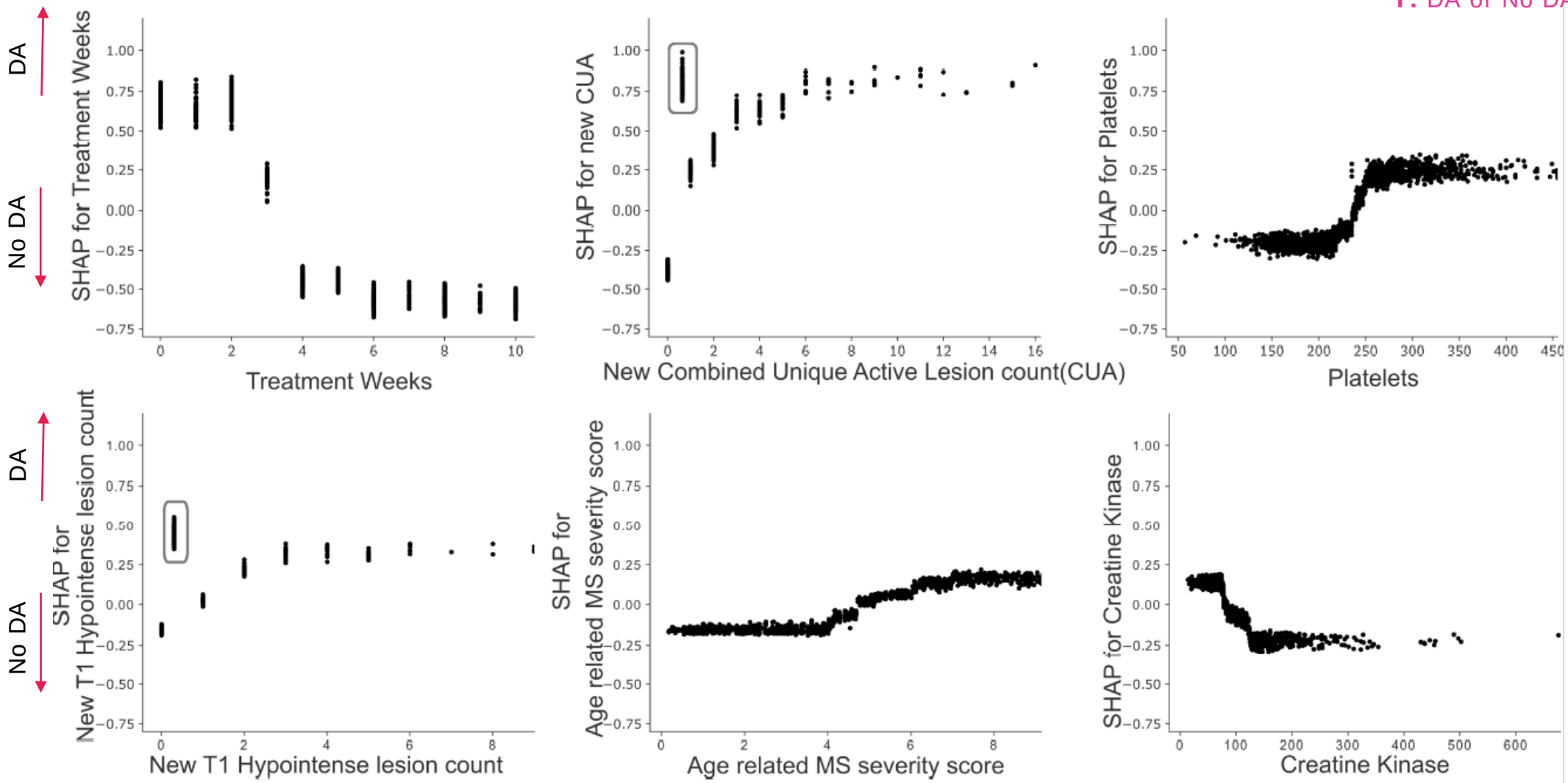
DA: Disease Activity

*Basu et. al. CPT:PSP 2022*

# Global relationships between top predictive covariates and the output variable
## SHAP dependence plot (P3-T-24)

**X:** Covariate ranges

**Y:** DA or No DA




DA: Disease Activity

*Basu et. al. CPT:PSP 2022*

# Outline

1 **Background and data**

Analysis objectives, MS disease activity definition, clinical trial data

2 **ML methods**

Machine Learning framework

3 **ML-based prediction of MS disease activity**

Model performance, covariate importance

4 **Conclusions**

Ongoing work and next steps

# Conclusions

- Placebo population has higher prevalence of disease activity compared to Cladribine treated population.
- 3 month sustained EDSS progression is the most informative criterium, contributing to 42% of detection of patients with disease activity
- T-24 models achieve 80% balanced accuracy[1] and AUROC[2] for disease activity prediction.
- There is a strong overlap of top predictive covariates among the models, with T-12 models showing similar trends as the T-24 models.
- The top predictors of disease activity are Cladribine treatment duration, New Combined Unique Active (CUA) lesion count, New T1 hypointense lesions count, ARMSS, as well as other well understood prognostic factors such as time since first symptom, age of onset in P4 models.

[1] Balanced accuracy: mean of sensitivity and specificity
[2] Auroc: area under RO curve

- ARMSSS: age-related Multiple Sclerosis severity score
- P3 : Phase 3 covariates
- P4 : Phase 4 covariates

# Discussion and next steps

- A dependent output variable encompassing multiple clinical endpoints can provide a more complete picture of disease activation and/or progression

- Cladribine high dimensional data set enabled us to present a feasibility study on how data driven ML models could help in future as a clinical decision support tool
  - interpretable prediction about how clinically available covariates drive the probability of future DA
  - Importance of quality MRI evidence in line with MAGNIMS guidelines for MS
  - Flexible framework for inclusion of newer neurological measures of disability progression e.g., timed 25-meter walk, 9 hole-peg-test

- Dynamic prediction models like recurrent neural networks (RNN) taking as input longitudinal covariates and predicting an updated probability of the risk of disease activity are under exploration

- The developed workflow is being applied to Phase 4 data being generated in few cladribine MS studies

# Take home message

**While ML enables improved predictions mining large datasets, interpretability methods can provide more transparent understanding of the model and results, increasing trust**

# Acknowledgements


- Sreetama Basu, Alain Munafo, Pascal Girard
- Ali-Frederic Ben-Amor, Sanjeev Roy
- Karthik Venkatakrishnan
- Lisa Benincosa
- Nektaria Alexandri, Elisabetta Verdun Di Cantogno, Ursula Boschert, Pierre Yves Berclaz
- MAVENCLAD Project Team

# Questions