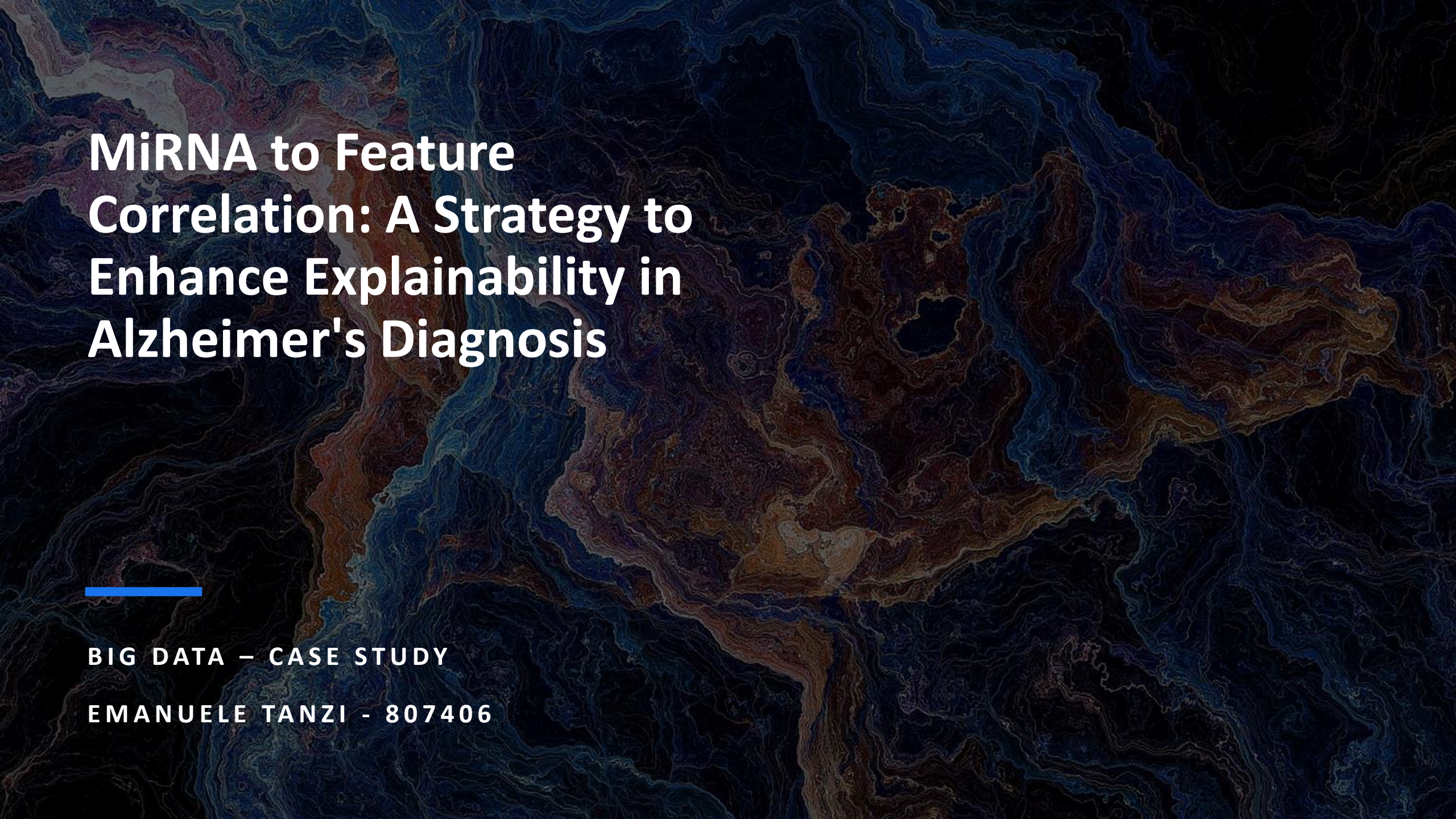


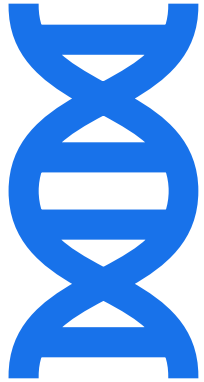
MiRNA to Feature Correlation: A Strategy to Enhance Explainability in Alzheimer's Diagnosis

BIG DATA – CASE STUDY

EMANUELE TANZI - 807406

Table of contents

Introduction: genetic analysis in medical diagnosis



Genetic information
improves early detection and understanding of diseases

Molecular-level analyses
help identify early pathological signatures
(e.g. gene expression, mutation, miRNAs)

Biomarkers

Quantifiable indicators of biological processes (e.g. diseases)

Useful in diagnosis, monitoring, treatment evaluation

Ideal biomarkers:

Non-invasive

Sensitive to early changes

Mechanistically linked to pathology

microRNA (miRNA)

Small, non-coding RNAs, regulate gene expression

Stable in blood, CSF, saliva

miRNA expression pattern reflect:

- Neuroinflammation
- Synaptic dysfunction
- Amyloid-beta metabolism

miRNA-based AD Diagnosis: Ongoing Research

Recent studies explore:

- Distinction between MCI, AD, and healthy aging
- miRNA involvement in neuroinlammation, synaptic disfuction and mitochondrial regulation

Bioinformatics and AI methods used to:

- Map miRNA-gene networks linked to disease progression
- Identify multi-omic biomarker panels (miRNAs, lipids, imaging)

Goal: improve early detection, disease staging, and personalized therapeutic decisions

Introduction of the applied methodology

Starting point: solutions based on complex classifiers, black boxes

Goal: enhancing interpretability, in terms of miRNA expressions

Workflow

1. Data Cleaning
2. Compute mapping between miRNA expressions and embeddings
3. Training a Random Forest model
4. SHAP for interpreting classifications
5. Mapping + SHAP: explainability based on miRNA

Dataset Description

Data obtained as results of a previous study!

subj_embeddings_train.csv

Patient x miRNA embedding

df_concat_final.csv

patient x (miRNA expressions + personal info)

feature_importance_prod_final.csv

miRNA embedding x importance (in classification)

Personal info;

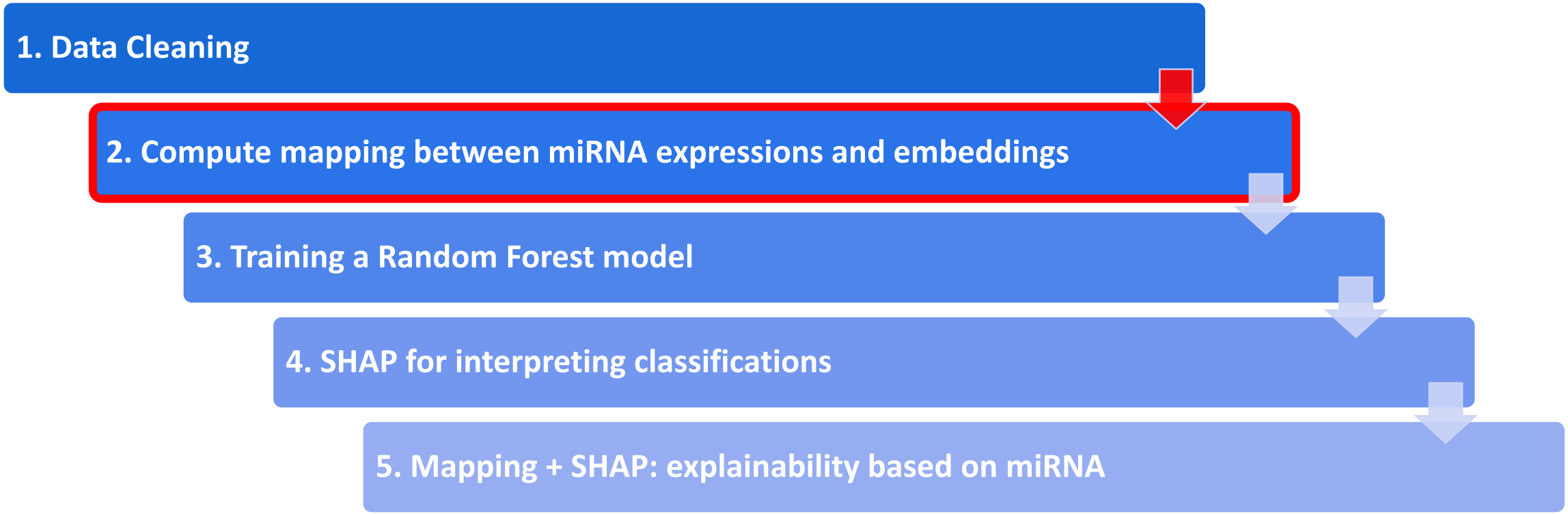
- Age
- Sex
- APOE4
- Diagnosis
  - ***AD***: *Alzheimer's Disease*
  - ***MCI***: *Mild Cognitive Impairment*
  - ***NC***: *Normal Control*

Data Cleaning

$$P = X \cdot H$$

- **H** represents the graph embeddings, relationship between embeddings and miRNA expressions
  - $H \in \mathbb{R}^{2558 \times 128}$
- **P** extracted from ***subj_embeddings_train.csv***
  - $P \in \mathbb{R}^{1256 \times 128}$
- **X** extracted from ***df_concat_final.csv***
  - $X \in \mathbb{R}^{1256 \times 2558}$

Workflow


1. Data Cleaning
2. Compute mapping between miRNA expressions and embeddings
3. Training a Random Forest model
4. SHAP for interpreting classifications
5. Mapping + SHAP: explainability based on miRNA

Compute mapping:
miRNA expressions - embeddings

$$P = X \cdot H$$

- H matrix represents mapping between
  - miRNA expressions represented in X
  - Embeddings represented in H

Model: LassoCV
5-fold cross validation
10,000 max. iter.
StandardScaler on X and P

$$\min_{\beta}\left\{\frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\right\}$$

Idea

Use ***Lasso*** *(Least Absolute Shrinkage and Selection Operator)* to compute the matrix

Why Lasso?

Feature selection	• Automatically discards less relevant miRNAs by shrinking their weights to zero
Sparse & interpretable mapping	• Identifies a minimal set of miRNAs that explain each genetic feature
Dimensionality reduction	• Helps focus only on the most informative signals
Biological insight	• Links miRNAs directly to predictive features, improving explainability
Practical advantage	• Enables reconstruction of embedding features without access to the original transformation

Analysis of the most influential features

- *Feature_importance_prod_final.csv*
  - *Embeddings value, associated to their importance in diagnosing Alzheimer's Disease*
  - *Obtained from the previous study*

	# Feature	# Normalized_Importance
	Missing: 0 (0%) Distinct: 128 (100%)	Missing: 0 (0%) Distinct: 128 (100%)
	Min 0 Max 127	Min 0.003445242... Max 0.018365802...
0	0	0.0053105350935546
1	1	0.0091937406847042
2	2	0.0053960984977914
3	3	0.0076990480212761
4	4	0.0080110573974738
5	5	0.0052441348056182
6	6	0.0159724326935228
7	7	0.0042753879445231
8	8	0.0039425950125196
9	9	0.0070830059892691
10	10	0.0051468801432528
11	11	0.0077276063912849
12	12	0.0081217867605086
13	13	0.0057673064578015
14	14	0.0091953636333213
15	15	0.0069067101443335
16	16	0.0036760188786808
17	17	0.0074218359975064
18	18	0.0078048633728756
19	19	0.0065631170067309

First reduction (on patient - miRNA matrix)

First Lasso training:

- learn to reconstruct the top relevant genetic features in P
  - using the matrix **X.**

Goal:

- saving the matrix of weights
  - The result of training activity,

Why?

- Pinpointing the most relevant miRNA features
  - Reduce dimension of X, removing not influent miRNAs

Result:

- reduction on X
  - from (1256 x 2558) to (1256 x 300)

Training Lasso on the entire dataset

Second reduction (on weight matrix)

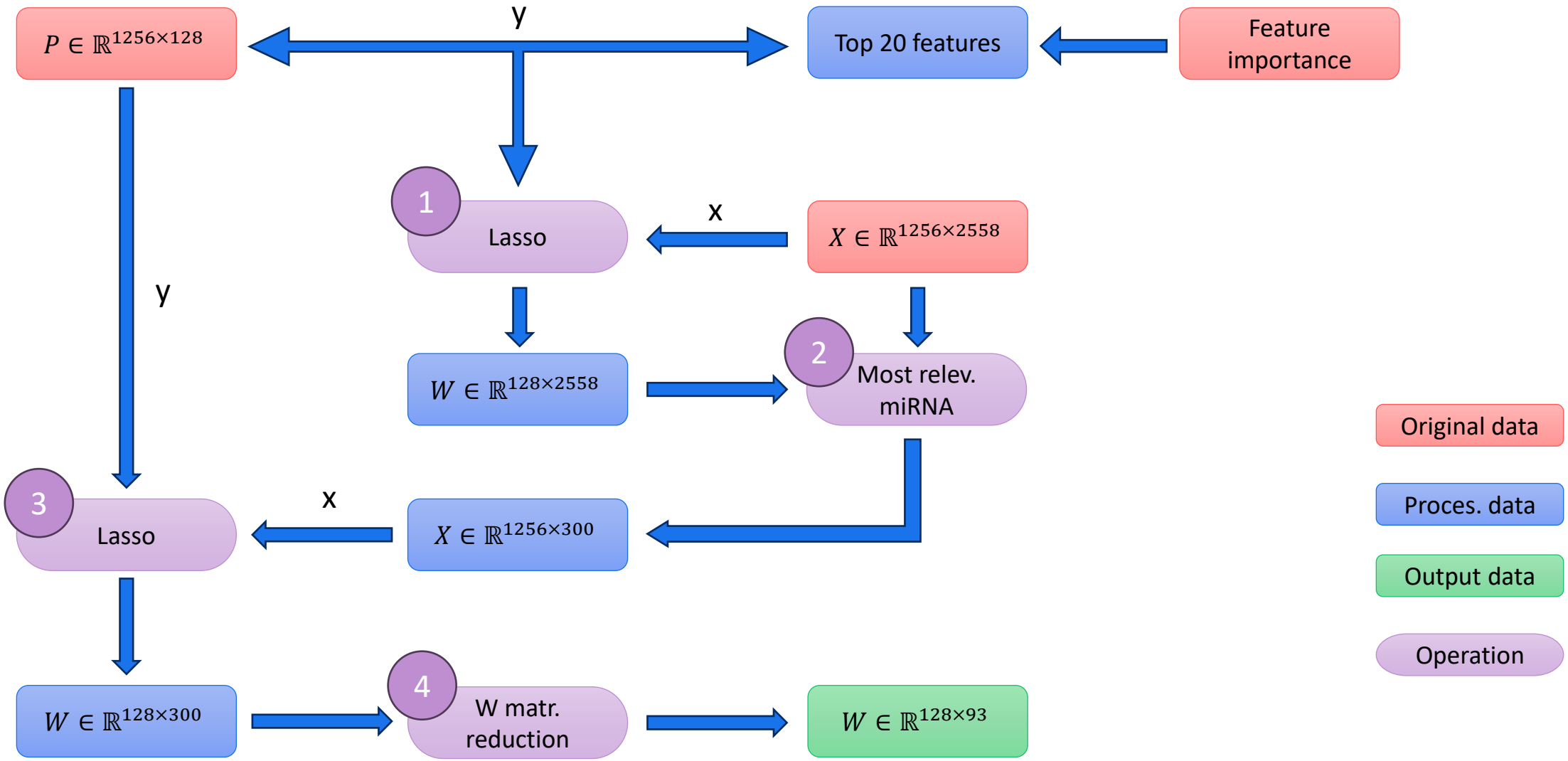
Thresholding strategy on weight matrix

- Threshold: for each row, x% of the maximum absolute

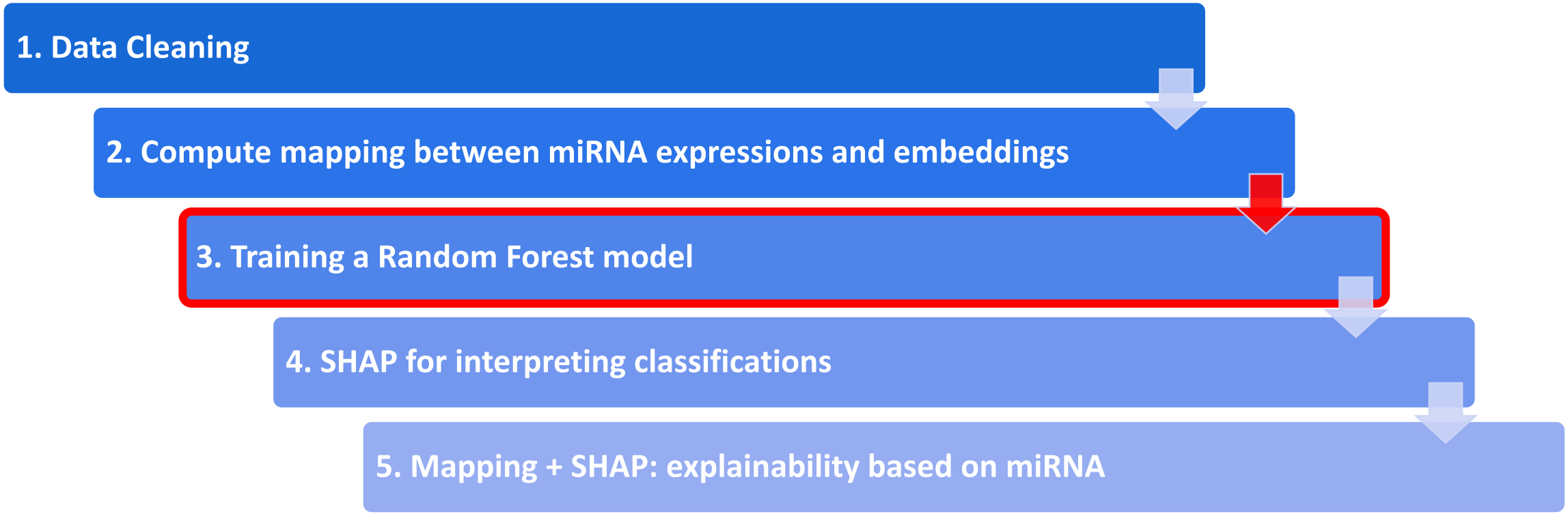
Checking the entries for each row

- Select and delete columns with only zero values
- Final result: **weight matrix (128 x 93)**
- Quality in reconstructing P:
  - $R^2 = 0{,}98 \qquad MSE = 0{,}02$

First step: miRNA-feature mapping computation



Workflow


1. Data Cleaning
2. Compute mapping between miRNA expressions and embeddings
3. Training a Random Forest model
4. SHAP for interpreting classifications
5. Mapping + SHAP: explainability based on miRNA

Random Forest classifier: setup

Purpose:

Establish a baseline for explainability

Included features:

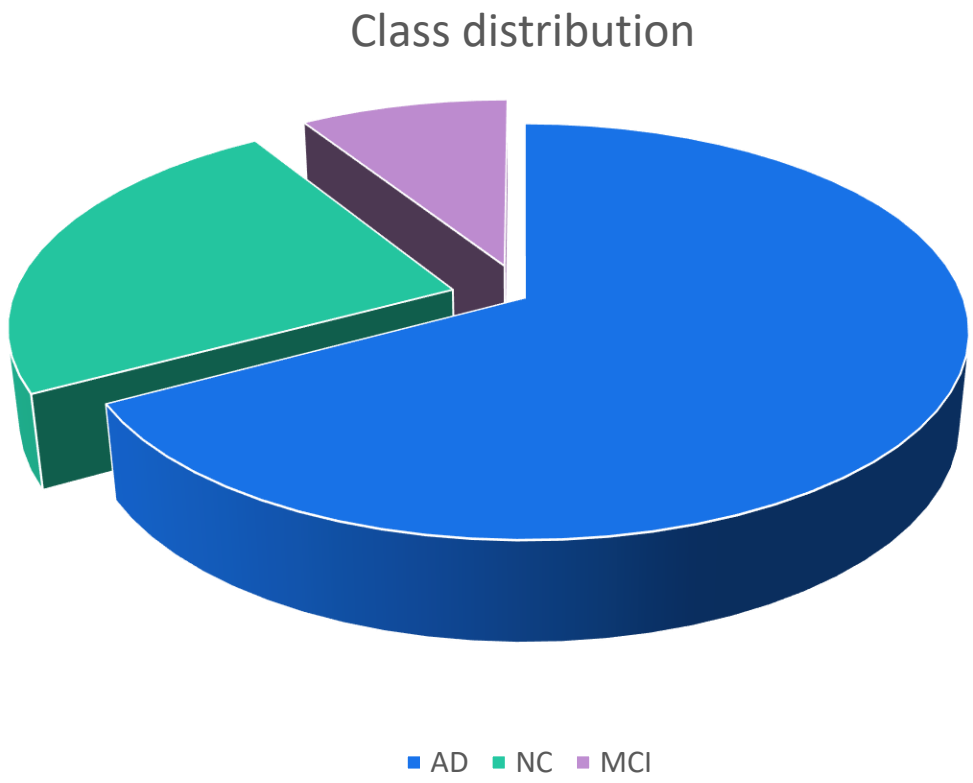
Genetic Embeddings

Patient metadata: age, APOE4

(From dataframe used to extract P)

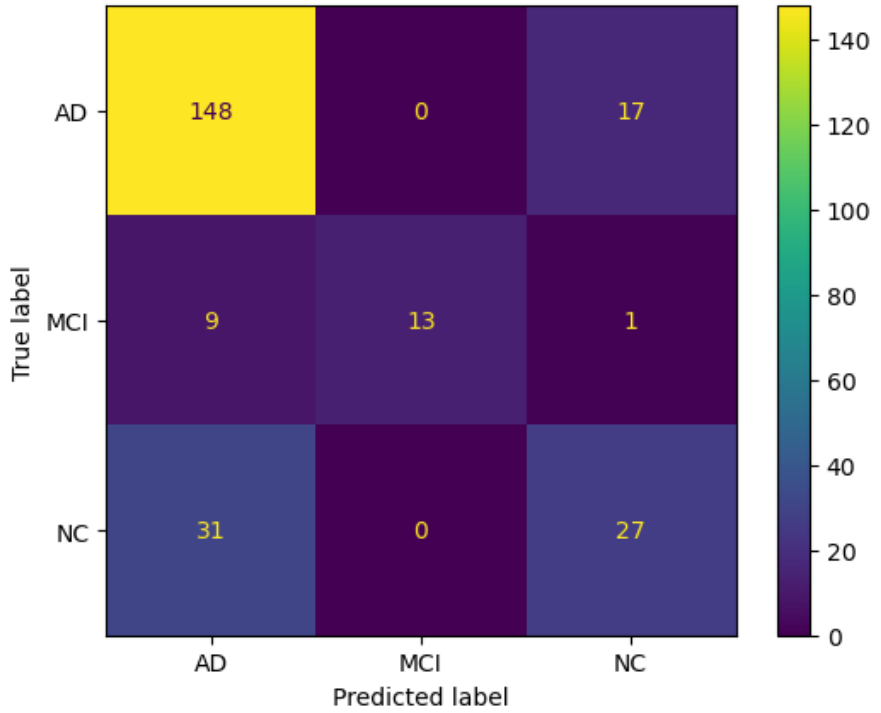
Traning pipeline

1) Removal of entries with missing personal data
   - 2% of the patients have missing data
2) 80/20 train-test split
3) Grid Search for hyperparameter tuning
4) Optimized for F1-macro (imbalanced classes)
   - AD: 67%
   - NC: 24%
   - MCI: 9%



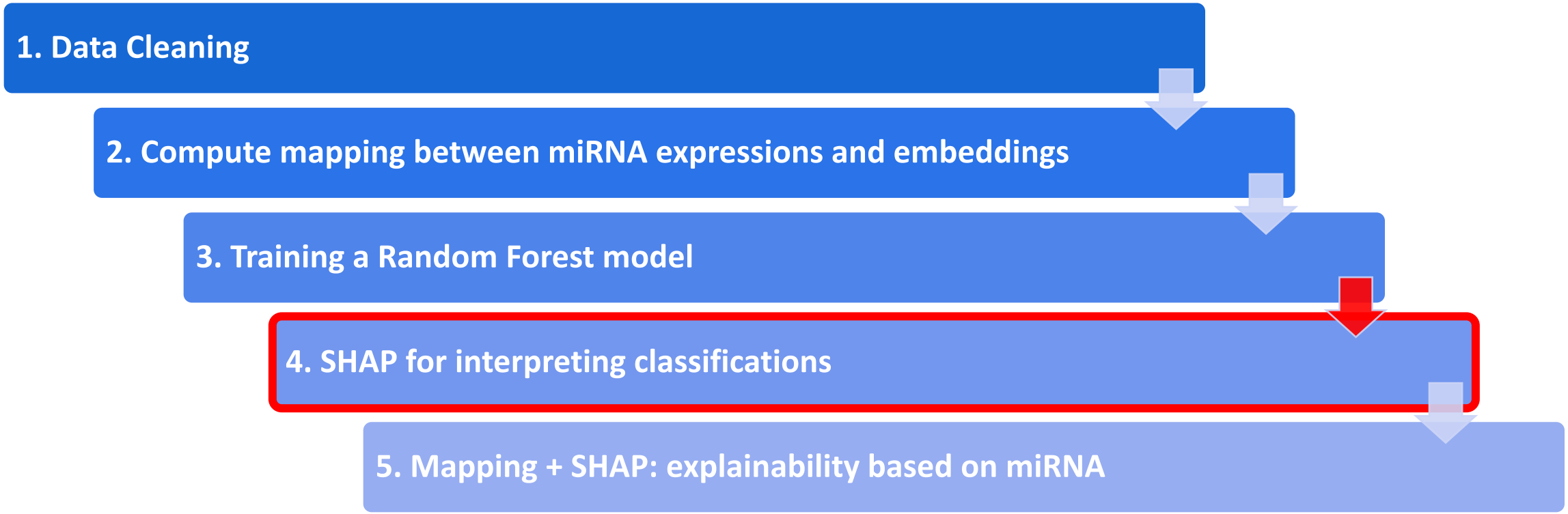
Best model: results

- Overclassification of AD
- MCI classified as AD still offer useful clinical insight
- Main goal: develop an explainability framework, not maximize accuracy



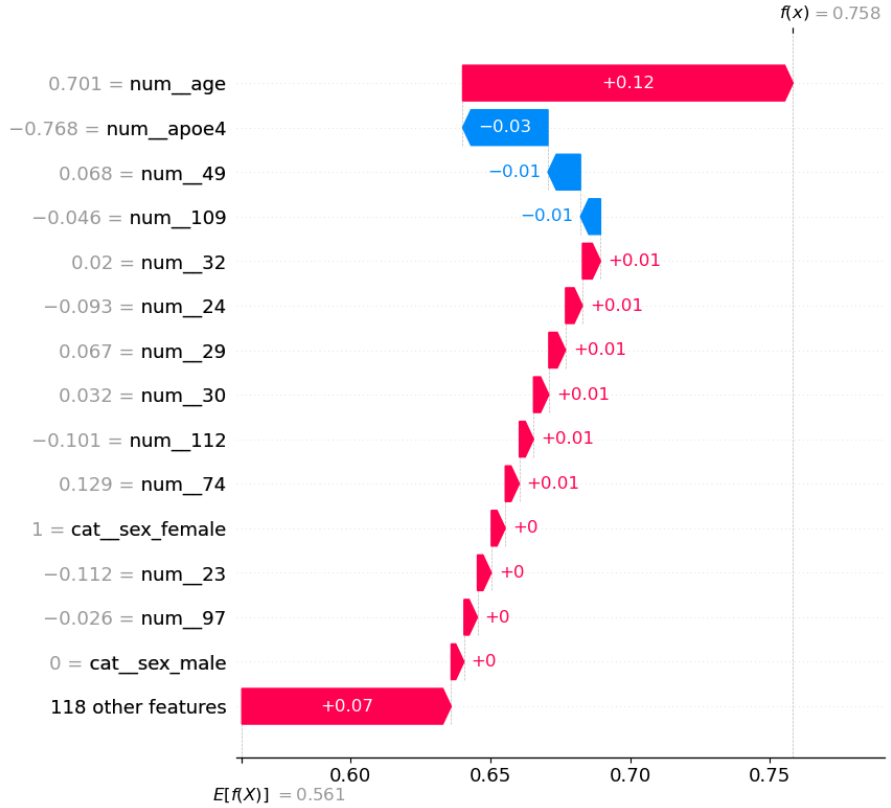
	Precision	Recall	F1-score	Support
AD	0,787	0,897	0,839	165
MCI	1,000	0,565	0,722	23
NC	0,600	0,764	0,754	58

Workflow

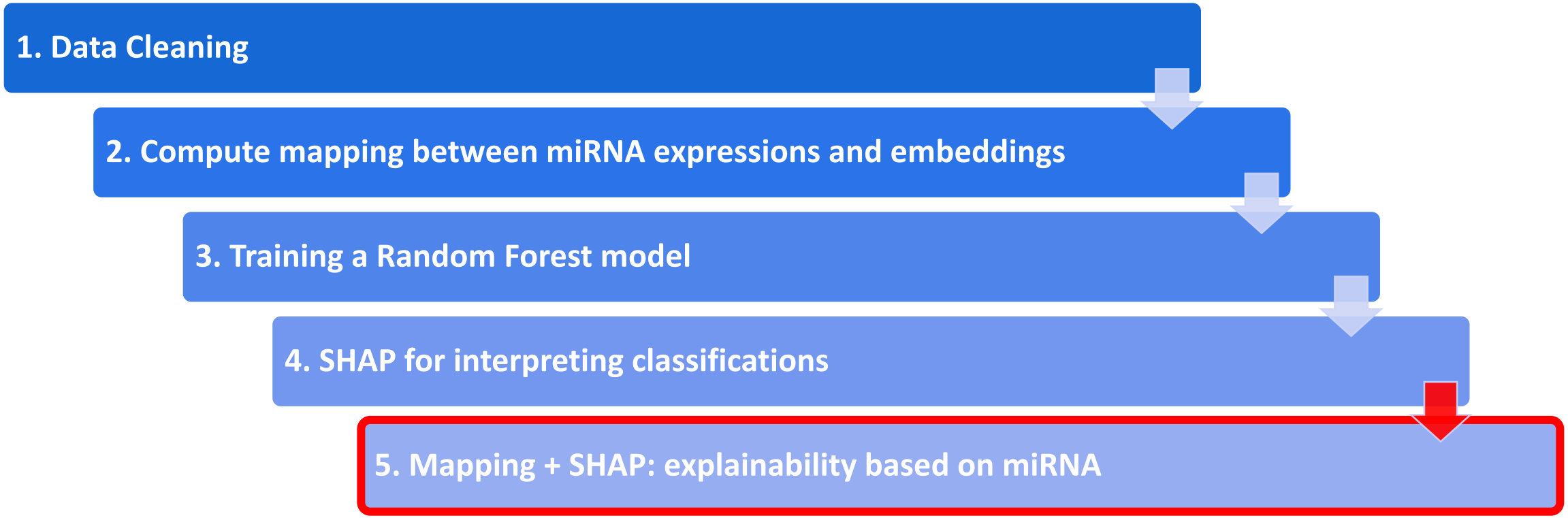

1. Data Cleaning
2. Compute mapping between miRNA expressions and embeddings
3. Training a Random Forest model
4. SHAP for interpreting classifications
5. Mapping + SHAP: explainability based on miRNA

SHAP for interpreting classifications

- **SHAP**: interpret feature contributions to each prediction
- Developed a function to:
  - Return the predicted class
  - List top influencing features, ranked by abs. SHAP values
  - Show both supporting and opposing evidence
  - Include raw feature values for better interpretation

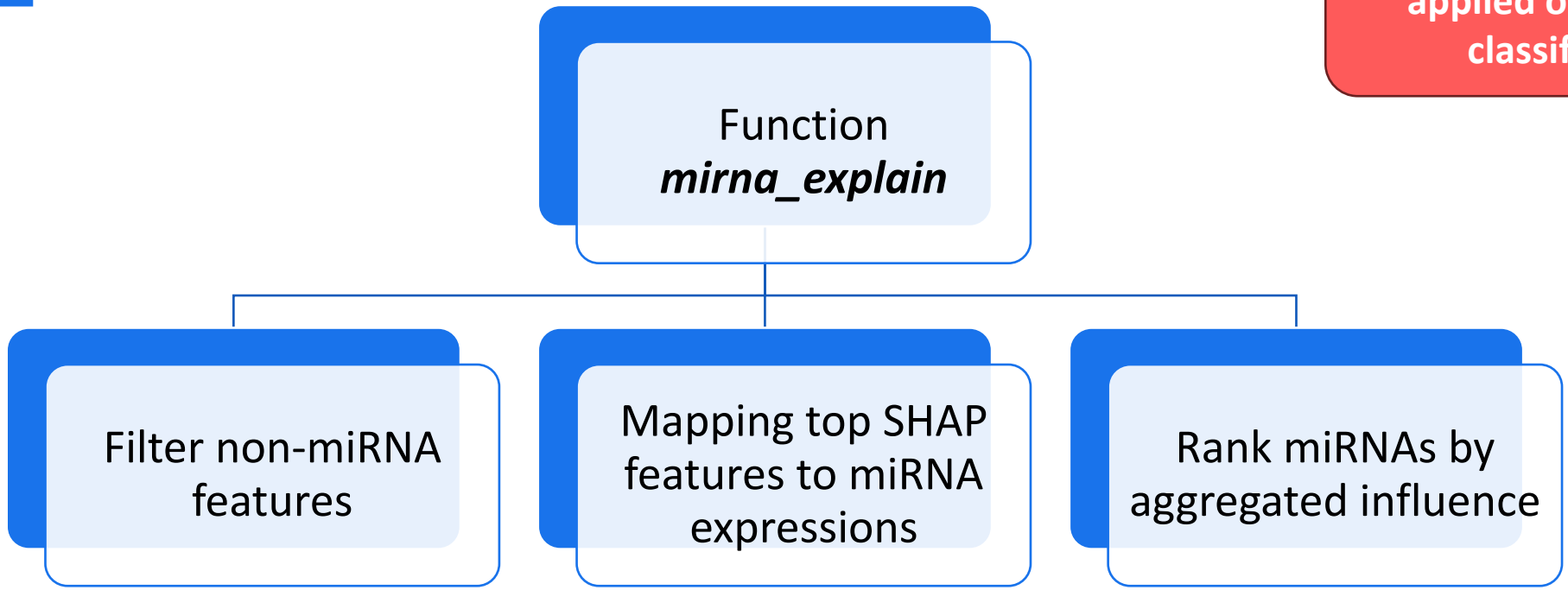


Workflow


1. Data Cleaning
2. Compute mapping between miRNA expressions and embeddings
3. Training a Random Forest model
4. SHAP for interpreting classifications
5. Mapping + SHAP: explainability based on miRNA

Mapping + SHAP: explainability based on miRNA

mirna_explain **to be applied on the single classifications**



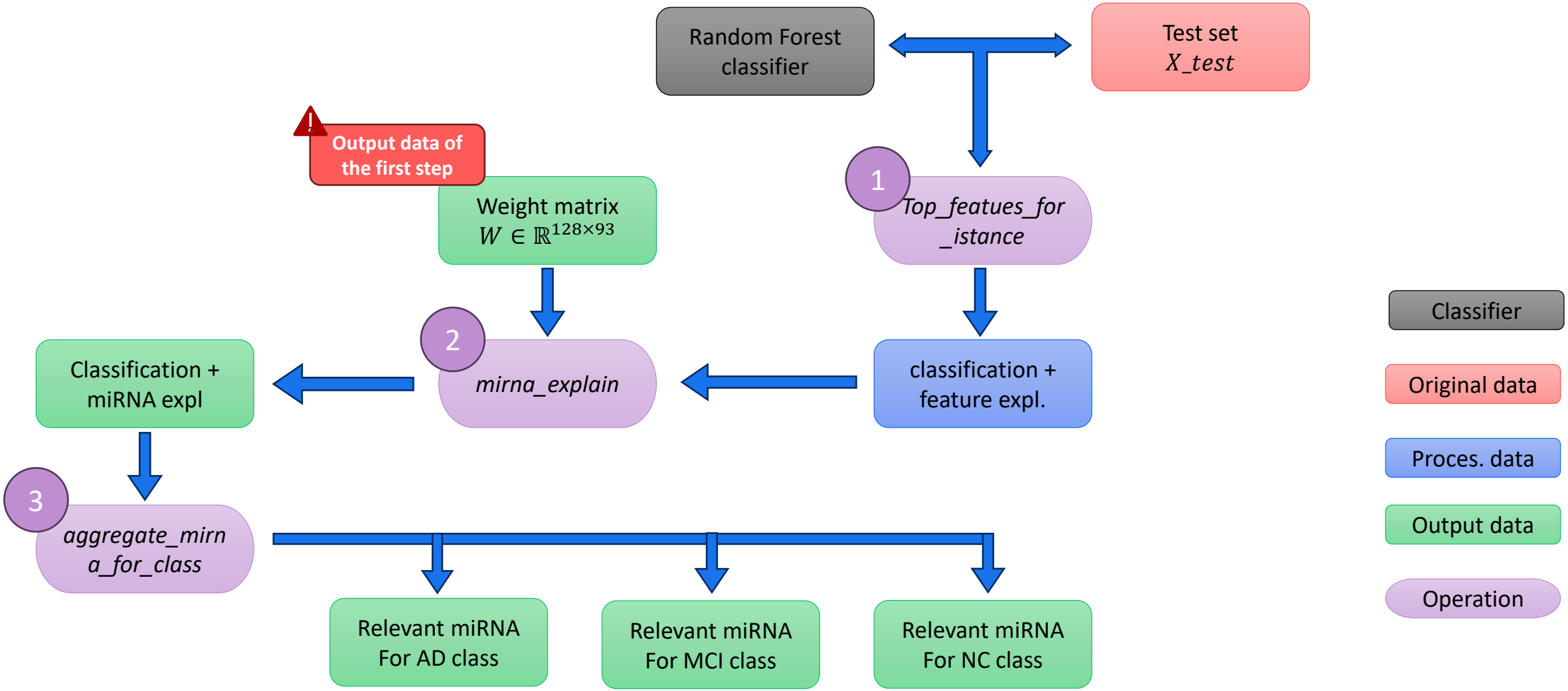
$$agg\_weight\ (m) = \sum_{f \in \mathcal{F}_m} W_{f,m} \cdot SHAP(f)$$

$$agg\_raw\ (m) = \sum_{f \in \mathcal{F}_m} W_{f,m} \cdot RAW(f)$$

miRNA interpretation for correct classifications

- Searching patterns across correct classifications
- Function ***aggregate_mirna_for_class***
  - For each class, using *mirna_explain* results:
    - **mean_weight**: tendency to support class
    - **Std_weight**: consistency of influence
    - **Positive_ratio**: how often it supports the predicted class
      - $\frac{1}{N}\sum_{i=1}^{N} 1(aggr\_weight_i(m) > 0)$
    - **Mean_raw**: average expression value in selected case
    - **Level_vs_avg**: deviation from population average

Second step: explaination for diagnosis



Results

- Two main goals:

ASSOCIATE PREDICTION WITH A PLAUSIBLE EXPLANATION, BASED ON MIRNA BIOMARKERS

IDENTIFY MIRNA BIOMARKERS THAT MAY BE RELEVANT FOR DIAGNOSIS

How to interpret results: SHAP

	feature	shap_value	raw_value	
0	num__age	0.11807506435215145	0.7014892076044722	+
1	num__apoe4	-0.03047287971166497	-0.7684837404432598	-
2	num__49	-0.011635854921354393	0.06833541148984061	-
3	num__109	-0.007178682055379113	-0.045846352374075514	-
4	num__32	0.006494414656083336	0.019938056881912882	+
5	num__24	0.0060228473392271555	-0.09323352282947651	+
6	num__29	0.005978601435954261	0.0668188105949285	+
7	num__30	0.005455914735912586	0.03198035339015115	+
8	num__112	0.0050926949311178756	-0.10139664498823846	+
9	num__74	0.005021357545629144	0.12856700535843185	+

- **The observation of the age feature**, which is significantly above the population average, as indicated by the row_value, pushes towards the predicted class (in the example, *AD*) since the positive shap_value.

- **The observation of the feature apoe4**, which is significantly below the population average, pushes against the predicted class, observing a negative shap_value.

How to interpret results: SHAP + mapping

	miRNA	aggregated_weight	agg_raw
0	hsa-miR-2861	0.011440480264205436	-0.11504813784851482
1	hsa-miR-6729-5p	0.008664085279676197	-0.3790482808194421
2	hsa-miR-1908-5p	0.005742331446907871	0.6307945470772633
3	hsa-miR-5787	0.0055785196181378635	0.31320730732942126
4	hsa-miR-150-5p	0.005062763113042427	0.3731062677960225
5	hsa-miR-1281	0.0042825065667354625	-0.008208151767773142
6	hsa-miR-6126	0.004256624269463154	0.20265186447676567
7	hsa-miR-6786-5p	0.004151927212852016	-0.028433771966038867
8	hsa-miR-92a-3p	0.0032325690591932512	-0.02422571122226973
9	hsa-miR-4306	0.003042857443553049	-0.012078345850617544

- **hsa-miR-2861** positively influenced the classification, being under-expressed relative to the population average.
- **hsa-miR-1908-5p** also contributed positively to the classification, but in this case through over-expression.

miRNA behavior: **AD class**

	miRNA	mean_weight	std_weight	positive_ratio	mean_raw	direction	level_vs_avg
8	hsa-miR-1908-5p	0.0042395262530651985	0.00862409626705597	0.7094594594594594	0.9970062496696276	pro-class	high
87	hsa-miR-7704	-0.004204262707793654	0.009632497183149897	0.2635135135135135	-1.342058296125548	anti-class	low
69	hsa-miR-6780a-3p	0.0031126545082062933	0.0064212977875222075	0.6891891891891891	1.0975439195919257	pro-class	high
61	hsa-miR-6729-5p	-0.0023608174315384025	0.011207023553603672	0.40540540540540543	-0.5714663604836641	anti-class	low
89	hsa-miR-8072	-0.0018427904331291016	0.003550289512466748	0.3581081081081081	-0.5561753363800935	anti-class	low
26	hsa-miR-3937	-0.0016977874777172684	0.0031805748558630747	0.24324324324324326	-0.5905679670898752	anti-class	low
48	hsa-miR-5196-5p	0.0016888161860586584	0.0043187601861377125	0.6148648648648649	1.009707223726697	pro-class	high
40	hsa-miR-4734	-0.0016145116066542701	0.005768183915825981	0.35135135135135137	-1.2382792355219963	anti-class	low
51	hsa-miR-5787	0.0015044805612350525	0.005955225501346829	0.6216216216216216	0.30345303676793106	pro-class	high
50	hsa-miR-557	-0.0013776356728777092	0.0021118352990586184	0.25	-0.23202197323295792	anti-class	low

miRNA behavior: **MCI class**

	miRNA	mean_weight	std_weight	positive_ratio	mean_raw	direction	level_vs_avg
8	hsa-miR-1908-5p	0.0042395262530651985	0.00862409626705597	0.7094594594594594	0.9970062496696276	pro-class	high
87	hsa-miR-7704	-0.004204262707793654	0.009632497183149897	0.2635135135135135	-1.342058296125548	anti-class	low
69	hsa-miR-6780a-3p	0.0031126545082062933	0.0064212977875222075	0.6891891891891891	1.0975439195919257	pro-class	high
61	hsa-miR-6729-5p	-0.0023608174315384025	0.011207023553603672	0.40540540540540543	-0.5714663604836641	anti-class	low
89	hsa-miR-8072	-0.0018427904331291016	0.003550289512466748	0.3581081081081081	-0.5561753363800935	anti-class	low
26	hsa-miR-3937	-0.0016977874777172684	0.0031805748558630747	0.24324324324324326	-0.5905679670898752	anti-class	low
48	hsa-miR-5196-5p	0.0016888161860586584	0.0043187601861377125	0.6148648648648649	1.009707223726697	pro-class	high
40	hsa-miR-4734	-0.0016145116066542701	0.005768183915825981	0.35135135135135137	-1.2382792355219963	anti-class	low
51	hsa-miR-5787	0.0015044805612350525	0.005955225501346829	0.6216216216216216	0.30345303676793106	pro-class	high
50	hsa-miR-557	-0.0013776356728777092	0.0021118352990586184	0.25	-0.23202197323295792	anti-class	low

miRNA behavior: **NC class**

	miRNA	mean_weight	std_weight	positive_ratio	mean_raw	direction	level_vs_avg
87	hsa-miR-7704	-0.009493698908527621	0.008630809917318813	0.14814814814814814	-1.0860131564950821	anti-class	low
61	hsa-miR-6729-5p	-0.006854652565146517	0.011401637817560284	0.2962962962962963	-0.2580471864952984	anti-class	low
18	hsa-miR-3187-5p	0.005383880564298723	0.0036482417809702112	0.9629629629629629	-0.0373804639818614	pro-class	low
8	hsa-miR-1908-5p	0.005309776549170138	0.00751482894137399	0.8518518518518519	0.5037096746740907	pro-class	high
24	hsa-miR-371a-5p	0.00529521147206174	0.007580141587175979	0.7777777777777778	0.41195219005574224	pro-class	high
40	hsa-miR-4734	-0.00430354218637012	0.0065273434706966055	0.25925925925925924	-0.9968942629297947	anti-class	low
51	hsa-miR-5787	0.003693659659411346	0.006396170650358696	0.6666666666666666	0.13198493177358908	pro-class	high
1	hsa-miR-1237-5p	0.003651341778599176	0.005525834462808761	0.8518518518518519	-0.29984877784275205	pro-class	low
22	hsa-miR-3663-3p	-0.003441508304852742	0.004330847875828473	0.25925925925925924	-0.2774664146901906	anti-class	low
89	hsa-miR-8072	-0.0030040544218789063	0.004169926870930897	0.25925925925925924	-0.4372135229424948	anti-class	low

Influence of miRNAs in predictions: how to interpret example results

miR-1908-5p and **miR-6780a-3p** consistently act as **pro-AD** markers, characterized by a positive mean weight and a high positive ratio (*mean_weight > 0.004*, *positive_ratio > 0.7*).

miR-3187-5p emerges as a strong **pro-NC** marker (*positive_ratio = 0.96*), with a neutral role in the other two classes.

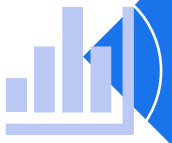
miR-7704 shows a **negative association with both AD and NC**, with a consistently negative mean weight (*mean_weight < -0.01*), suggesting a potential anti-class effect.

Conclusions: study summary

1. Development of a pipeline to link model features to miRNA expressions for Alzheimer's classification
2. Used Lasso regression to build a sparse, interpretable mapping from miRNAs to genetic features
3. Trained a new Random Forest model to enable explainability analysis
4. Applied SHAP to trace predictions back to relevant miRNAs
5. Identified recurring miRNA patterns associated with the three classes
6. Contribution: a reproducible framework that enhances model transparency and supports biomarker discovery

Future Works

- The applied methodology is model-agnostic, compatible with any black-box classifier
  .
- Possible future works involve the following aspects:



Extend analysis, using new datasets

- Larger
- More diverse

Explore more powerful models

- Also different from Random Forest
- Focused on early-stage detection

Integrate biological prior knowledge

- To provide a better selection on miRNA expressions
- Focus on biologically plausible and statistically relevant expressions

Thanks for the attention!

Also, thanks to Dr. Veronica Buttaro for the support!

EMANUELE TANZI

E.TANZI2@STUDENTI.UNIBA.IT