

THE MEDICAL INFORMATICS PLATFORM

MIP USER MANUAL

Table of Contents

1. Introduction.....	2
2. MIP User Guidelines	2
2.1 General Navigation with the MIP	2
2.2 Variables Exploration, Analysis Model and Experiment Design	3
2.2.1 EE: Epidemiological Exploration.....	3
2.2.2 IA: Interactive Analysis.....	5
2.2.3 BSD: Biological Signatures of Diseases.....	6
2.2.4 Online Resources	8
3. Other MIP Functionalities	8
3.1.1 Writing Articles	8
3.1.2 Accessing my Saved Articles and Models	8
3.1.3 Third Party Applications	8
4. Glossary	11

1. Introduction

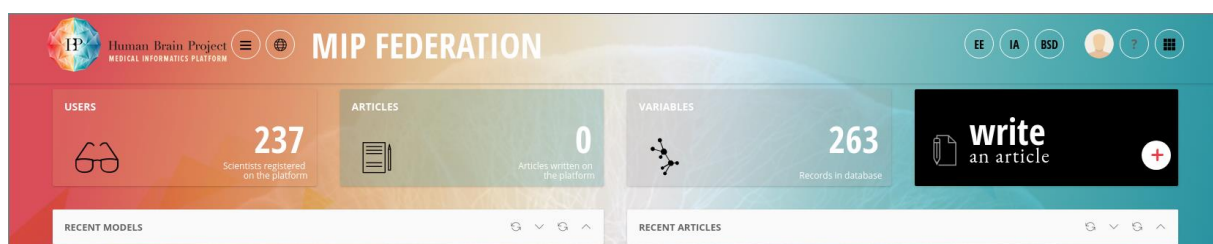
The Medical Informatics Platform (MIP) aims to provide tools to analyse medical data and advance more rapidly in understanding the neurological and psychiatric diseases. The users can access the platform via a Web user interface (UI) (password restricted) where they can run exploratory data analyses, create and share analysis models, execute descriptive statistics, inferential statistics and machine-learning algorithms on user-defined analysis models, as well as collaboratively write articles.

2. MIP User Guidelines

2.1 General Navigation with the MIP

After log-in, the platform opens with the main dashboard.

The dashboard shows a summary of statistics, users, available variables, as well as the latest three saved analysis models and articles (those of the current user or shared among all users). From here, the user may also start writing articles (description of data analyses performed).



At any time the user can return to this page by clicking on the HBP logo on the top left corner.

From the top banner, user can at any time access:

- **My Data:** personal dashboard displaying own work (saved analysis models and articles)
- **My Community:** all work labelled for sharing by any user within its MIP community
- **Functionalities:**
 - Epidemiological Exploration (EE)
 - Interactive Analyses (IA)
 - Biological Signatures of Diseases (BSD)
 - Personal Profile
 - Third-party web applications



2.2 Variables Exploration, Analysis Model and Experiment Design

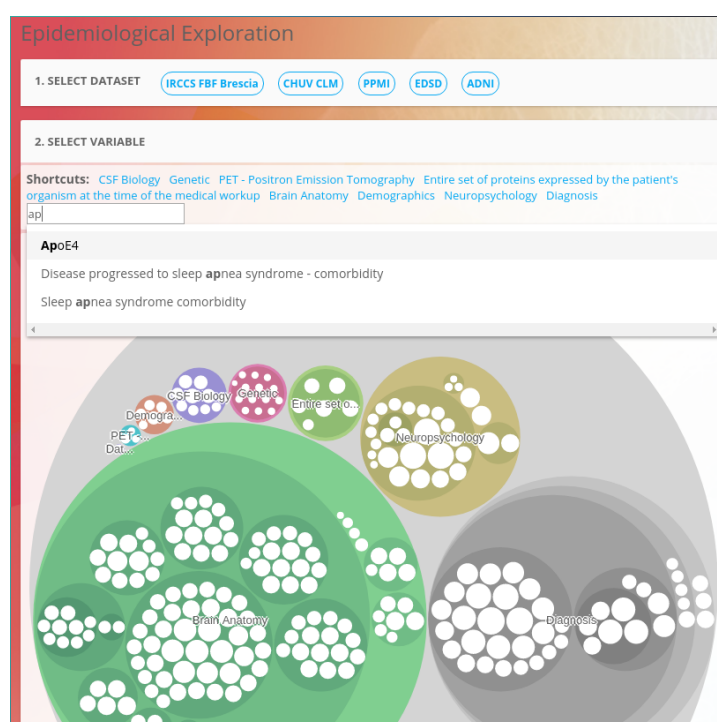
The MIP provides the following functionalities:

- **EE**: Epidemiological Exploration - allows exploration of the available variables, including visualisation of variables' types, selection of variables, visualisation of variables' descriptive information, and definition of analysis models (selection of response and explanatory variables for data analysis);
- **IA**: Interactive Analyses - provides descriptive summary statistics for the defined analysis models, in tabular and graphical formats;
- **BSD**: Biological Signatures of Disease - provides selection and configuration of data analysis algorithms - descriptive and inferential statistics, machine-learning and validation, using analysis models defined with IA functionality.

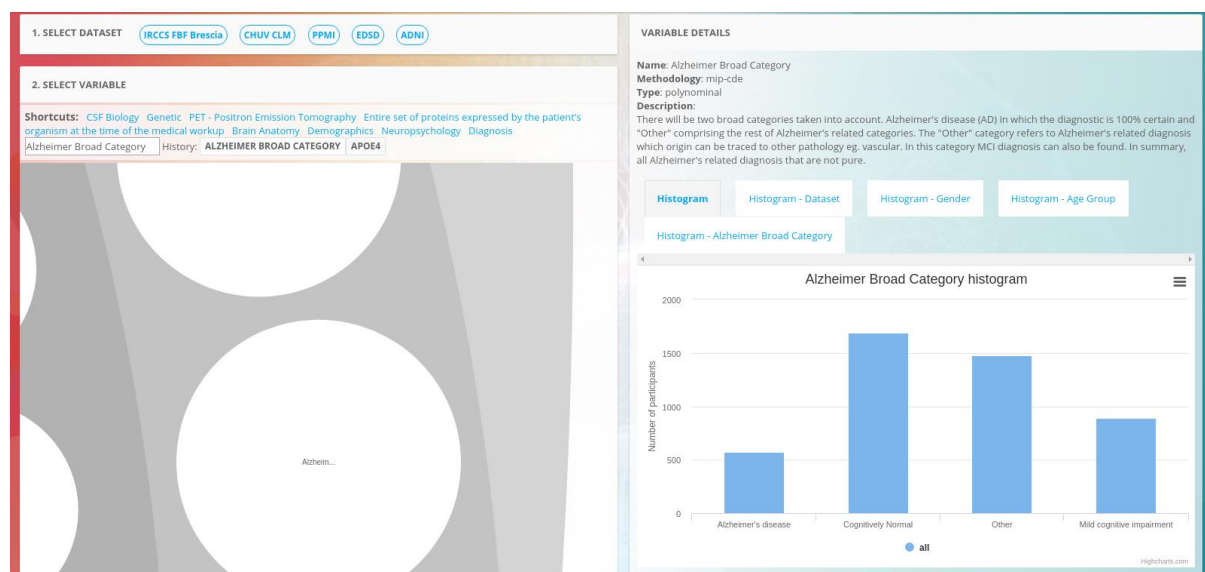
2.2.1 EE: Epidemiological Exploration

The exploration of variables is done through their representation in a circle-pack design. We will call it the MIP Variable Space.

User may also search for a variable by typing its name directly in the search box above the MIP Variable Space (see screen shot below, searching for APOE4 as an example)



Some variables have descriptive information associated with it. They are displayed in the right-sided panel of the MIP Variable Space (see next screenshot).



User may stop variables exploration here, or continue, to define an analysis model by selecting variables from the MIP Variable Space, as variable (dependent variable), co-variables (independent variables) and filters. It is possible to select a group of variables in one step. See screenshot below.

Variables used for setting up grouping and conditions should be selected as filter variables.

3. ADD VARIABLES TO MODEL



2.2.1.1 Define an Analysis Model

To define an analysis model, user needs to:

1. Search for the variables of interest in the MIP Variable Space, and click on the desired variable, or a group of variables.

Let's take an example: to predict changes in the volume of the Hippocampus in Alzheimer's disease, with respect to age and gender:

- Select the variable Left Hippocampus in the MIP Variable Space and then click on "+ AS VARIABLE" in the model table to define it as a response variable.
- Then select Age Years and Gender variables in EE and click "+ AS COVARIABLES", to define them as explanatory variables. Then do the same for Alzheimer Broad Category. Variables will be automatically classified as "NOMINAL" or "CONTINUOUS" depending on their respective types.

User may at any time remove variables from the Configuration table by clicking the “X” sign. Similarly, user may restore any previously searched variable from the “History” line.



- When finished with analysis model definition, just click on the Review Model button at the bottom of the screen.

2.2.2 IA: Interactive Analysis

The statistical analysis of the model is run in the MIP. The results are summarized in appropriate tables and visualizations. Statistical description of the defined analysis model is provided in a table and in appropriate visualisations.

Model

Filter variable

Variable

lethippocampus

CoVariables

alzheimerbroadcategory

gender

subjectageyears

Datasets

IRCCS FBF Brescia

CHUV CLM

PPMI

EDSD

ADNI

Table

Boxplot

Heatmap

VARIABLES	IRCCS FBF BRESCIA	CHUV CLM	PPMI	EDSD	ADNI
Demographics					
Gender	1960	699	714	474	1066
F	1194	355	256	247	485
M	766	344	458	227	581
Age Years	58.26 (20.00-92.00), std: 17.13	71.17 (45.00-93.00), std: 10.69	61.93 (31.00-85.00), std: 10.19	70.04 (36.00-87.00), std: 9.00	73.26 (55.00-90.00), std: 7.36
Diagnosis					
Alzheimer Broad Category	1960 (Null count: 176)	699	714	474 (Null count: 106)	1066
CN	1240	32	-	151	268
MCI	201	36	-	76	576
Other	192	572	714	-	-
AD	151	59	-	141	222
Limbic					
Left Hippocampus	2.98 (0.58-4.07), std: 0.38	2.91 (0.86-4.47), std: 0.38	3.21 (2.37-4.18), std: 0.29	2.99 (1.30-4.45), std: 0.36	2.93 (1.82-3.97), std: 0.35

User can explore the defined analysis model by selecting/deselecting datasets, adding conditions (filters), and selecting visualisations.

To add conditions and grouping, click on the “filter variable” link on the left. That opens a pop-up window for configuring individual rules, their grouping and for selecting their combinations using “and/or” buttons. Conditional and grouping rules are applicable only to variables selected as “filter variables” in the EE functionality. See screenshot below.

Configure filtering query

AND OR

Alzheimer Broad Category

equal

Alzheimer's disease

DELETE

+ ADD RULE

+ ADD GROUP

SET FILTER

CLOSE

After setting up grouping and conditional rules, the analysis model's data is updated. Corresponding tabular and visual summary statistics reflect the analysis model changes.

To perform an experiment, user needs to save the analysis model by giving it a name. He can also share it with the MIP community.

YOUR MODEL'S TITLE

My Model

☐ Share this model

SAVE

↓

2.2.3 BSD: Biological Signatures of Diseases

To configure the experiment:

- 1- Click "Run Machine Learning Experiment", on the IA screen.

RUN MACHINE LEARNING EXPERIMENT

- 2- In the next BSD screen, choose a method among the "Available Methods", see screenshot below.

Run an Experiment on **alz-hippocampus** RELATED EXPERIMENTS ▾

Available Methods

JSI HEDWIG

DISTRIBUTED K-MEAN

Predictive Model ▾

SGD NEURAL NETWORK

SGD LINEAR MODEL

NAIVE BAYES

K-NEAREST NEIGHBOR

GRADIENT BOOSTING

SGD Neural Network method ▾

Your Experiment ▾

Experiment Parameters ▾

SGD Neural Network

Hidden layer sizes:

The ith element represents the number of neurons in the ith hidden layer. Pass integers separated by comma.

Activation:

Activation function for the hidden layer.

Alpha:

L2 penalty (regularization term) parameter.

Initial learning rate:

The initial learning rate used. It controls the step-size in updating the weights.

ADD METHOD

Model ▾

Select model

Variable

lefthippocampus

CoVariables

alzheimerbroadcategory
gender
subjectageyears

Filters

- 3- For some methods user can configure parameters of the algorithm.
- 4- Click on “Add method” to add it to the methods selected for the experiment. Several methods can be added to an experiment.
- 5- User can choose to train and validate the experiment on various datasets. See screenshot below.

Training and validation ▾

Training

☒ IRCCS FBF Brescia
☒ CHUV CLM
☒ PPMI
☐ EDSD
☐ ADNI

Validation

☐ IRCCS FBF Brescia
☐ CHUV CLM
☐ PPMI
☒ EDSD
☒ ADNI

- 6- When ready to run the experiment, user needs to give it a name and then to click on RUN EXPERIMENT.

The results of the experiment are presented in textual, tabular or visual format depending on the type of methods chosen and their implementation.

2.2.4 Online Resources

A video demoing the MIP is available on YouTube:

<https://www.youtube.com/watch?v=MNWExzouMJw&t=61s>

3. Other MIP Functionalities

3.1.1 Writing Articles

- 1- Go to main MIP Dashboard (click the HBP icon on the top left-hand side);
- 2- Click on “Write an Article”;
- 3- Write the article: use the editor to add a title, abstract and content to your article;
- 4- You can also drag and drop results of your models (or others’ if are shared) from the left-hand side into the content of your article;
- 5- Give a name to your article, save it and, optionally, share it.

3.1.2 Accessing my Saved Articles and Models

User can at any time access the already written articles and saved analysis models, via:

- My Data - all own work, shared or unshared;
- My Community - all work shared within the MIP community;

User can preview articles, save them to a file system accessible from his computer or open them for re-editing.

3.1.3 Third Party Applications

3rd Party Application are made by users to provide some insights on specific models or visualisations.

3.1.3.1 3D Biological Rules:

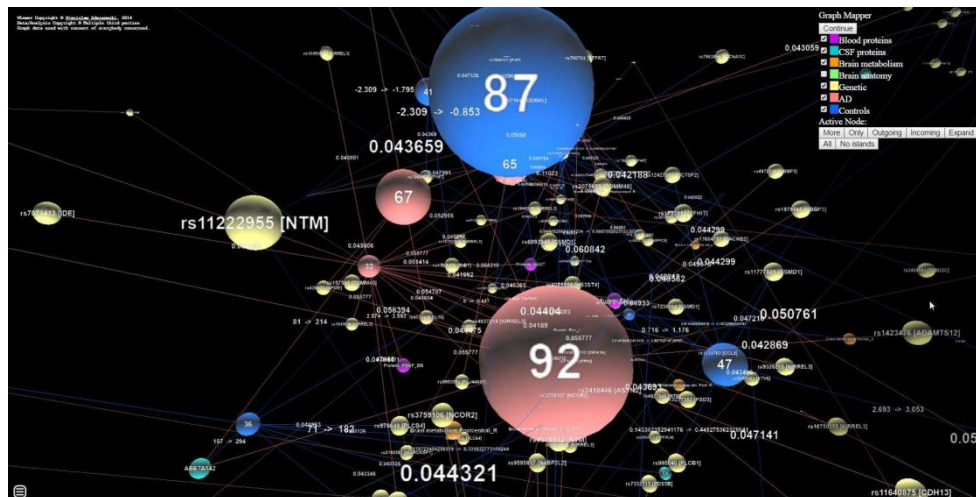
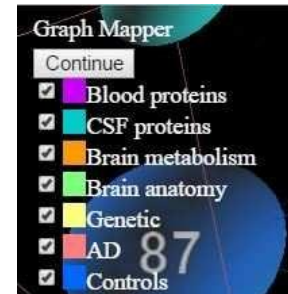
Navigate in a 3D world of variables.

The 3D application shows the results of a rule-based clustering algorithm applied to Alzheimer's disease patients to identify homogeneous subgroups of patients. The hypothesis is that patients in each subgroup have the same underlying cause of the disease. The rule-based algorithm aims to explain the variability between individuals and describes a population by a group of “local over-densities”.

These are defined as subspaces over combinations of variables: blood proteins (magenta), CSF proteins (aqua-blue), Brain metabolism (orange), brain anatomy (green) and genetic (yellow).

The red spheres represent AD subgroups and the blue ones healthy controls. The number in a sphere indicates the number of subjects belonging to the subgroup. The edges show rules between the spheres and variables that define each subgroup.

User can also use the left mouse button for rotation, the middle button for zoom and the right mouse button for translation.



User can select and deselect the different variables clicking on the variables tick to make them appear or disappear in the 3D world.

For the genetic variable click “More” to get additional information from the <http://www.ensembl.org/> database.

3.1.3.2 2D Biological Rules

A graphical view of biological rules.

Like the 3D application, 2D application shows the results of a rule-based clustering algorithm applied to Alzheimer's disease patients to identify homogeneous subgroups of patients. The hypothesis is that patients in each subgroup have the same underlying cause of the disease. The rule-based algorithm aims to explain the variability between individuals and describes a population by a group of “local over-densities”.

These are defined as subspaces over combinations of variables: genetic (yellow), brain anatomy (green), brain metabolism (orange), CSF proteins (aqua-blue) and blood proteins (magenta).

In this application user can select a variable on the 2D Map and get information on the right side.



2D Biological Rules

Network Legends

- Genetic
- Brain Anatomy
- Brain Metabolism
- AD
- Control
- CSF Protein
- Blood Protein

Information

Name:
 Volume:RightCerebralCortex
 Category: Brain Anatomy
 More...

Key Nodes and Associations:

- Node 31 (Blue):**
 - Volume:RightTransverse
 - rs78505 [RSP04]
 - rs13148572 [PDS5A]
 - rs12503572 [PDS5A]
 - rs2267779 [GRIN2A]
 - rs1362321 [GRIN2A]
 - rs1806201 [GRIN2B]
 - rs2643350 [PCSK6]
 - rs9953092 [FLJ44881]
 - rs7671946 [DDX60L]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]
 - rs9953092 [FLJ44881]
 - rs7671946 [DDX60L]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]
- Node 47 (Blue):**
 - Protein:FA2
 - rs2075650 [TOMM40]
 - rs1777685 [CSMD1]
 - rs2255953 [ETV6]
 - rs5993945 [CSMD1]
 - rs12427568 [CCT5P2]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]
 - rs9953092 [FLJ44881]
 - rs7671946 [DDX60L]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]
- Node 65 (Red):**
 - Volume:RightBanksts
 - rs1806201 [GRIN2B]
 - rs2643350 [PCSK6]
 - rs9953092 [FLJ44881]
 - rs7671946 [DDX60L]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]
 - rs9953092 [FLJ44881]
 - rs7671946 [DDX60L]
 - rs157580 [TOMM40]
 - rs12509086 [DDX60L]
 - rs3815501 [BZW1]

4. Glossary

Model: analysis models: Set of variables, co-variables, filters, training and validation datasets.

Epidemiological Exploration (EE): Exploration of variables, their distribution related to datasets, genders, age groups and diagnosis.

Interactive Analyses (IA): Summary statistics of the selected model by dataset.

Biological Signatures of Diseases (BSD): Statistical and machine learning algorithms.

Variable: Dependent or target variable.

Co-variable : Independent variable, or predictor.

- **Nominal:** Categorical type.
- **Continuous:** Numerical (real, integer).

Filter: Include or exclude specifics subjects.

Dataset: Set of patients or subjects related data provided by specific medical institutes.

Federation: Set of algorithms designed to aggregate intermediate results from different datasets.

Distributed: Set of algorithms performed either on different datasets and outputting parallel results, or predictive models trained iteratively on different datasets.

Experiment: Set of algorithms applied on a model.

Methods: Algorithms.