YouForAll

ALTA SCUOLA
POLITECNICA
POLITECNICO DI MILANO I POLITECNICO DI TORINO

near

User Manual

Contents

About NEAR

NEAR fits in the Alta Scuola Politecnica XVII cycle multidisciplinary project called YouForAll: Your digital twin for aLLowing a healthy society (Y4A). Y4A is born from a collaboration of ASP (Alta Scuola Politecnica) and Dedalus Italia s.p.a (the leading provider of health and diagnostic software in Europe and one of the largest in the world) with two main aims:

- Model a digital twin paradigm for each citizen, profiling the data, through AI methodologies, based on specific health targets
- Define cloud-based solutions, for allowing the seamless and automatic integration among the personal digital twin management and the clinical information systems, so that the institutional health systems can guarantee citizens with targeted prevention, early enrolment on and the activation of personalized care processes.

NEAR is a Data Driven predictive model derived from the training of Artificial Neural Networks (ANN).
The aim of NEAR model is to reshape the knowledge of a Neural Network in an explainable model which can be interpreted and evaluated from doctors and clinician.
Unlike classic ANN models, NEAR does not require all the input parameters used for training. This flexibility has the price of a loss of reliability

in the calculated score proportional to the number of missing values. Thanks to its interpretability combined with the power of the data driven approach and graceful degradation, this model aims to pave the way for the Digital Twin in healthcare.

The Group

Four students compose the team:

- **Kassem Karim:** the Team Leader and a biomedical engineer from the Politecnico di Milano.
- **Cavallo Andrea:** an informatics engineer from the Politecnico di Torino
- **Fassino Davide:** a mathematical engineer from the Politecnico di Torino
- **Vergani Andrea Mario:** an informatics engineer from the Politecnico di Milano

The students worked under the supervision of the academic tutors:

- **Prof. Della Valle Emanuele:** informatics professor at the Politecnico di Milano.
- **Prof. Deriu Marco:** biomedical professor at the Politecnico di Torino

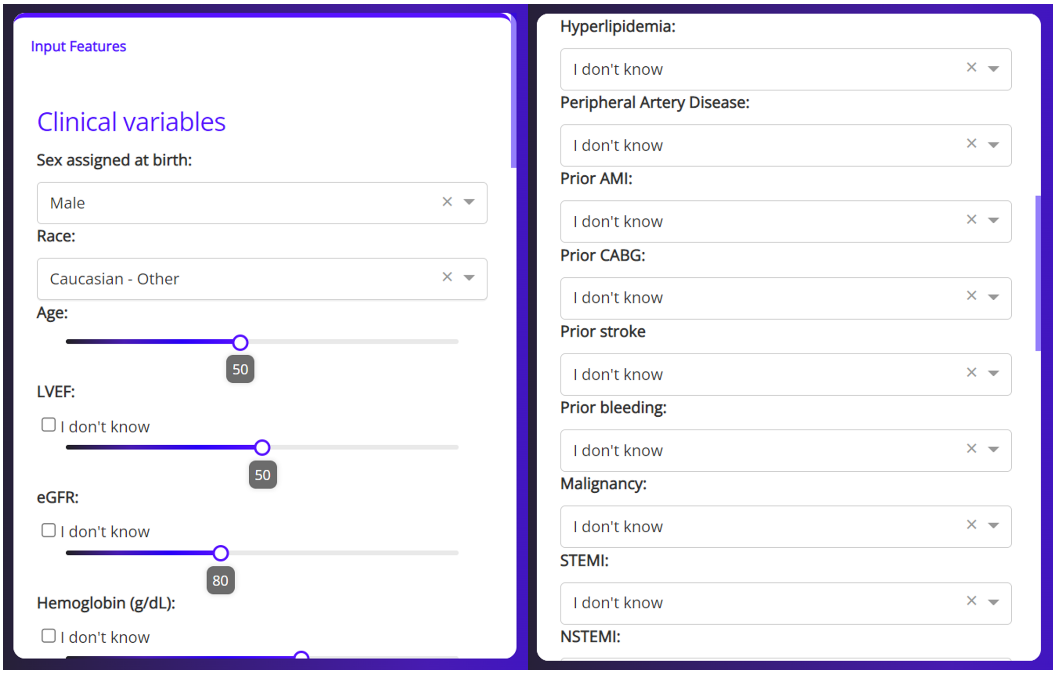
with the collaboration of:

- **Dedalus Italia s.p.a Architecture Team**

Chapter 1

Data Input

1.1 Single Patient Data Input



Input Features

Clinical variables

Sex assigned at birth:
Male

Race:
Caucasian - Other

Age:
50

LVEF:
□ I don't know
50

eGFR:
□ I don't know
80

Hemoglobin (g/dL):
□ I don't know

Hyperlipidemia:
I don't know

Peripheral Artery Disease:
I don't know

Prior AMI:
I don't know

Prior CABG:
I don't know

Prior stroke
I don't know

Prior bleeding:
I don't know

Malignancy:
I don't know

STEMI:
I don't know

NSTEMI:

Figure 1.1: On the left the "Input Box" first view with the selectors while on the right there is the "hypertension" dropdown selector open.

The “Input Features Box” (along with the “Upload File Box”) is one of the 2 possible methods to insert data into the App. In this frame all the variables are divided into 3 main categories accessible by using the blue scrolling bar on the right side of the box:

- Clinical
- Therapeutical
- Angiographic

If the input is numerical a selector allow the user to insert the correct value in the specific interval, otherwise in case of categorical features the input is a dropdown list with all the possible categories.
With the exception of only the first three features: Sex assigned at birth, Race and Age (used for the computation of creatinine from eGFR) all other inputs are not mandatory and can be disabled by selecting the option “I don’t know” in the dropdown list or above the slider [Fig. 1.2].
The data inserted into this box will be processed by the NEAR models and the results will be shown in the "Analysis Box" (section chapter 2) and if there are enough data in the "PDTA Box" (for more information look at section section 1.1)

1.2 Tabular Input

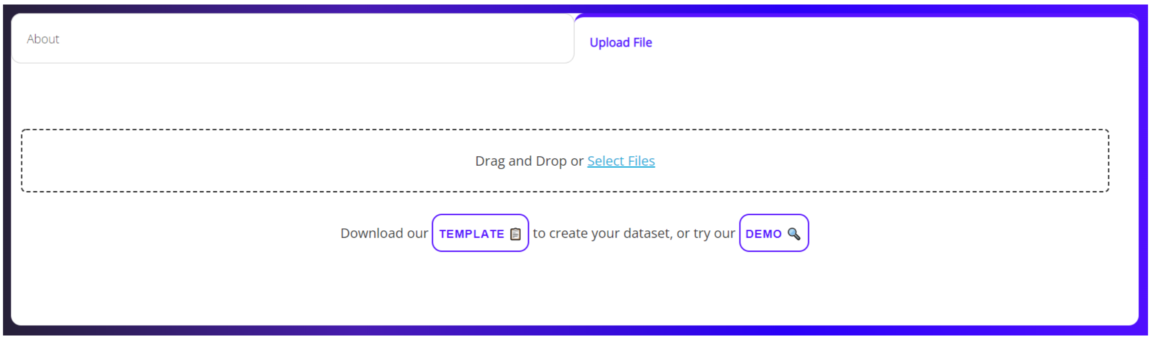
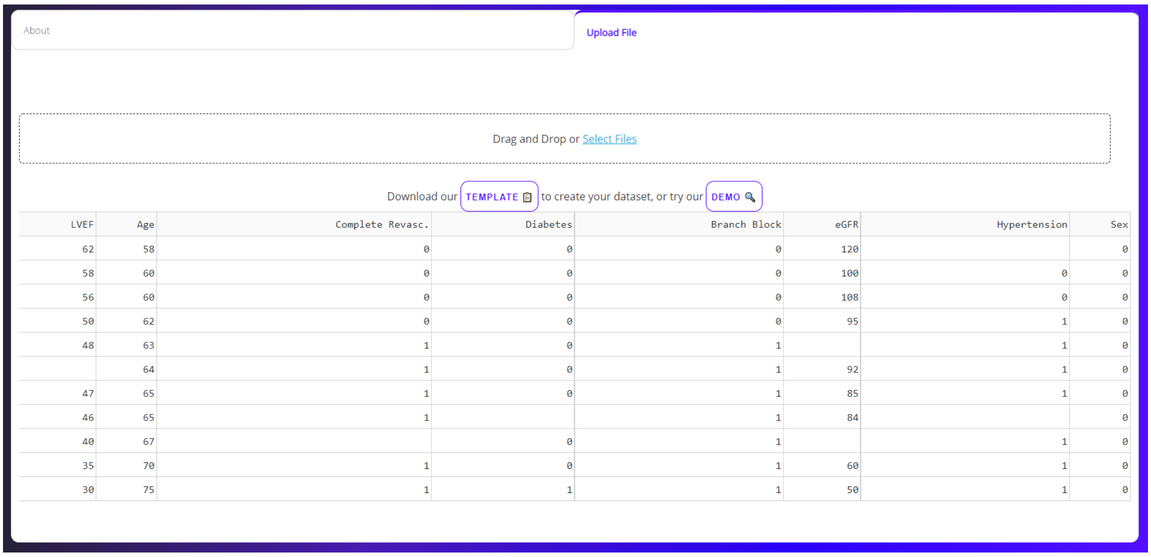

About
Upload File
Drag and Drop or Select Files
Download our TEMPLATE to create your dataset, or try our DEMO

Figure 1.2: The "Upload File" tab with the "Template" and the "Demo" button.

In the "Upload File" tab of the multiple input box it's possible to upload a table in which columns represent the patient features while the row the patient evolution in time (each row is a set of new values).
If the clinician want to generate a new dataset it's possible to download the template in .csv (Excel compatible file) using the button "Template" [Fig. 1.2)].
In case the clinician want to upload a pre-existing dataset it's simple as drag and drop the file in the upload region of the tab. Otherwise by clicking the "Demo" [Fig. 1.2)] button it's possible to upload a demo dataset with made-up data used for the plotting generation only. 1.3)]



Download our TEMPLATE to create your dataset, or try our DEMO

LVEF	Age	Complete Revasc.	Diabetes	Branch Block	eGFR	Hypertension	Sex
62	58	0	0	0	120		0
58	60	0	0	0	100	0	0
56	60	0	0	0	108	0	0
50	62	0	0	0	95	1	0
48	63	1	0	1		1	0
	64	1	0	1	92	1	0
47	65	1	0	1	85	1	0
46	65	1		1	84		0
40	67		0	1		1	0
35	70	1	0	1	60	1	0
30	75	1	1	1	50	1	0

Figure 1.3: Demo table with the made-up data

In the future version of the NEAR Web App the table will be also interactive!

For the temporal data analysis and graphs look at chapter chapter 4.

Chapter 2

Patient Data Analysis

The "Patient Data Analysis BOX" (positioned near to the input box on the right) is the core of NEAR, in this frame there is all the knowledge extracted from the model.
The BOX itself is divided into 2 different tabs: the **Model Bleeding** tab and the **Model Death** one.
Each tab refers to a different neural imputed model generated using the same patient population and has 3 sections:

- **Prediction**
- **Model's Risk Interpretation**
- **Model's Parameters**

The sections and the graphical representation of the data generated by NEAR are the same for both models hence the following explanations will use the Death Model as example but the same concepts are valid also for the Bleeding one.

2.0.1 Model Prediction

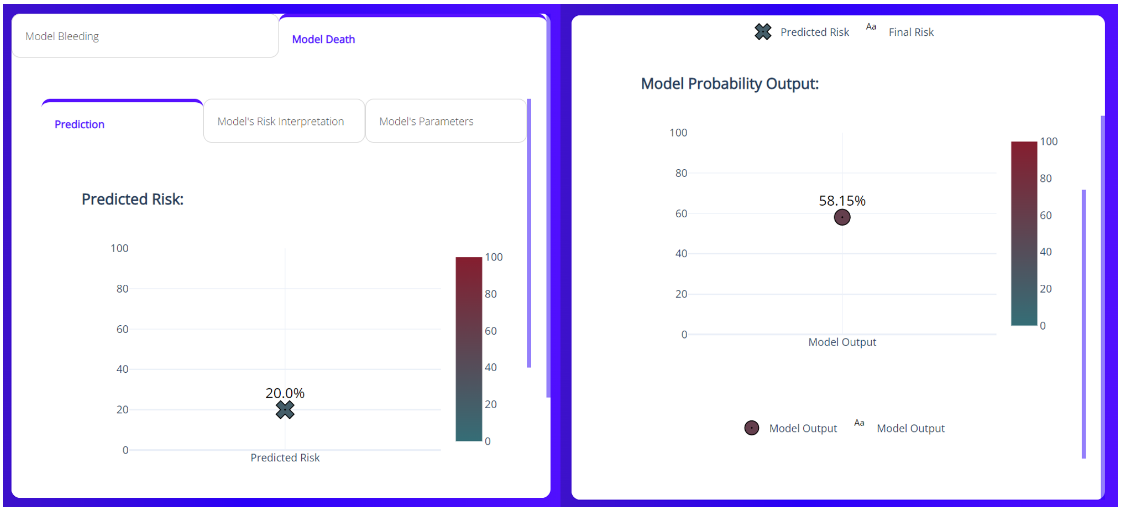


Figure 2.1: On the left the Calibrated Risk score for the event, on the right the raw prediction obtained from the model.

Despite the fact that the output of a machine learning algorithm (or in our case the Imputed Model output) being between 0 and 1 may resemble a probability, this value cannot be used as it is, since it may not express the true probability of the event. To obtain a calibrated score, it is necessary to select an appropriate regressor (or calibrator) that has to be fitted on the data. (For more information about the Probability Calibration problem look at our Paper –> ...)

For this reason in this section the app provides to the clinician both the calibrated risk and the model prediction.

The first plot visible to the clinician represent the real and calibrated risk of the event (in this case Death) computed by the model and based on the patients' population used to validate the Neural Network.

Scrolling down the section it's possible to look ad the probability generated by NEAR before the calibration. By changing the input values also

these 2 probabilities will change, however due to the isotonic regressor used for the calibrated score (left plot on Fig.), the plot will show a finite and discrete numbers of possible risk percentages while the model output (right plot on Fig.) will show a continuous range of values.
Technicalities aside with the first plot the clinician is able to asses the realistic risk of the event while with the second one is possible to observe in details if the model prediction move towards the event (100) or not.

2.0.2 Model's Risk Interpretation

This section is the natural continuation of the previous one. With the plots described below it's possible to deeply understand which feature contributes to the final prediction, if the contribution is positive (the variable increase the risk of the event) or negative (the feature behave like a protective factor) and with how magnitude.
Also some useful informations about the importance of each features inserted by the clinician are provided along with the reliability of the prediction provided by the model (based on the number of missing values and the importance given by the models).

Cumulative Risk

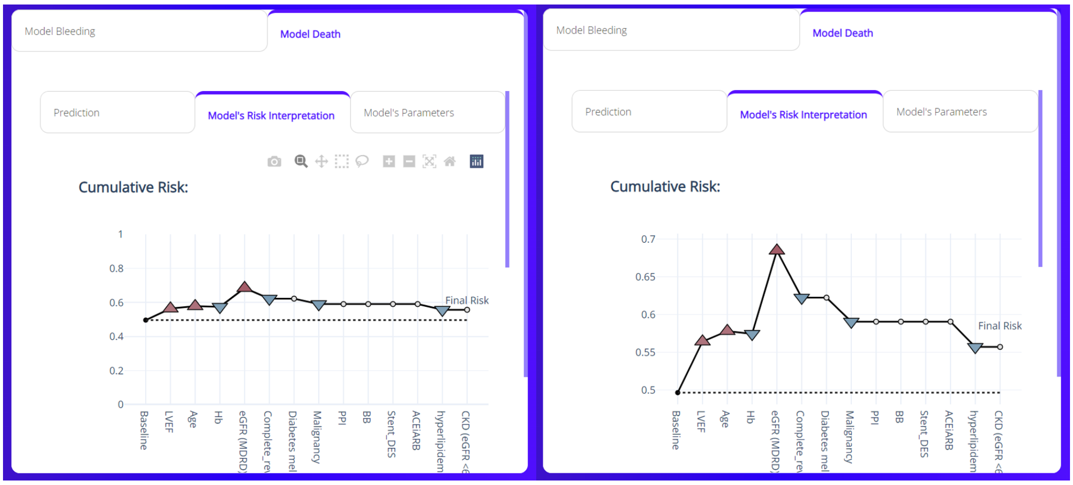


Figure 2.2: On the left the Cumulative Risk with the y-axis range 0-1, on the right the expanded axis after a double-click on the plot itself.

The cumulative plot show in detail how much each feature change the model prediction baseline (not the calibrated score). If the arrow is red the feature has a positive contribution while the blue arrow mean a negative one (positive = higher risk, negative = lower risk).

Also the magnitude of the contribution is represented by the slope of the previous segment associated with the feature in analysis. (The slope more evident in the expanded axis plot, right view Fig. 2.2).

The white dot are the unknown inputs (that did not contribute to the risk).

Features Contribution

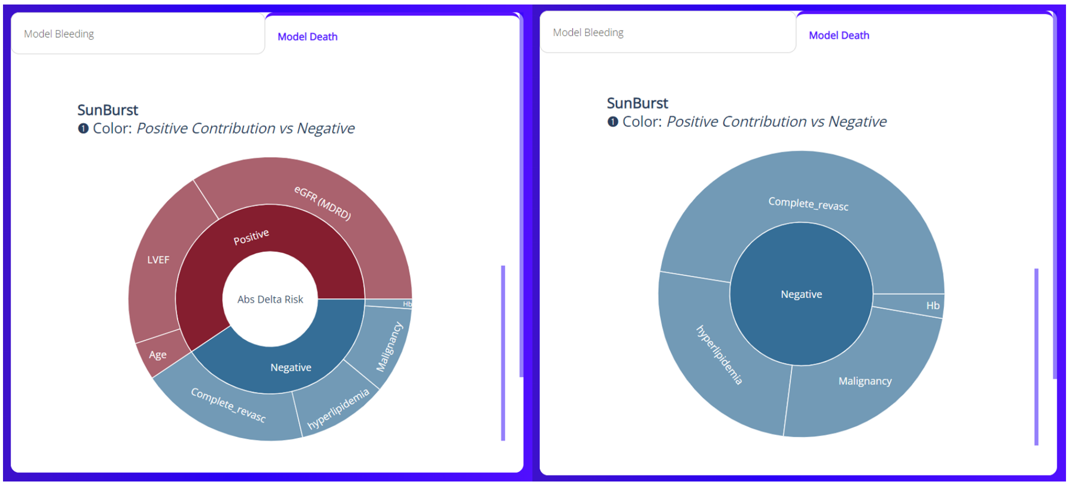


Figure 2.3: On the left the Features Contribution Plot with both positive and negative values. On the right the expanded view for the negative ones obtained by clicking on the "Negative" inner circular sector.

With this plot the contribution type and magnitude of each features is better represented. It's also possible to obtain the absolute contribution value by hovering over each sector. Also in this case it's possible to expand the view by clicking on the labels in the graph [Fig. 2.3]

Features Importance

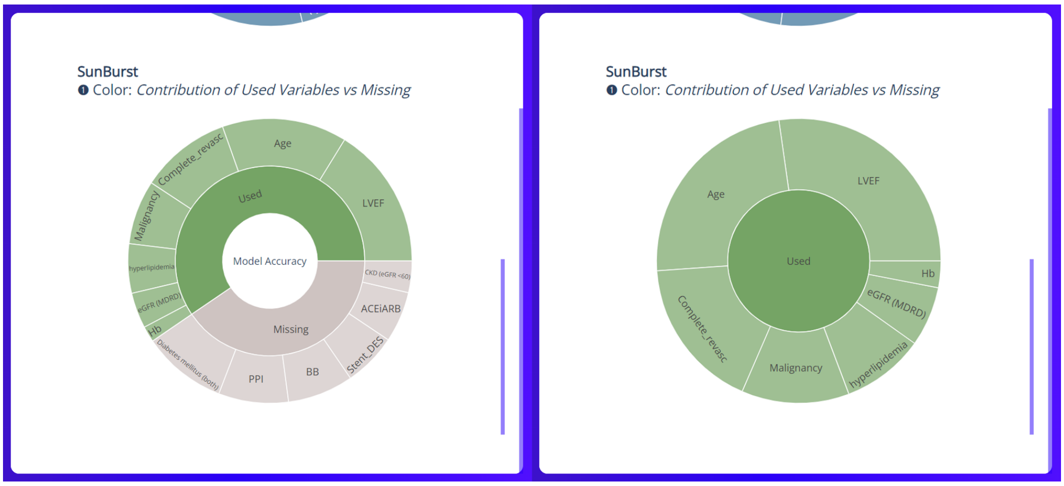


Figure 2.4: On the left the Features Importance Plot with both the "Used" and the "Missing" ones. On the right the expanded view for the used ones obtained by clicking on the "Used" inner circular sector.

Even if NEAR use more than 30 input features these are not mandatory. It's obvious that the more the inputs variable the more accurate the prediction is. Also the reliability of the prediction is associated on the importance given by the model to the features.

This plot show in gray all the features that can be used but are missing and in green the given inputs. The dimension of each sector is associated with the relevance of the feature in the prediction [Fig. 2.4].

With this plot the clinician can verify the reliability of the prediction (maximum reliability with all the features used) and understand which feature has to be included to improve the prediction (based on the variable importance).

2.0.3 Model's Parameters

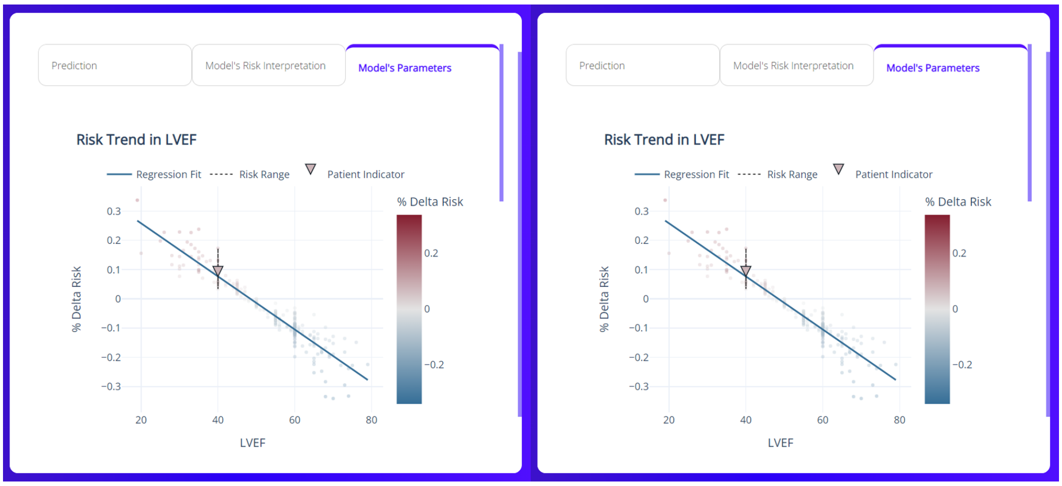


Figure 2.5: On the left the Partial Dependence Plot with the patient indicator (which color associate the patient risk) and the risk range of the population with the same value of the patient. On the right... TO DO THE BOX PLOTS

As explained in the NEAR Paper, each features delta risk (represented in the previous plots) is generated from a set of coefficients obtained by fitting a scatter plot named "Partial Dependence Plot" or looking at the median of the delta risk computed over a patients population. The partial dependence plot has on the x-axis the values of the numerical feature (Age, eGFR, LVEF...) and on the y-axis the delta risk assigned from the Neural Model to the specific patient (each dot in the plot is a patient from the cohort).

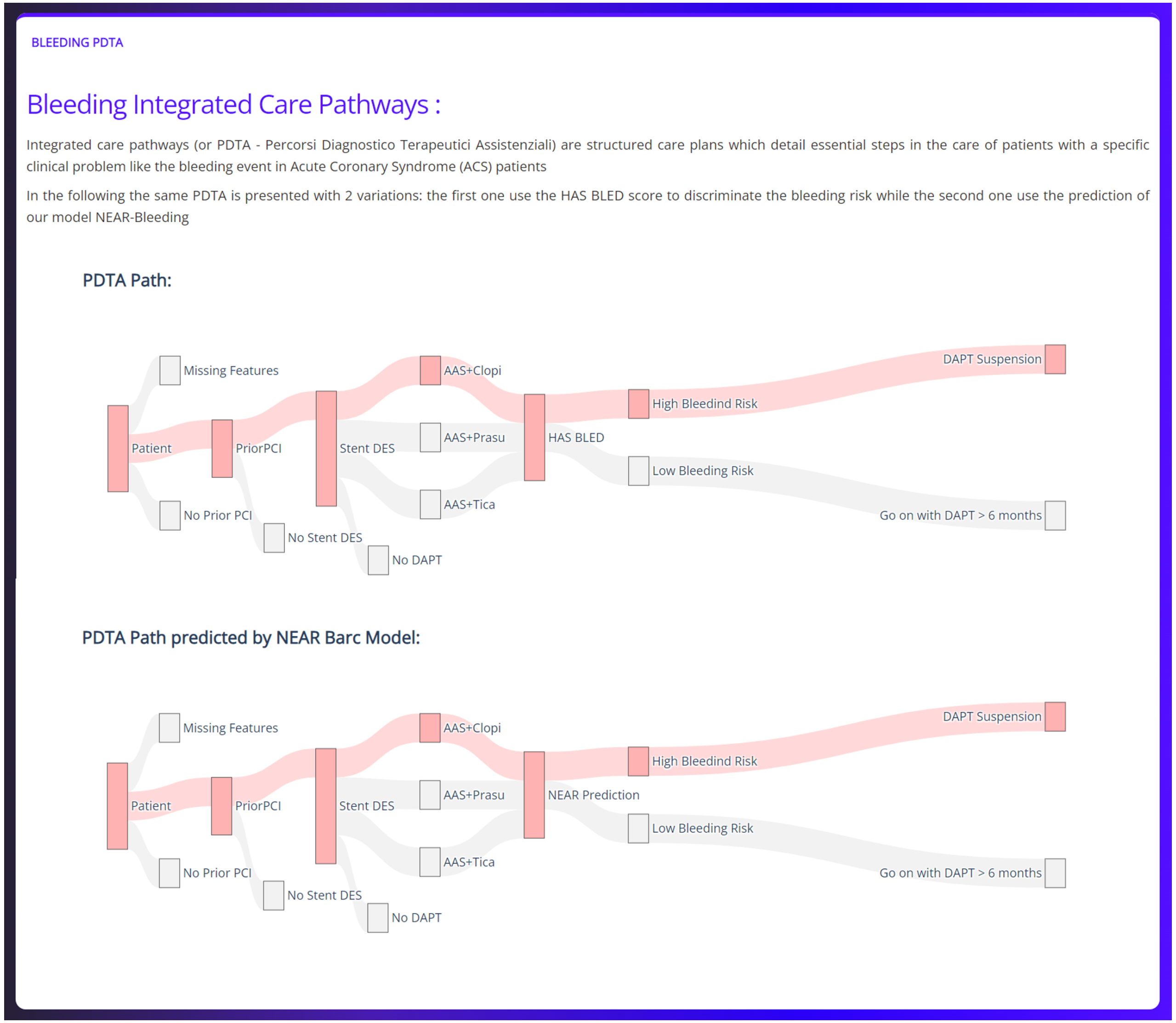
From this plot the clinician is able to retrieve useful information about the coefficients used for the risk model and also how the model relate a certain numerical feature to the event (linear relation, exponential, parabolic...)2.5].

For the categorical features (Yes / No) instead of a scatter plot 2 boxplots are used, one for each possible choice.

Chapter 3

PDTA Box

To DO



BLEEDING PDTA

Bleeding Integrated Care Pathways :

Integrated care pathways (or PDTA - Percorsi Diagnostico Terapeutici Assistenziali) are structured care plans which detail essential steps in the care of patients with a specific clinical problem like the bleeding event in Acute Coronary Syndrome (ACS) patients

In the following the same PDTA is presented with 2 variations: the first one use the HAS BLED score to discriminate the bleeding risk while the second one use the prediction of our model NEAR-Bleeding

PDTA Path:

PDTA Path predicted by NEAR Barc Model:

Figure 3.1: PDTA Complete

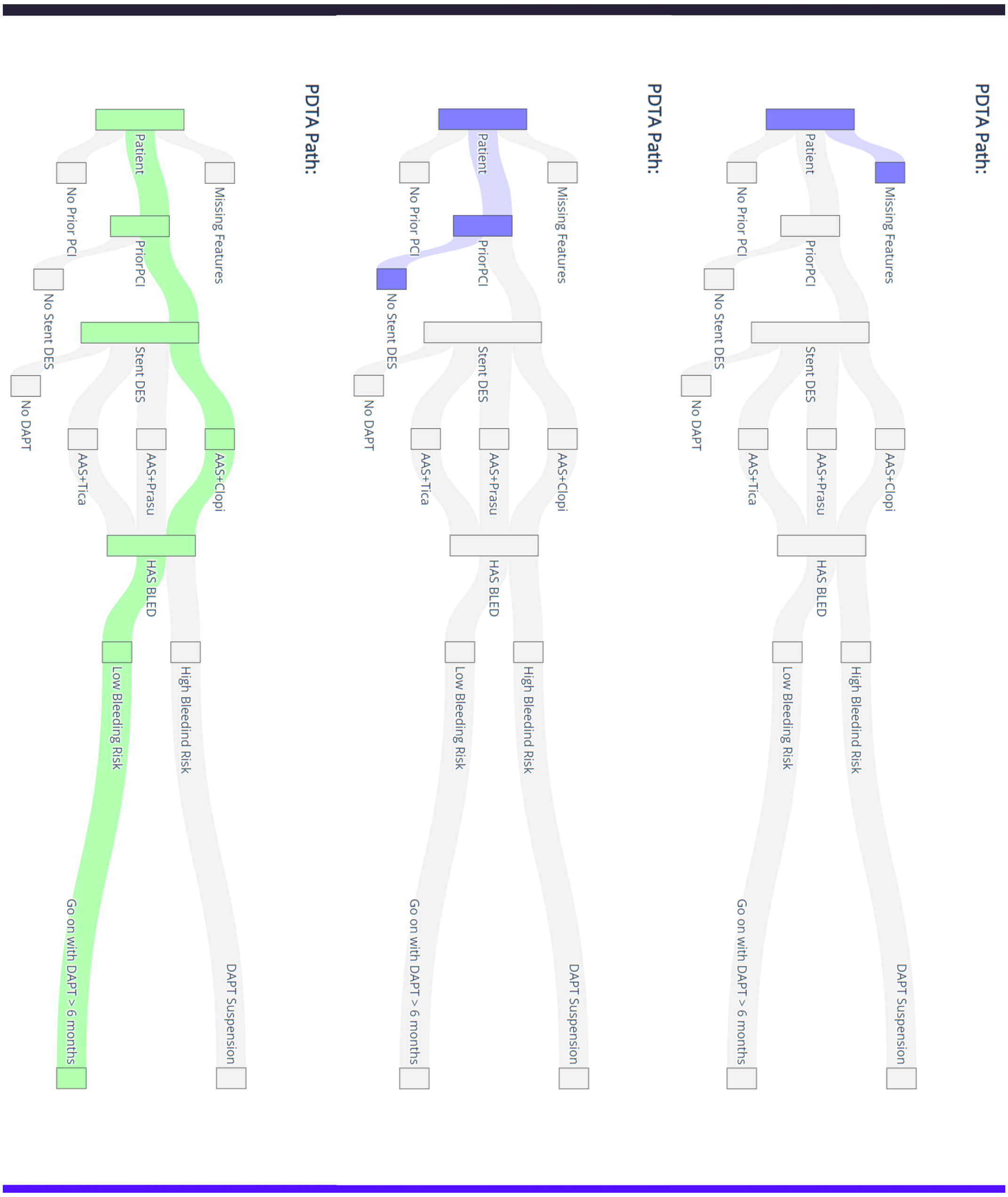


Figure 3.2: PDTA Examples

Chapter 4

Temporal Evolution