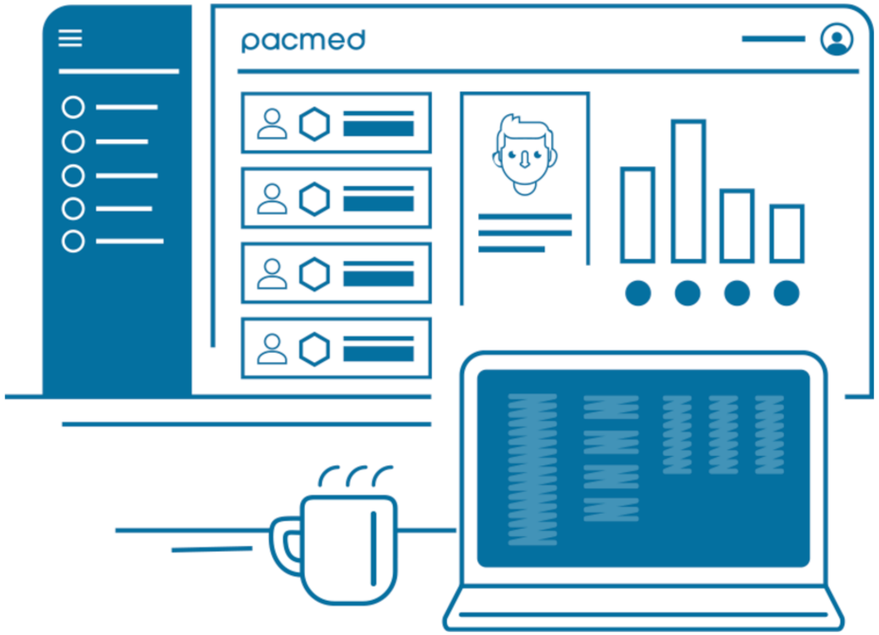
pacmed

ECP

Platform voor de
InformatieSamenleving



Interpretability of machine learning on the intensive care

SHAP and beyond

david.ruhe@gmail.com

Personal and precise healthcare through machine learning

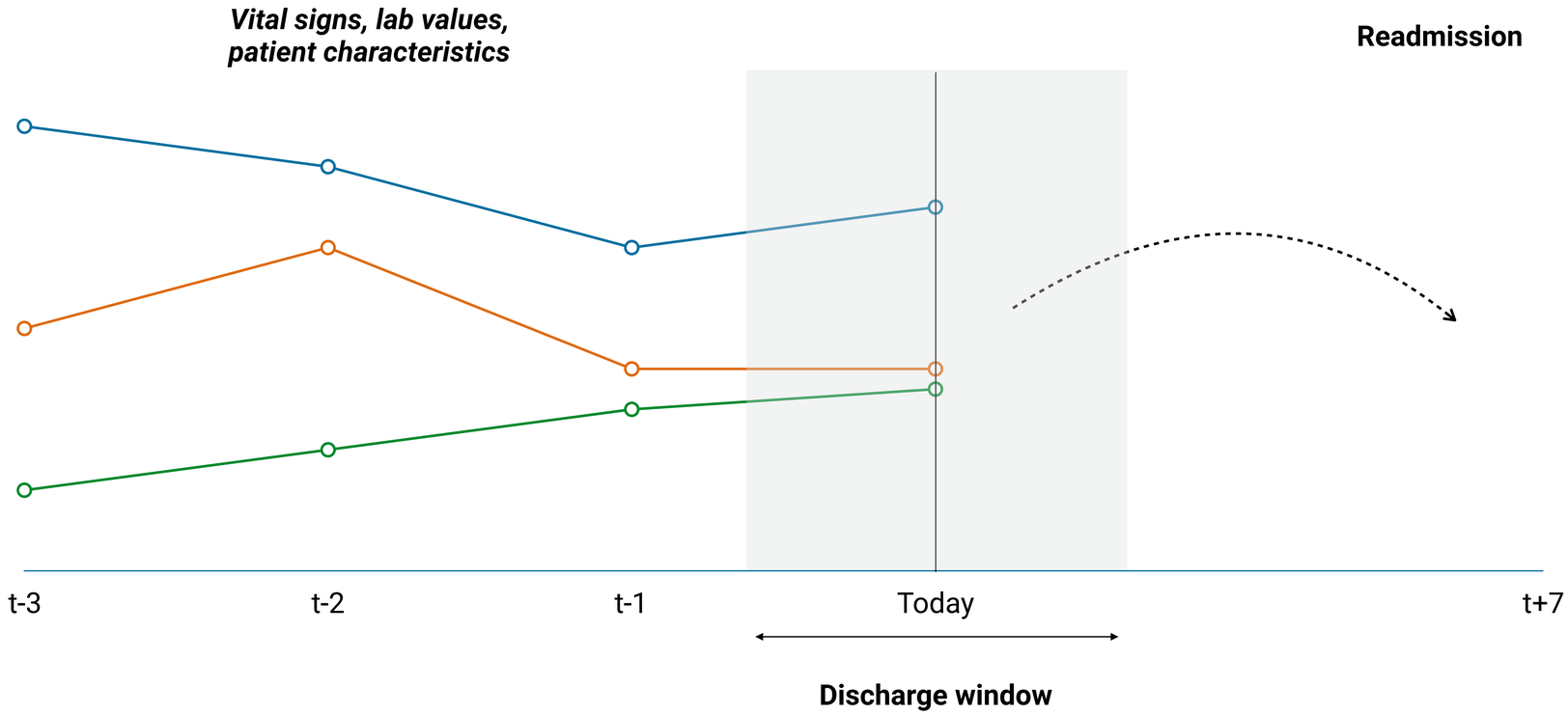
pacmed

Patients only get care that has the **highest probability** of **succes** for them

Facilitate a **continuous learning system** where each decision can be used to improve the next

Spend scarce **resources** as **efficiently** as possible

Use case: predicting intensive care readmission risk

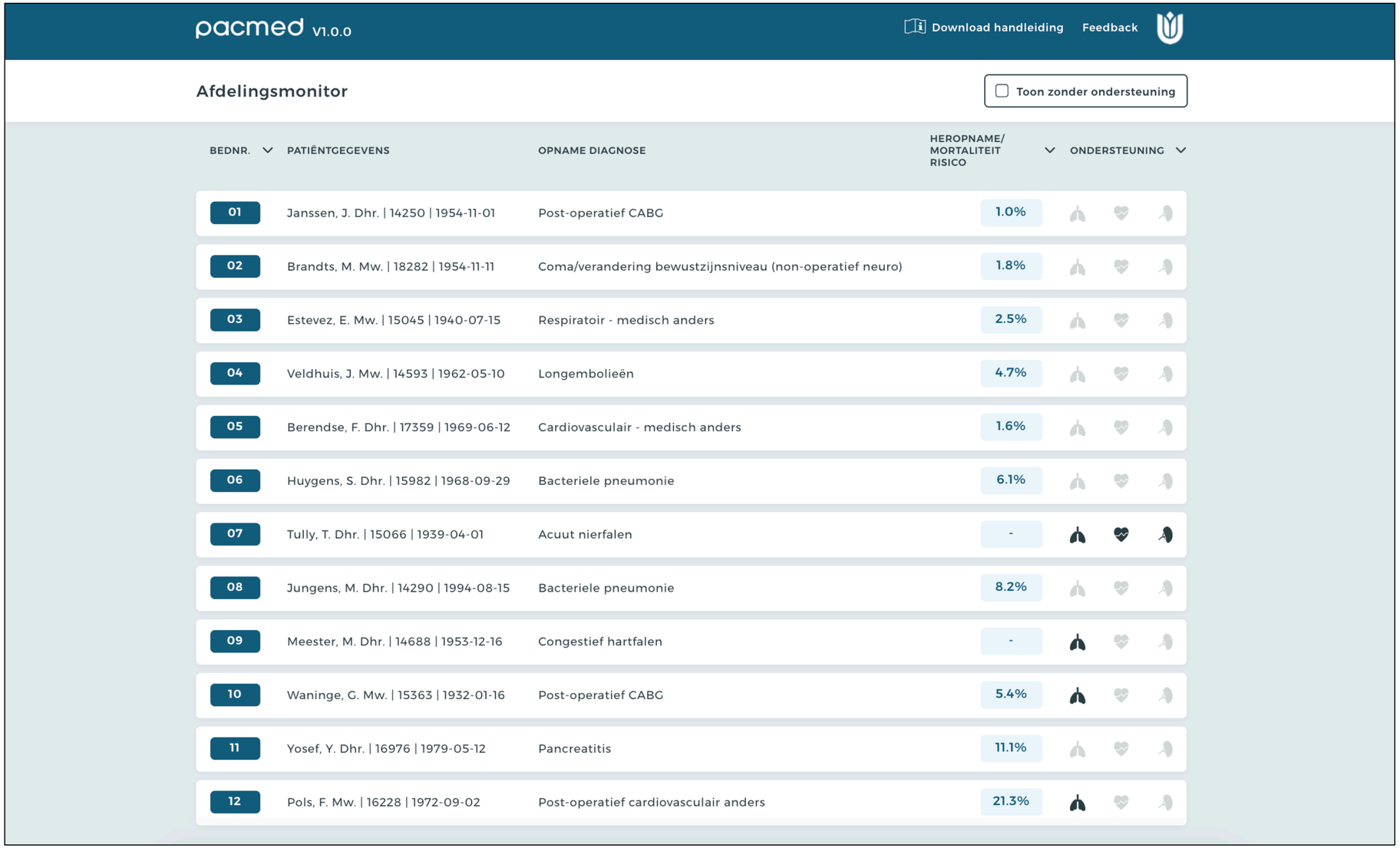


pacmed

ZIEKENHUIS ST ANTONIUS

Erasmus MC Universitair Medisch Centrum Rotterdam

Machine learning software for ICU readmission risk



So what is the goal of interpretability?

Gain **trust** in the model’s predictions

Able to place the model’s predictions in the right **medical context**; understand when to use the model and when not to use it

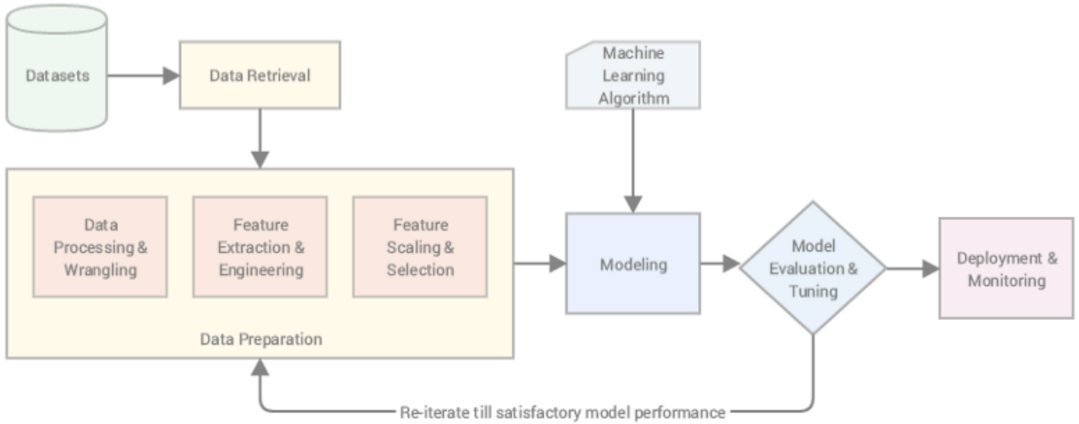
What are the different methods to achieve interpretability?

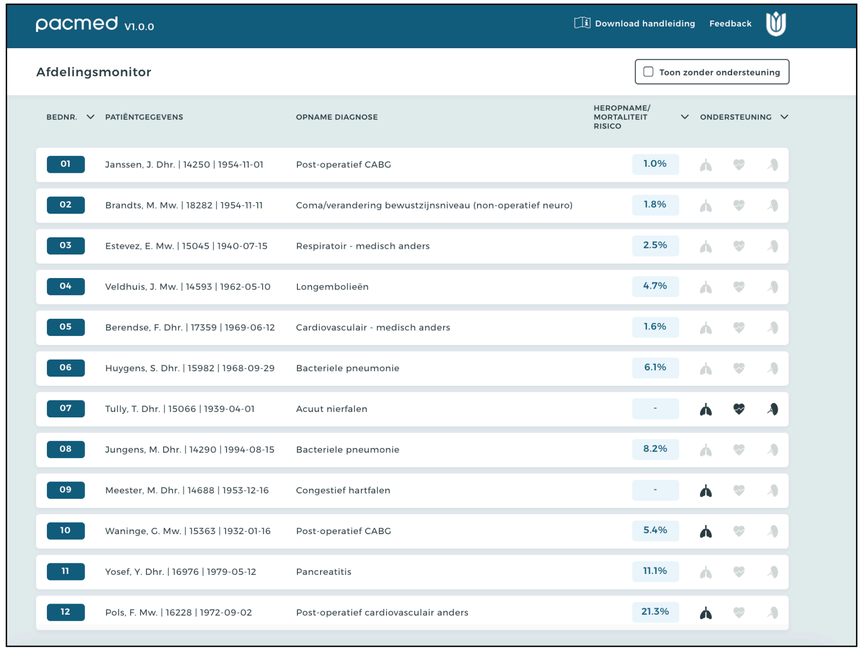
Transparency of the methodology and scope of the model

Global **understanding** of the workings of the **model**

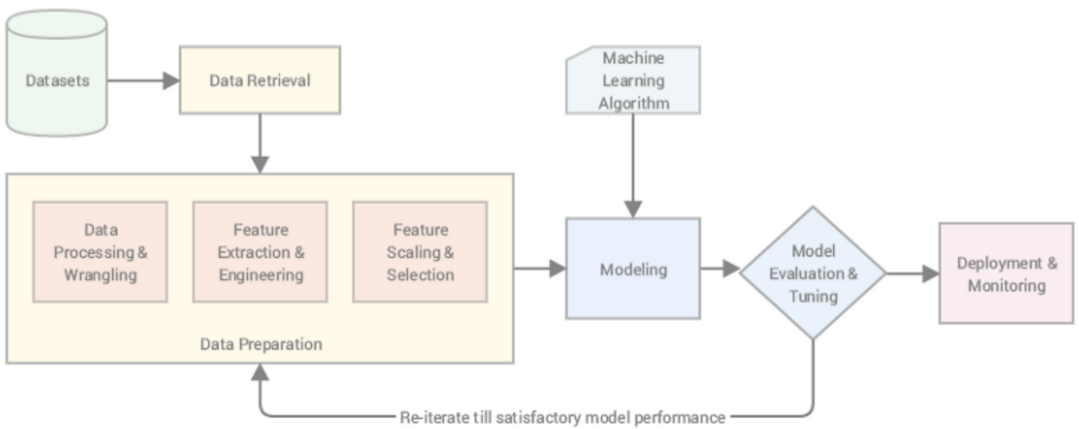
Specific **interpretability** of the model’s **predictions**

Inputs of the model are important for placing the model in the appropriate context





Inputs of the model are important for placing the model in the appropriate context



Has this type of patient ever been seen before, by the model?

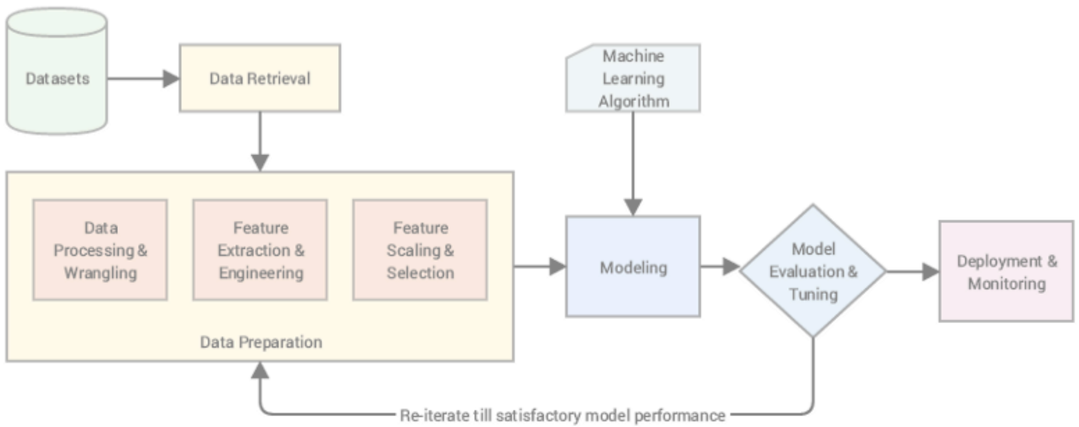
pacmed ... eiding Feedback

Afdelingsmonitor

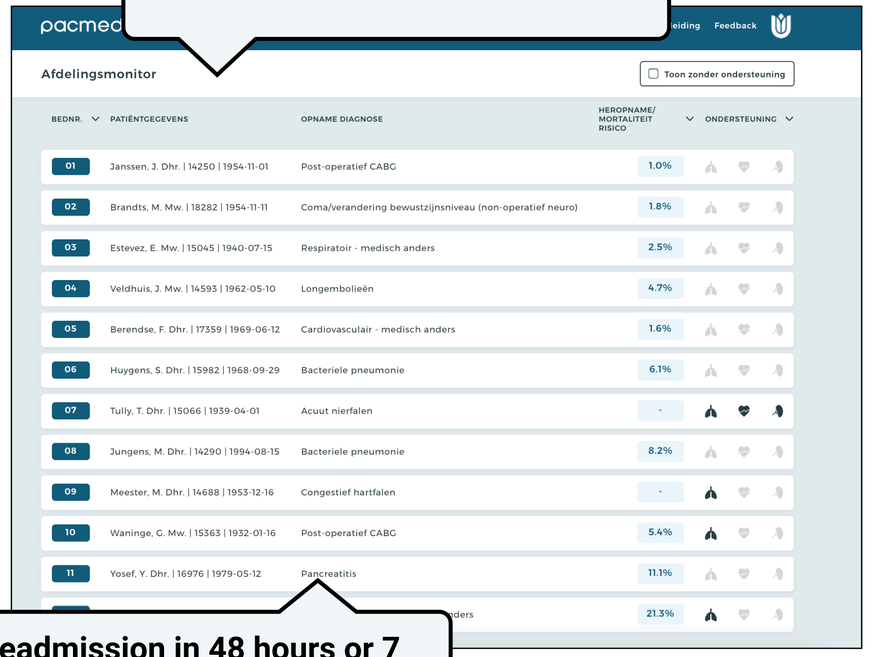
Toon zonder ondersteuning

BEDNR.	PATIËNTGEGEVENS	OPNAME DIAGNOSE	HEROPNAME/ MORTALITEIT RISICO	ONDERSTEUNING
01	Janssen, J. Dhr. | 14250 | 1954-11-01	Post-operatief CABG	1.0%	
02	Brandts, M. Mw. | 18282 | 1954-11-11	Coma/verandering bewustzijnsniveau (non-operatief neuro)	1.8%	
03	Estevez, E. Mw. | 15045 | 1940-07-15	Respiratoir - medisch anders	2.5%	
04	Veldhuis, J. Mw. | 14593 | 1962-05-10	Longembolieën	4.7%	
05	Berendse, F. Dhr. | 17359 | 1969-06-12	Cardiovasculair - medisch anders	1.6%	
06	Huygens, S. Dhr. | 15982 | 1968-09-29	Bacteriele pneumonie	6.1%	
07	Tully, T. Dhr. | 15066 | 1939-04-01	Acuut nierfalen	-	
08	Jungens, M. Dhr. | 14290 | 1994-08-15	Bacteriele pneumonie	8.2%	
09	Meester, M. Dhr. | 14688 | 1953-12-16	Congestief hartfalen	-	
10	Waninge, G. Mw. | 15363 | 1932-01-16	Post-operatief CABG	5.4%	
11	Yosef, Y. Dhr. | 16976 | 1979-05-12	Pancreatitis	11.1%	
12	Pols, F. Mw. | 16228 | 1972-09-02	Post-operatief cardiovasculair anders	21.3%	

Inputs of the model are important for placing the model in the appropriate context

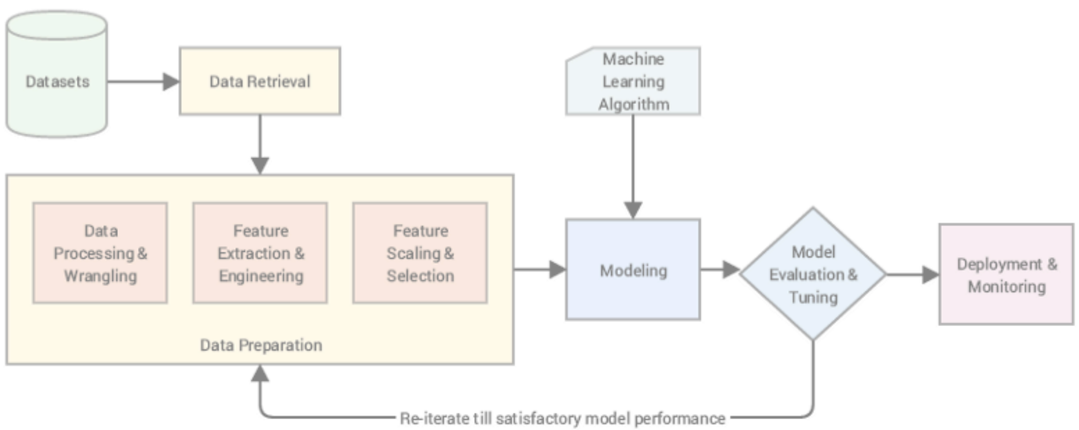


Has this type of patient ever been seen before, by the model?

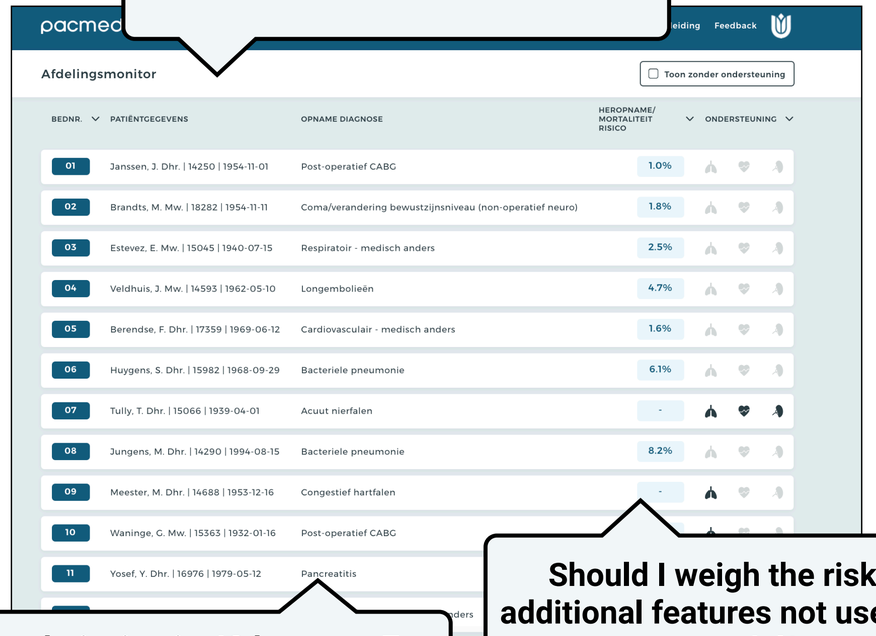


Readmission in 48 hours or 7 days?

Inputs of the model are important for placing the model in the appropriate context



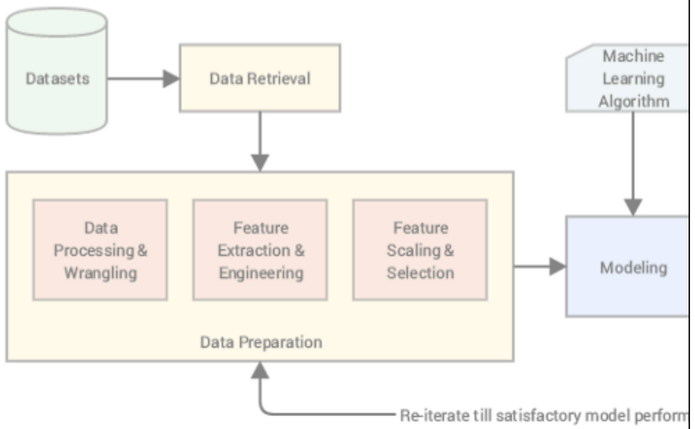
Has this type of patient ever been seen before, by the model?



Readmission in 48 hours or 7 days?

Should I weigh the risk with additional features not used in the model?

All these inputs are available to the doctor in an Instructions for Use



pacmed Critical

ICU Discharge Decision Support Tool

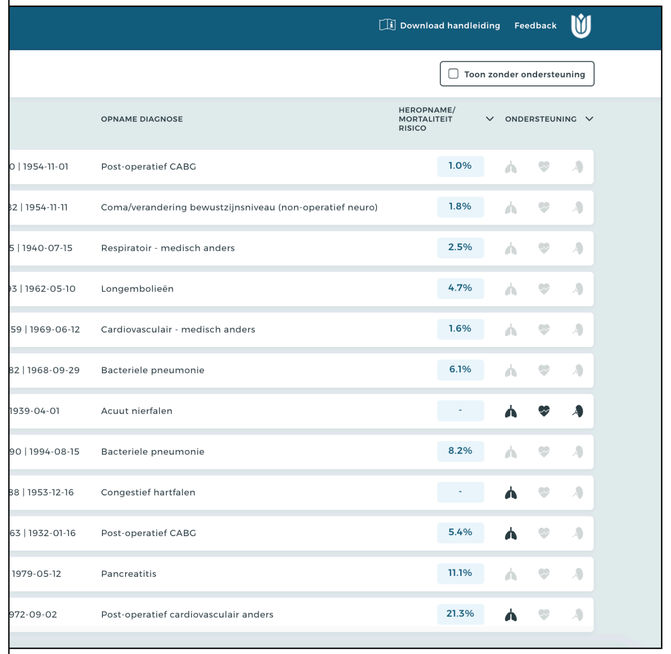
Instructions for Use

English

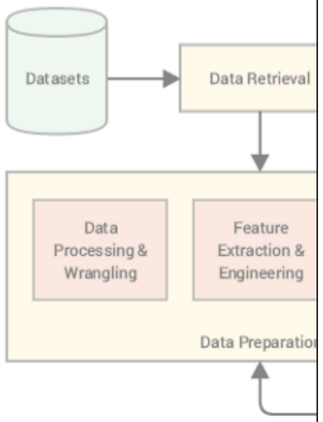
LOT Software version: 1.0.0

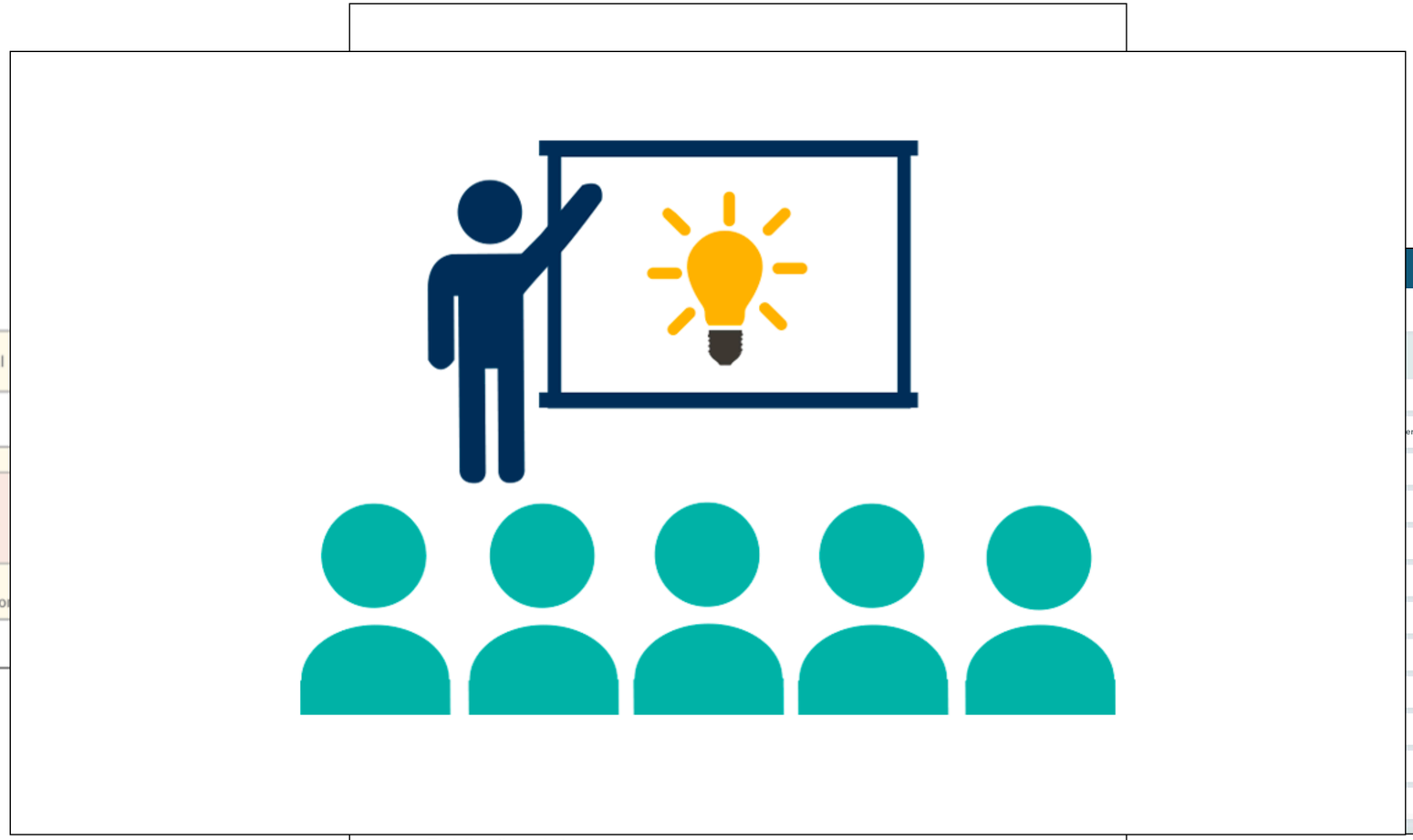
Software release date:

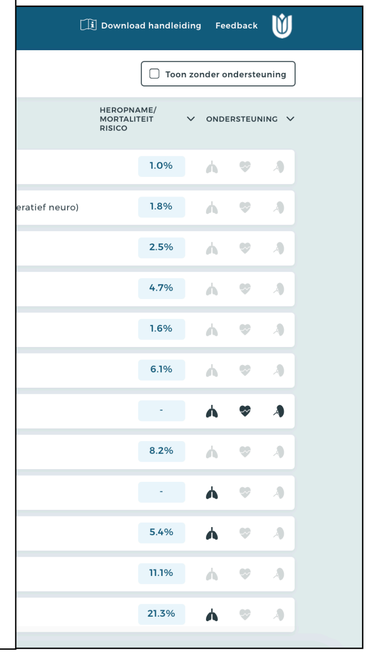
IFU version: 1.0
Publication date:



And doctors are trained to use the software







Publication date:

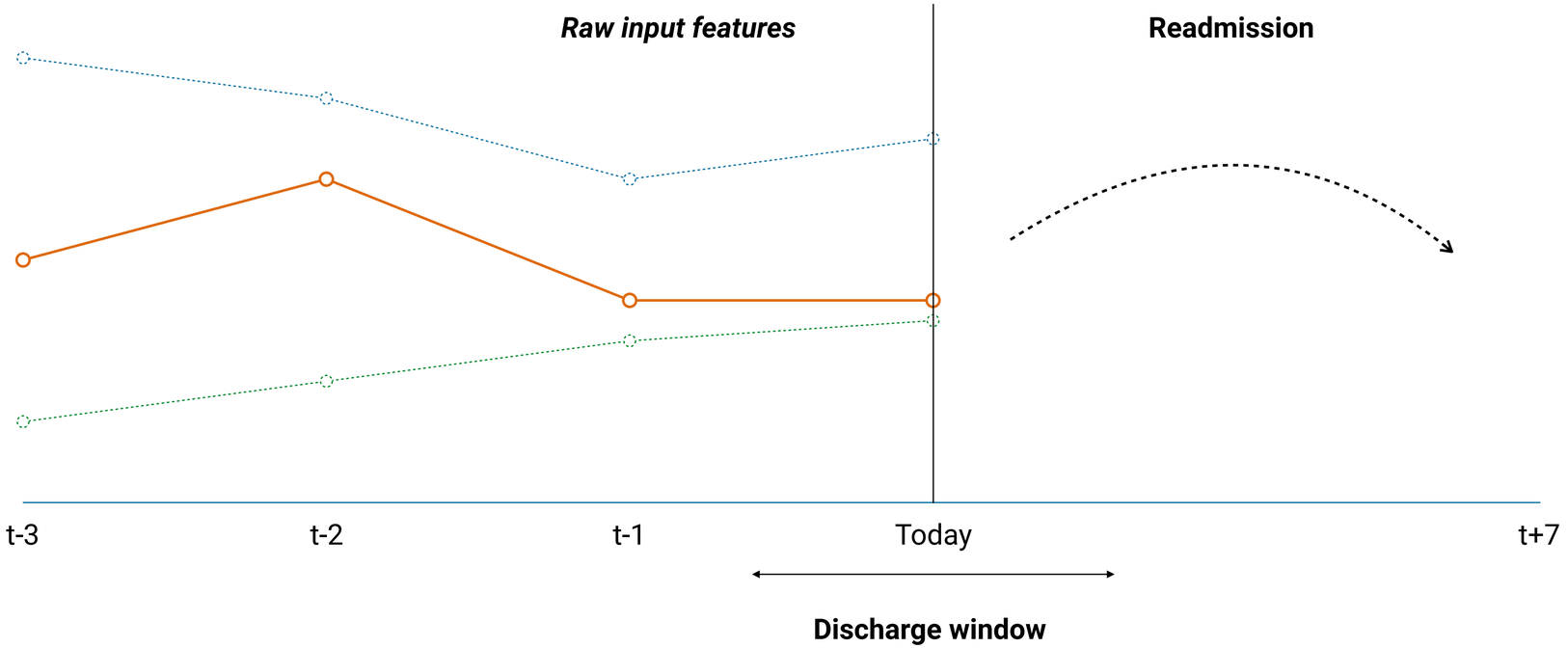
What are the different methods to achieve interpretability?

Transparency of the methodology and scope of the model

Global **understanding** of the workings of the **model**

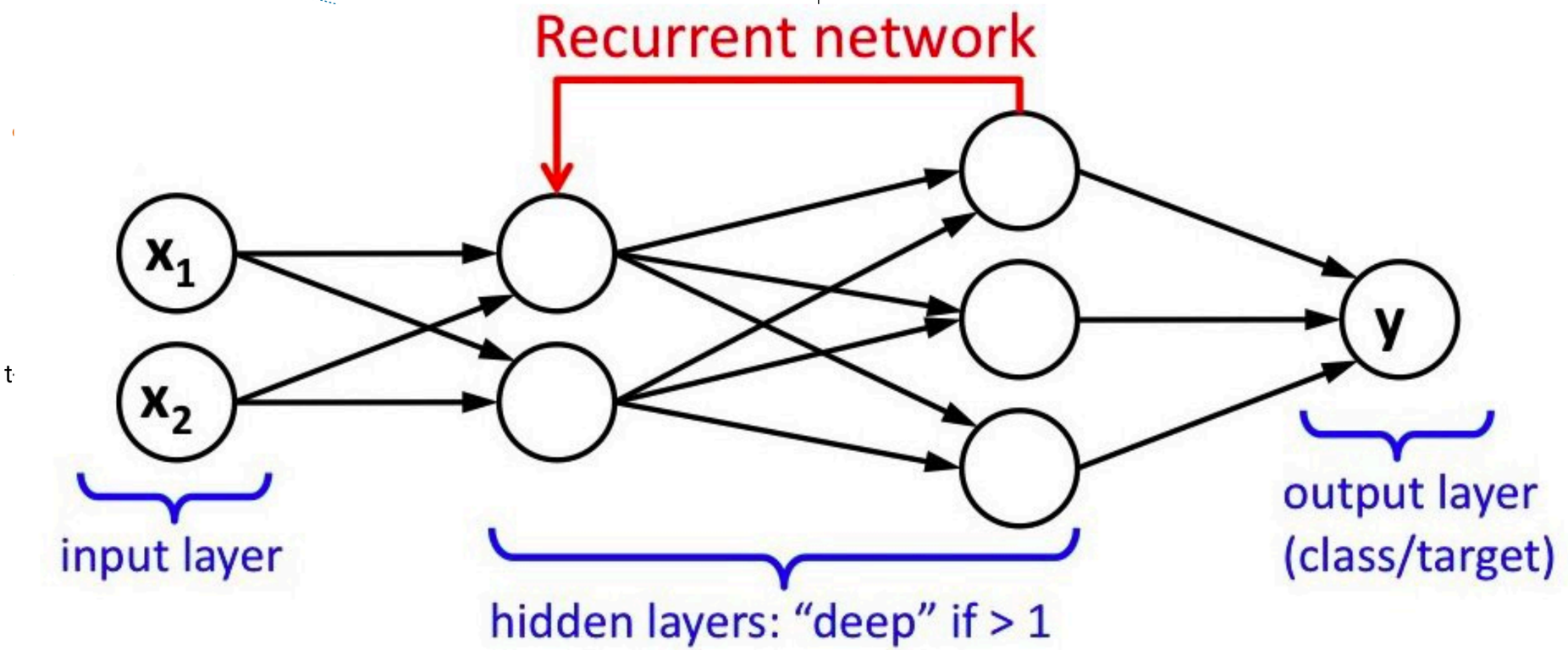
Specific **interpretability** of the model’s **predictions**

How to turn the many time series into information relevant for a model (and doctor)?

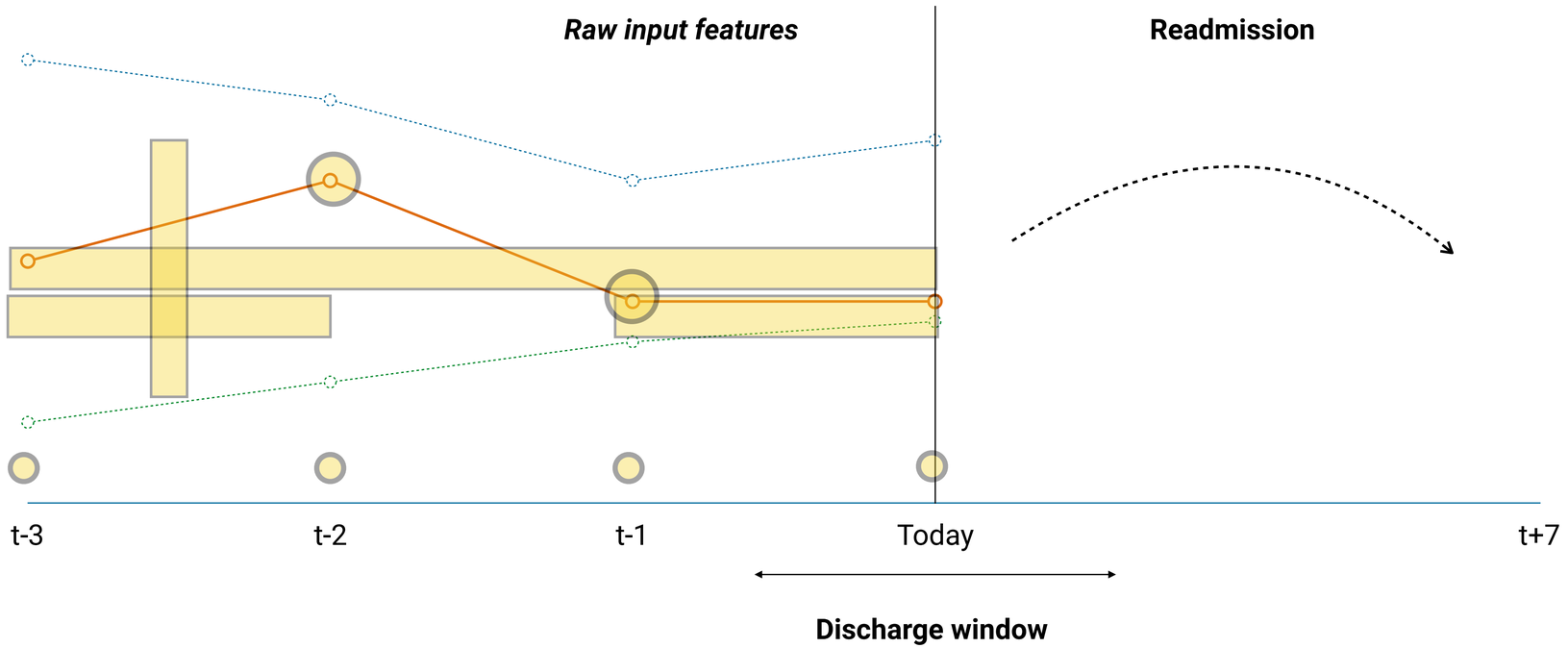


Recurrent neural nets?

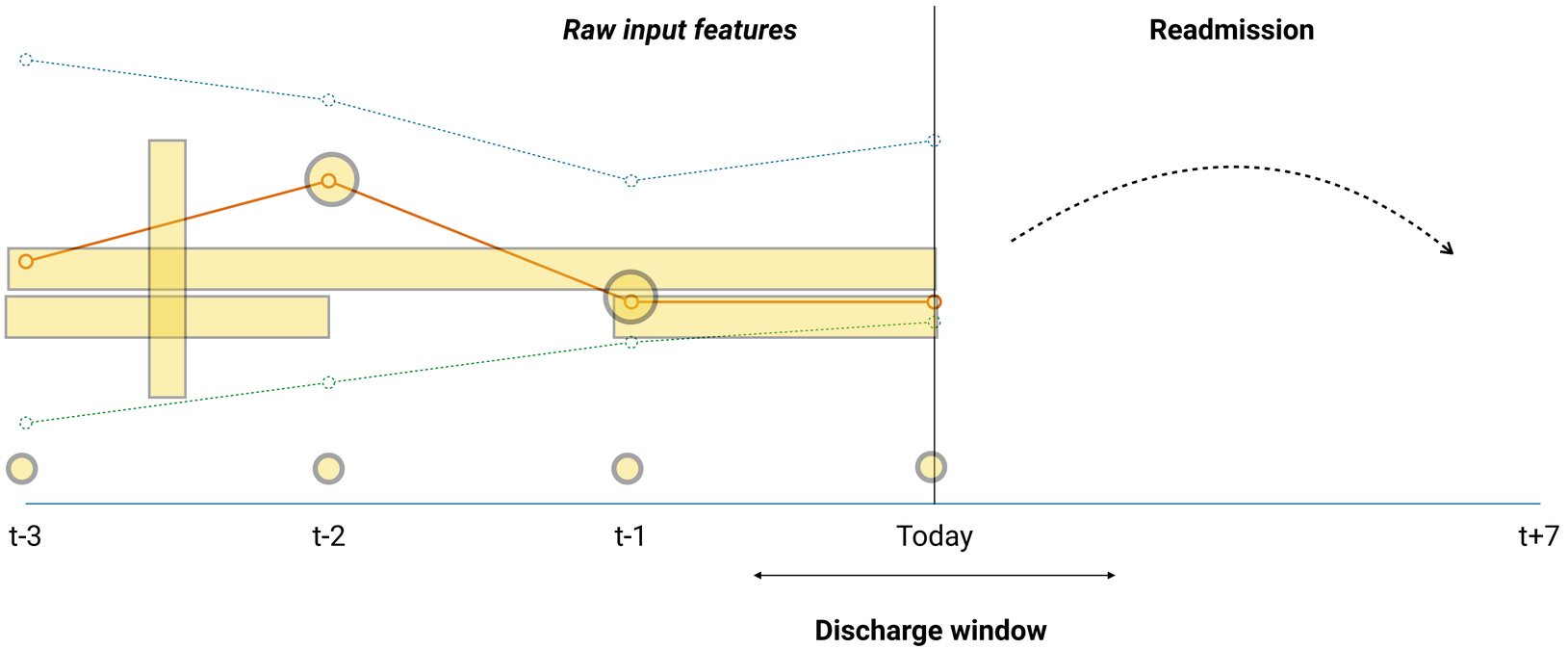
Raw input features | **Readmission**

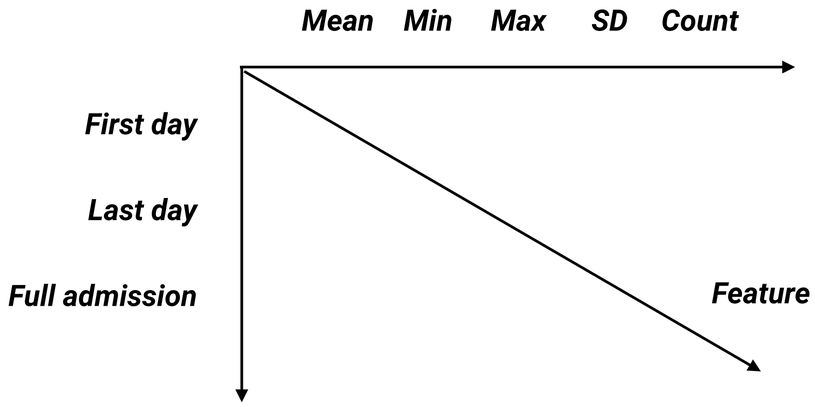


Or explicit feature engineering..

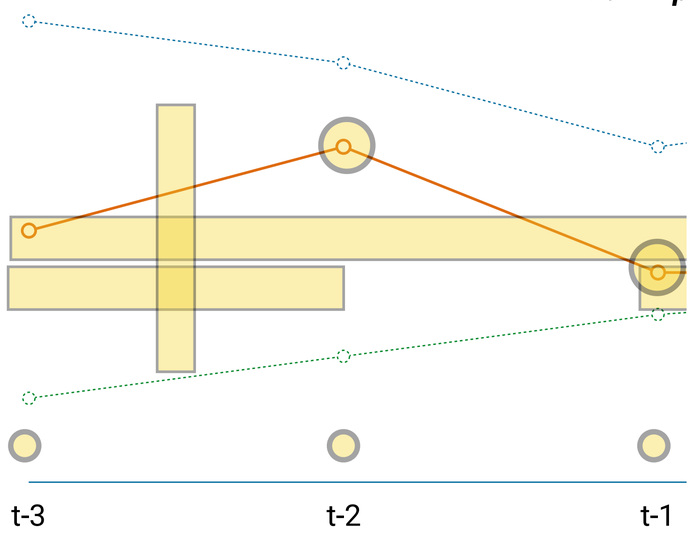


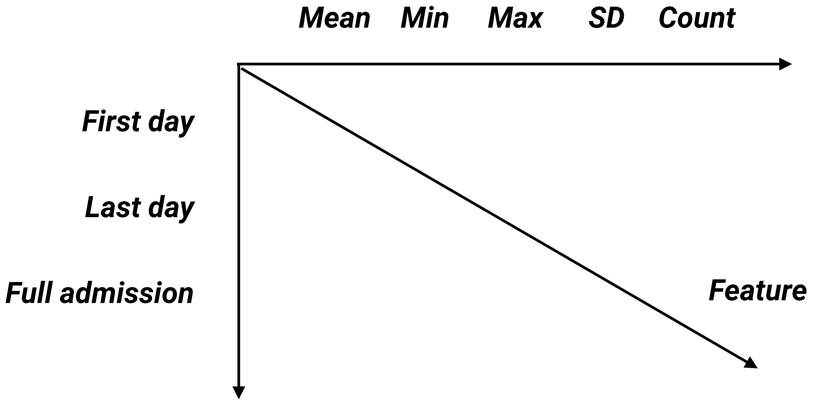
..which allows us to calculate simple and understandable values..





..for hundreds of medical parameters





Patient and admission characteristics

- Age, sex
- Length and weight at admission
- Department of origin
- Length of stay
- Number of prior admissions
- Time in the hospital before admission
- CPR code

Vital signs & device data

Respiration

- Respiratory rate
- Mechanical Ventilation
- Tidal Volume
- Expiratory minute Volume
- Respiration modus
- PEEP
- Piek druk
- Supplemental O2
- Fraction of inspired O2
- Type of O2 administration
- Peripheral O2 saturation

Circulation

- Blood pressure (diastolic and systolic, arterial and non-invasive)
- Pulmonary artery press. (diastolic and systolic)
- CVP
- PCWP wedge
- Heart rate
- Cardiac output
- Tidal volume (inspiratory and expiratory)
- Heart rhythm & ectopic
- Shock index
- Temperature peripheral

Other

- CVVH
- Lines and drains

Medication categories

- Alimentary tract and metabolism
- Antibiotics
- Blood and blood-forming organs
- Cardiovascular
- Musculoskeletal system
- Nervous system
- General (sondevoeding)

Lab values

Blood gas analysis

- Base excess
- O2 content in blood
- Arterial O2 saturation
- pH
- Part. press. (O_2 & CO_2)
- Actual bicarbonate

Chemistry

- Natrium, Kalium
- Chloride
- Calcium, ion. Calcium
- Magnesium
- Fosfaat
- Creatinine
- CK
- EST and CRP
- Blood glucose
- Blood lactate
- Amylase
- Serum albumine
- BUN_creatinine
- NT-ProBNP

Haematology

- Hb, Ht
- White blood cell count
- MCH, MCV
- Erythrocytes
- Thrombocytes
- Lymphocytes
- Leucocytes
- Baso, eo and neutro
- Reticulocytes
- PT, APTT

Cardiac enzymes

- CK-MB
- Troponin-T

Liver tests

- ALAT and ASAT
- GGT, AF
- LDH
- Bilirubine

Urinalysis

- Natrium, Kalium
- Ureum

Clinical observations

- CAM, DOS, RASS, NAS
- GCS
- Pupil size and reaction
- Cough stimulant
- Urine output
- Number of bronchial toilets

We can start with global feature importance (for understandable features)

Feature importance Gradient Boosting model

in feature list | in model

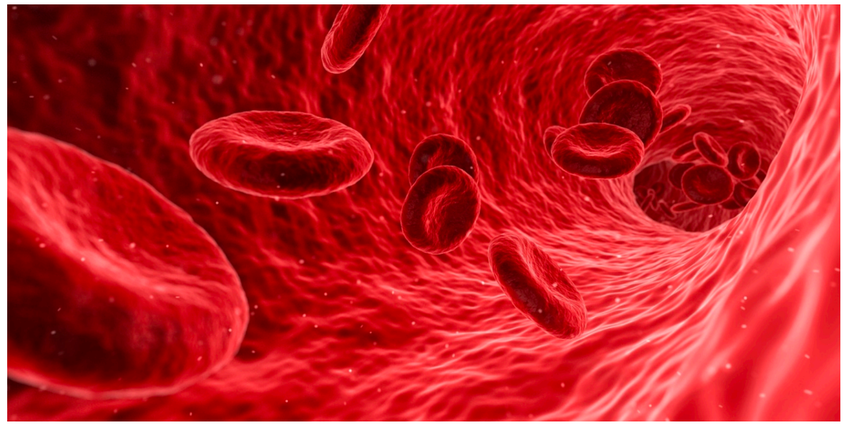
Feature (category)

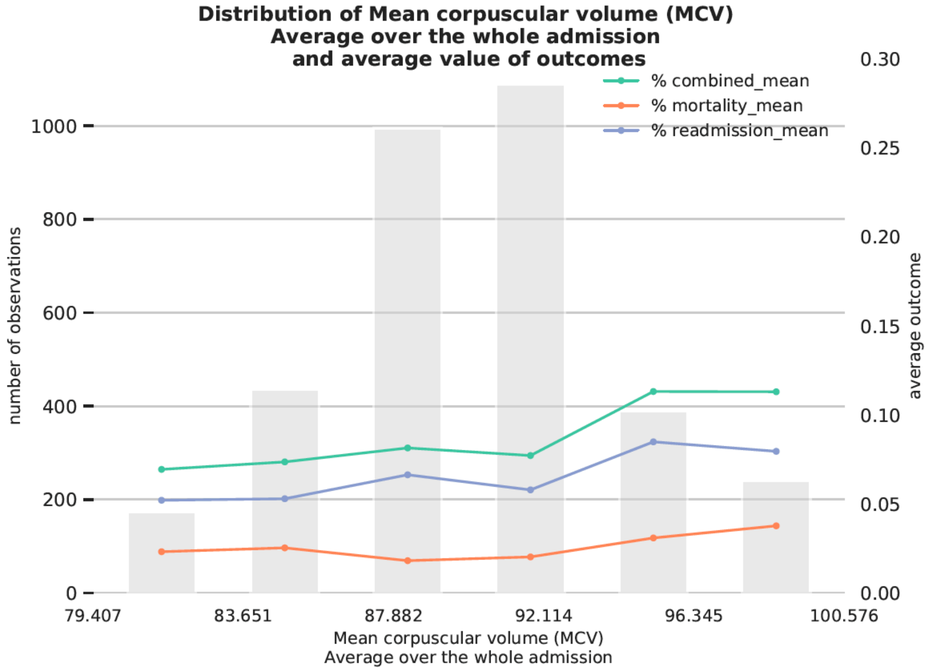
category	subcategory	feature
0. lab values	00. neurology	amylase
0. lab values	00. neurology	base_excess
0. lab values	00. neurology	bicarbonate
0. lab values	00. neurology	bilirubin
0. lab values	00. neurology	creatinine
0. lab values	00. neurology	glucose
0. lab values	00. neurology	lactate
0. lab values	00. neurology	pco2
0. lab values	00. neurology	ph
0. lab values	00. neurology	po2
0. lab values	00. neurology	spo2
0. lab values	00. neurology	troponin
0. lab values	02. haematology	aptt
0. lab values	02. haematology	baso
0. lab values	02. haematology	baso_perc
0. lab values	02. haematology	eo
0. lab values	02. haematology	eo_perc
0. lab values	02. haematology	erythrocyten
0. lab values	02. haematology	esr
0. lab values	02. haematology	hemaglobin
0. lab values	02. haematology	hematocrit
0. lab values	02. haematology	leukocyten
0. lab values	02. haematology	lymfo
0. lab values	02. haematology	lymfo_perc
0. lab values	02. haematology	mch
0. lab values	02. haematology	mcv
0. lab values	02. haematology	mono
0. lab values	02. haematology	mono_perc
0. lab values	02. haematology	neutro
0. lab values	02. haematology	neutro_perc
0. lab values	02. haematology	prothrombin_time
0. lab values	02. haematology	reticulocyten
0. lab values	02. haematology	thrombo
0. lab values	04. chemistry	albumin
0. lab values	04. chemistry	bun_creatinine
0. lab values	04. chemistry	calcium
0. lab values	04. chemistry	calcium_ion
0. lab values	04. chemistry	chloride
0. lab values	04. chemistry	creatinine_kinase
0. lab values	04. chemistry	crp
0. lab values	04. chemistry	fosfaat
0. lab values	04. chemistry	kalium
0. lab values	04. chemistry	magnesium
0. lab values	04. chemistry	natrium
0. lab values	04. chemistry	nt_probnp
0. lab values	04. chemistry	protein_total
0. lab values	04. chemistry	ureum
0. lab values	05. liver tests	alat
0. lab values	05. liver tests	alkaline_fosf
0. lab values	05. liver tests	asat
0. lab values	05. liver tests	gamma_gt
0. lab values	05. liver tests	lactate_ld
0. lab values	06. urinalysis	kalium_urine
0. lab values	06. urinalysis	natrium_urine
0. lab values	06. urinalysis	ureum_urine
0. lab values	10. cardiac enzymes	mb_enzyme
0. lab values	11. blood gas analysis	o2_content

Aggregation type

Univariate relation between feature and outcome ‘medically validates’ the model

MCV





(HR: 1.901; 95% CI 1.357-2.662). Other variables associated with this outcome included age, tracheostomy and mean corpuscular volume (MCV) at ICU discharge. Similar results were obtained after the exclusion of unexpected deaths in the ward (HR 1.940; CI 1.312-2.871) and for in-hospital deaths (HR 1.716; 95% CI 1.141-2.580).

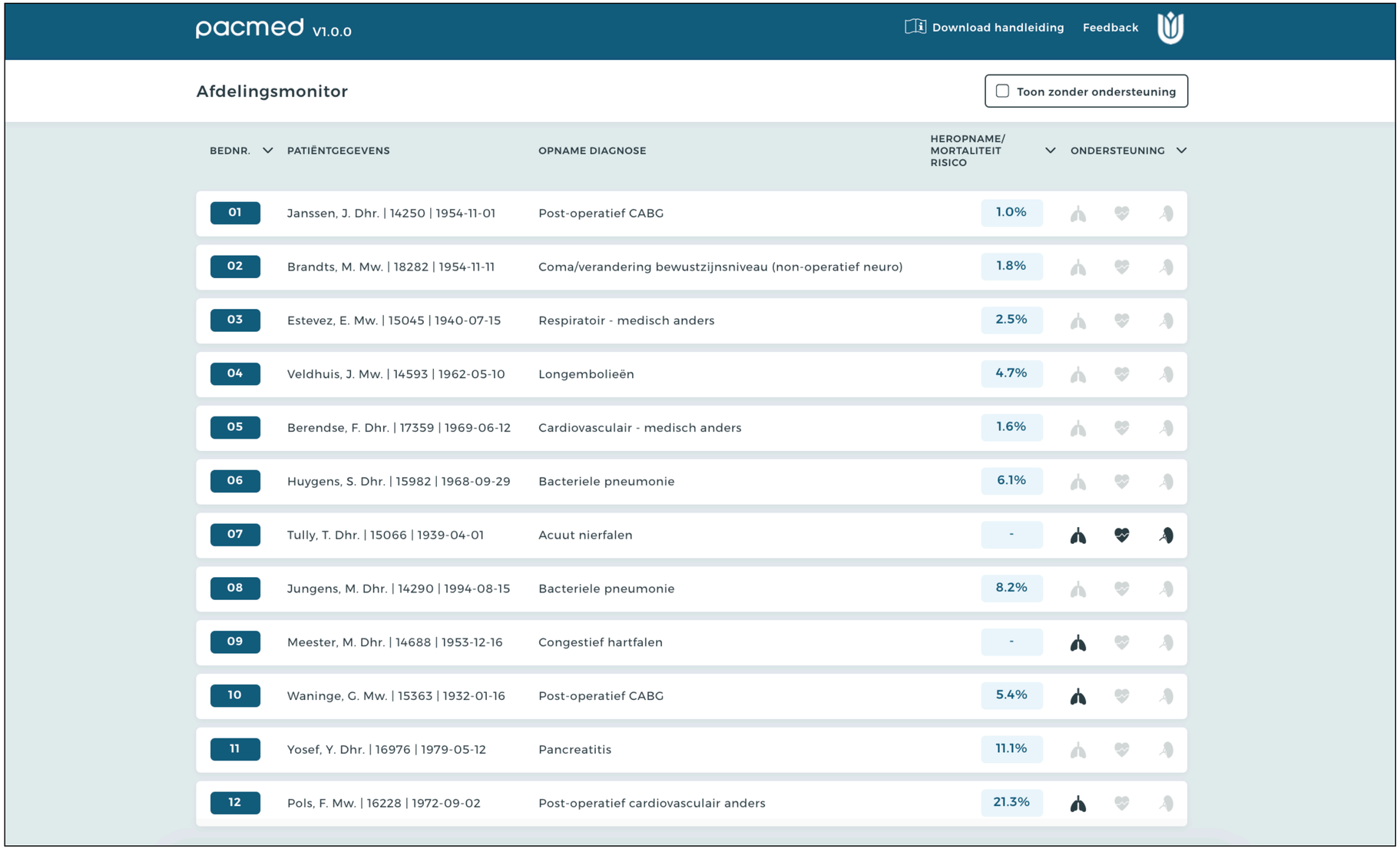
What are the different methods to achieve interpretability?

Transparency of the methodology and scope of the model

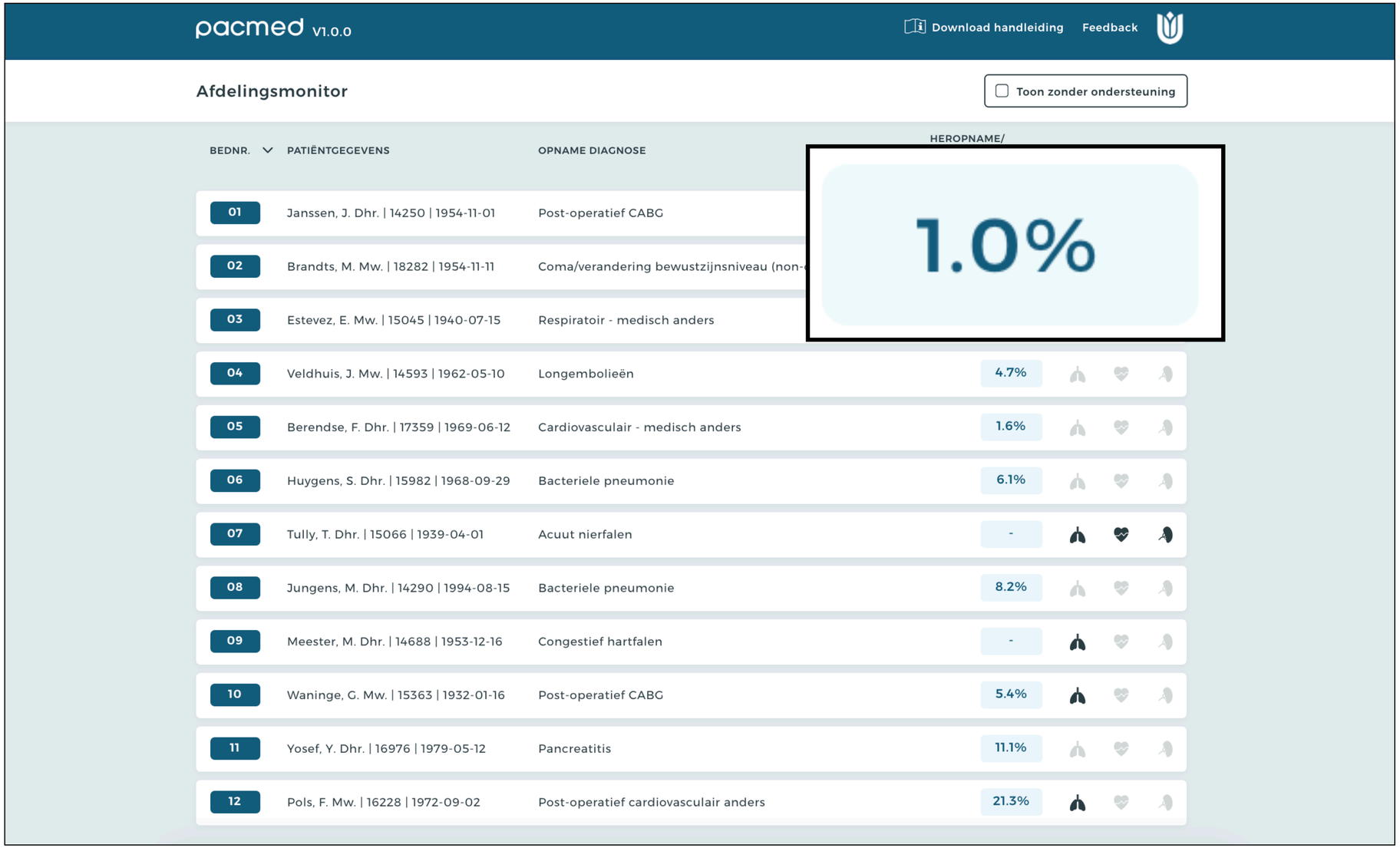
Global **understandability** of the workings of the **model**

Specific **interpretability** of the model’s **predictions**

If we ‘trust’ the model globally does sensible things; let’s start using the predictions?



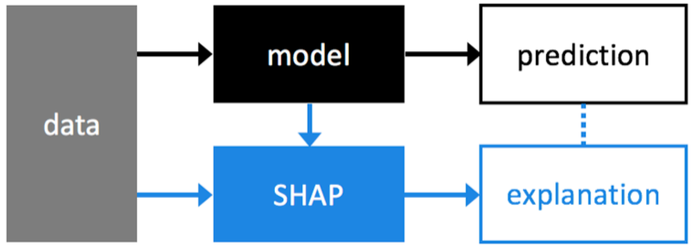
1.0% readmission risk: can I safely discharge this patient?



Let’s introduce SHAP

GitHub, Inc. [US] | https://github.com/slundberg/shap

README.md



build passing

SHAP (SHapley Additive exPlanations) is a unified approach to explain the output of any machine learning model. SHAP connects game theory with local explanations, uniting several previous methods [1-7] and representing the only possible consistent and locally accurate additive feature attribution method based on expectations (see our papers for details).

Install

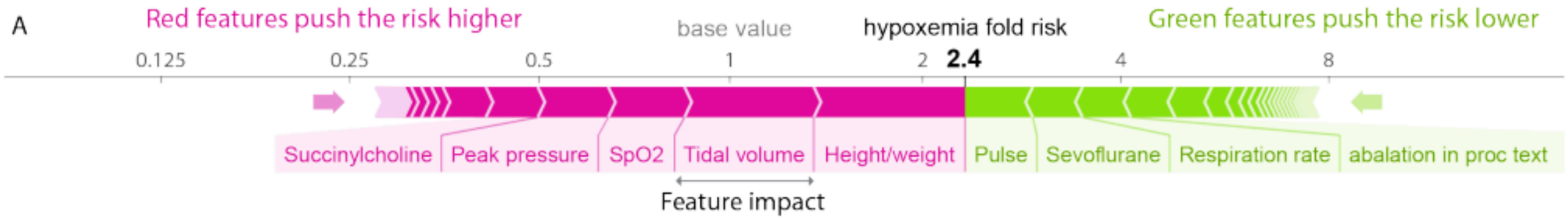
SHAP can be installed from either PyPI or conda-forge:

```
pip install shap
or
conda install -c conda-forge shap
```

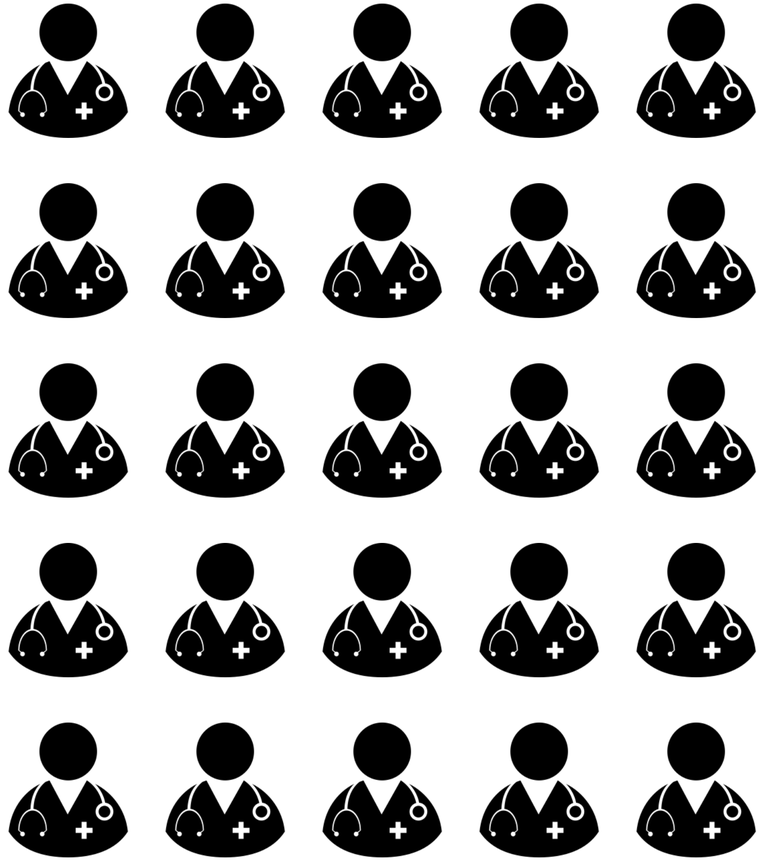
Example: features driving hypoxemia risk during surgery

Explainable machine learning predictions to help anesthesiologists prevent hypoxemia during surgery

Authors: Scott M. Lundberg[1], Bala Nair[2,6], Monica S. Vavilala[2,6], Mayumi Horibe[4], Michael J. Eisses[2,3], Trevor Adams[2,3], David E. Liston[2,3], Daniel King-Wai Low[2,3], Shu-Fang Newman[2], Jerry Kim[2,3], Su-In Lee[1,5]*



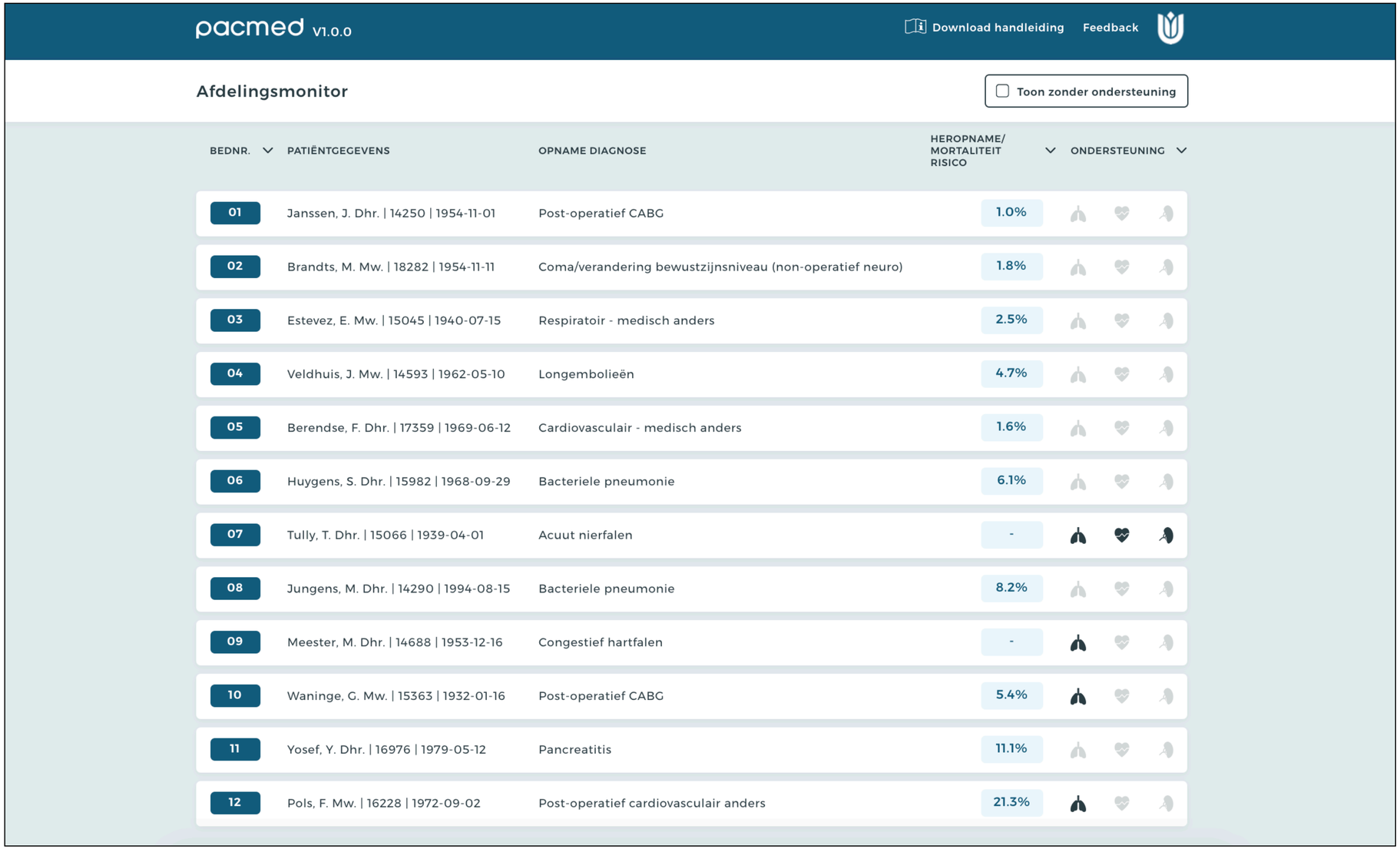
We tested our machine learning software with 25 intensive care doctors!



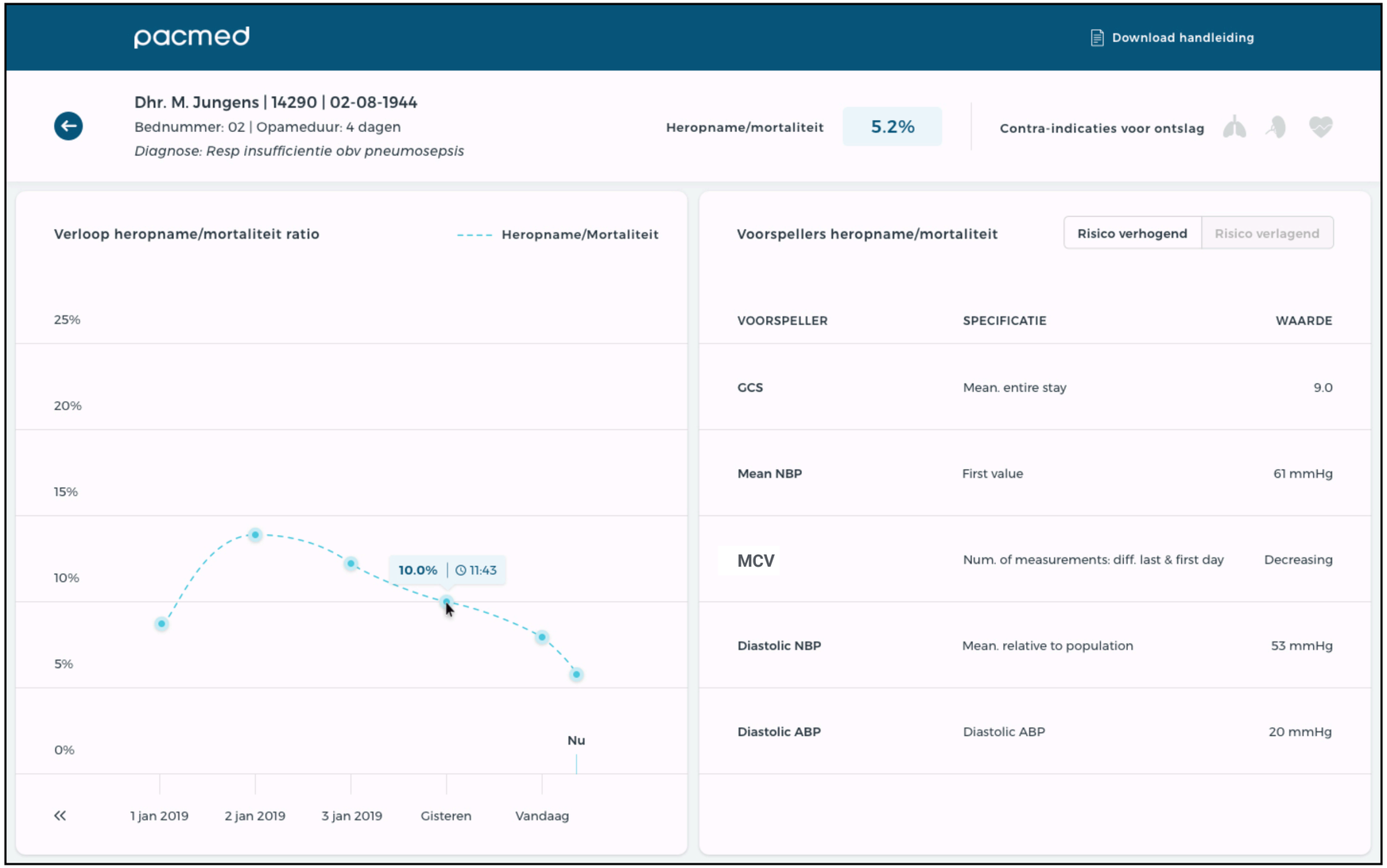
Amsterdam UMC
Universitair Medische Centra

UMC Utrecht

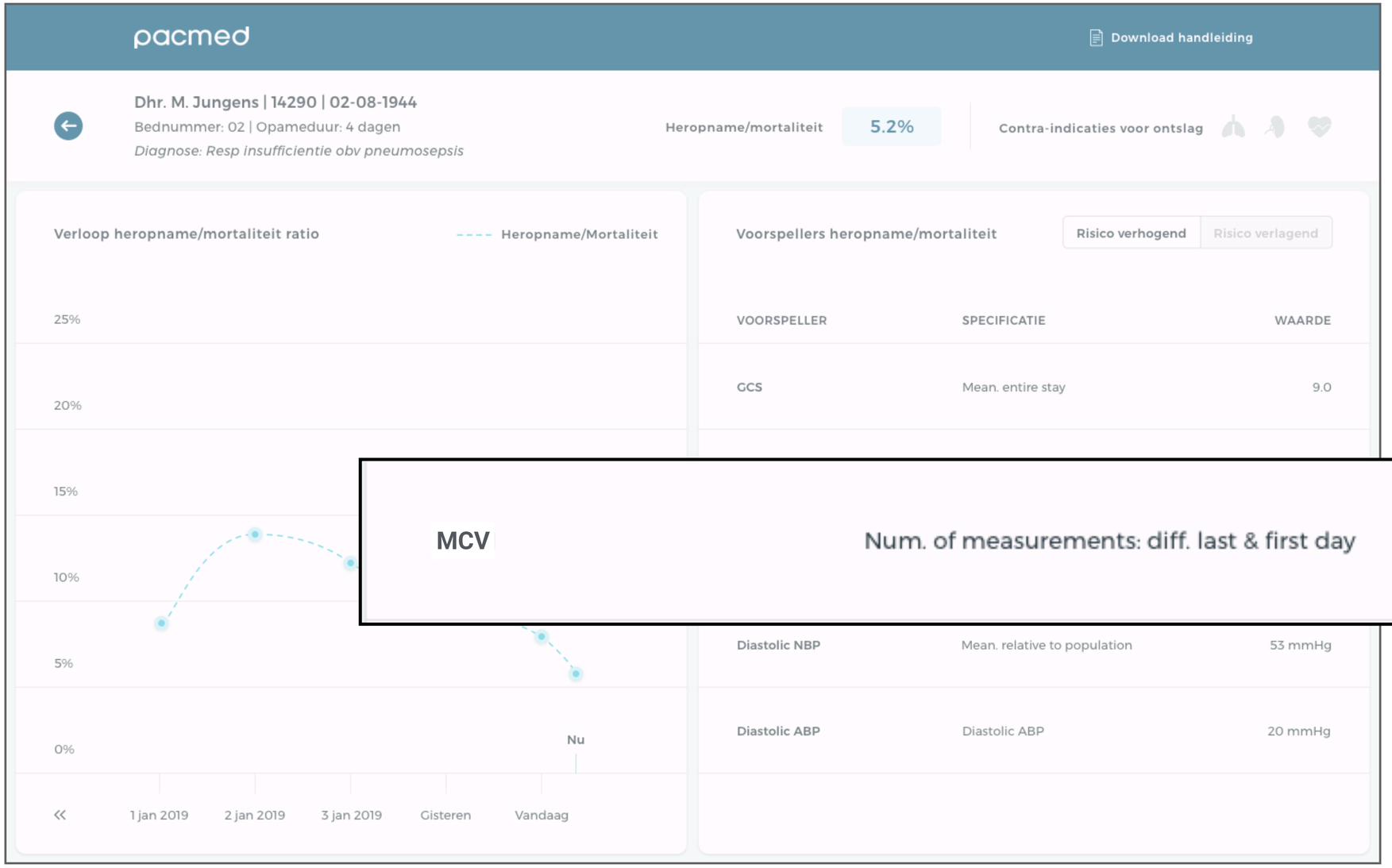
From the main screen, you can click on individual patients



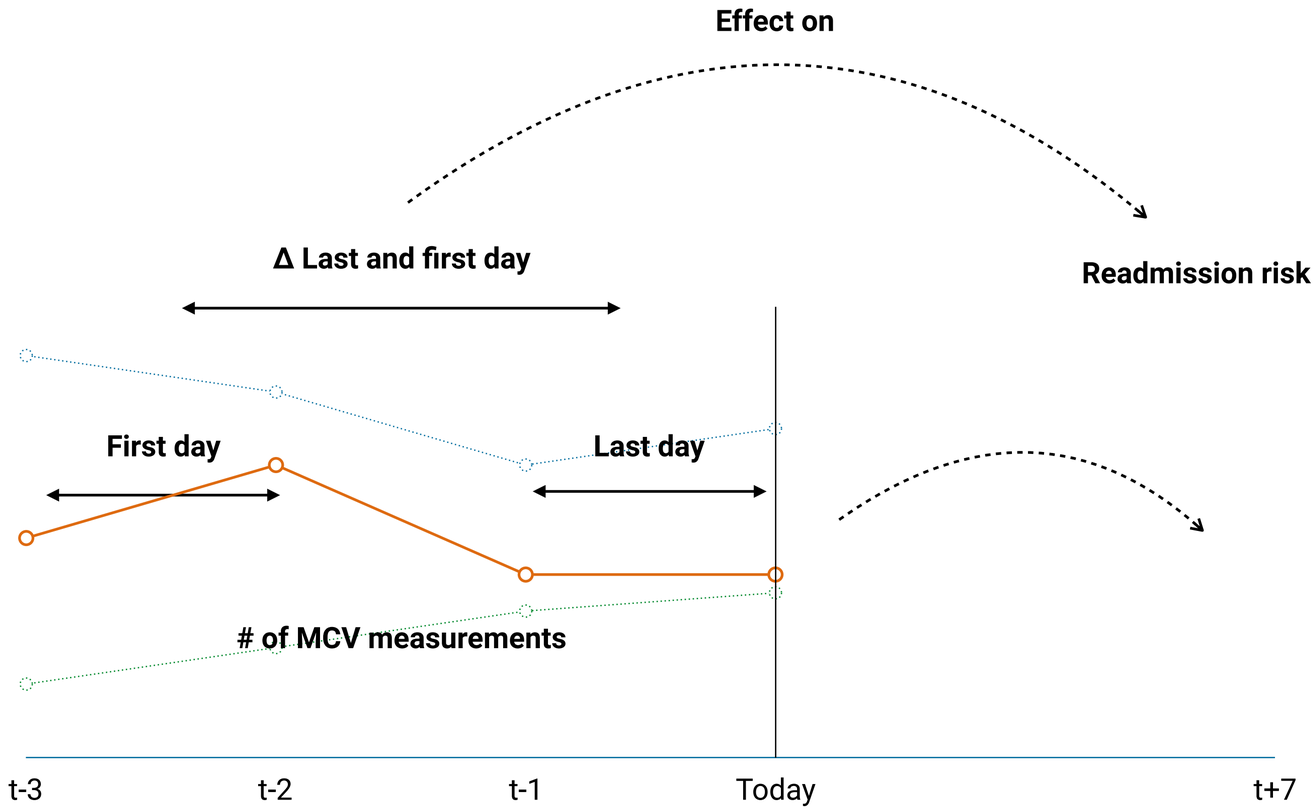
Attempt 1: Let’s add the top 5 predictive features (SHAP) to the user interface



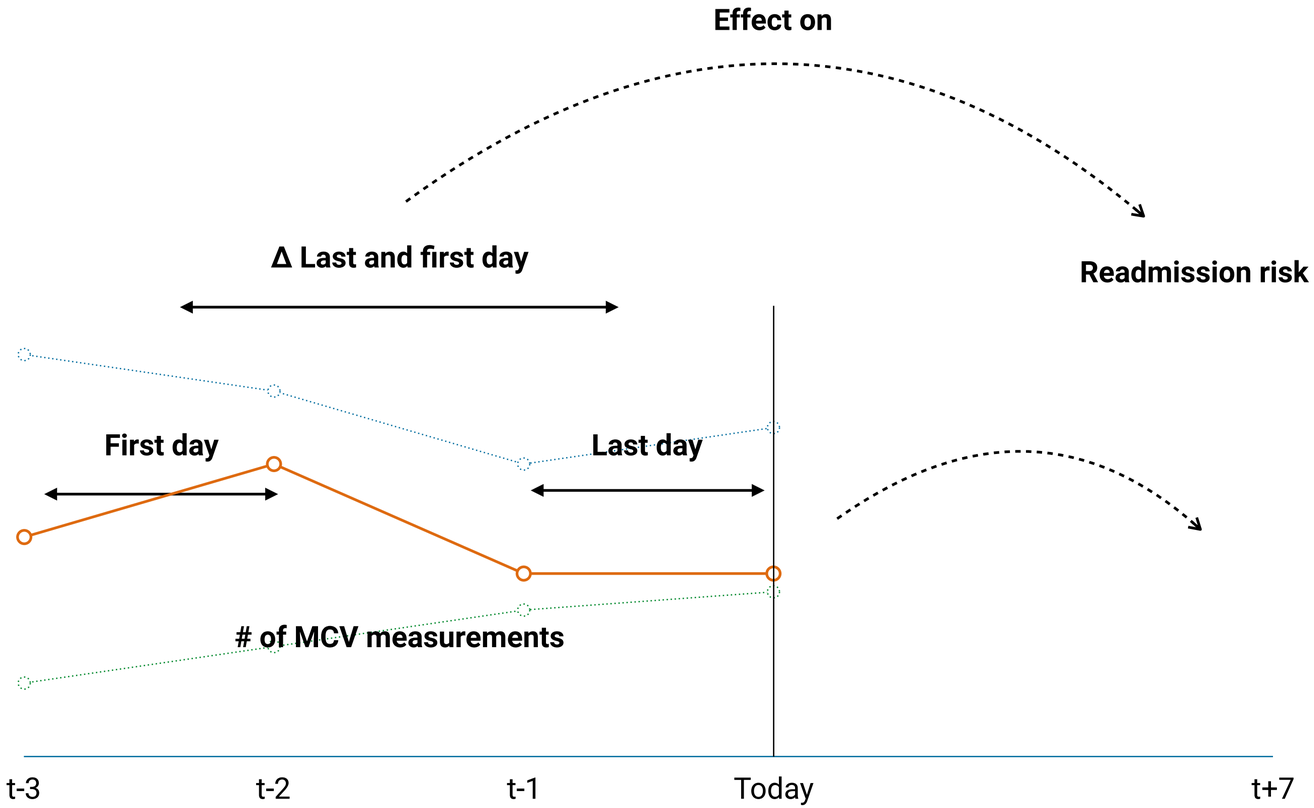
Attempt 1: Let’s add the top 5 predictive features (SHAP) to the user interface



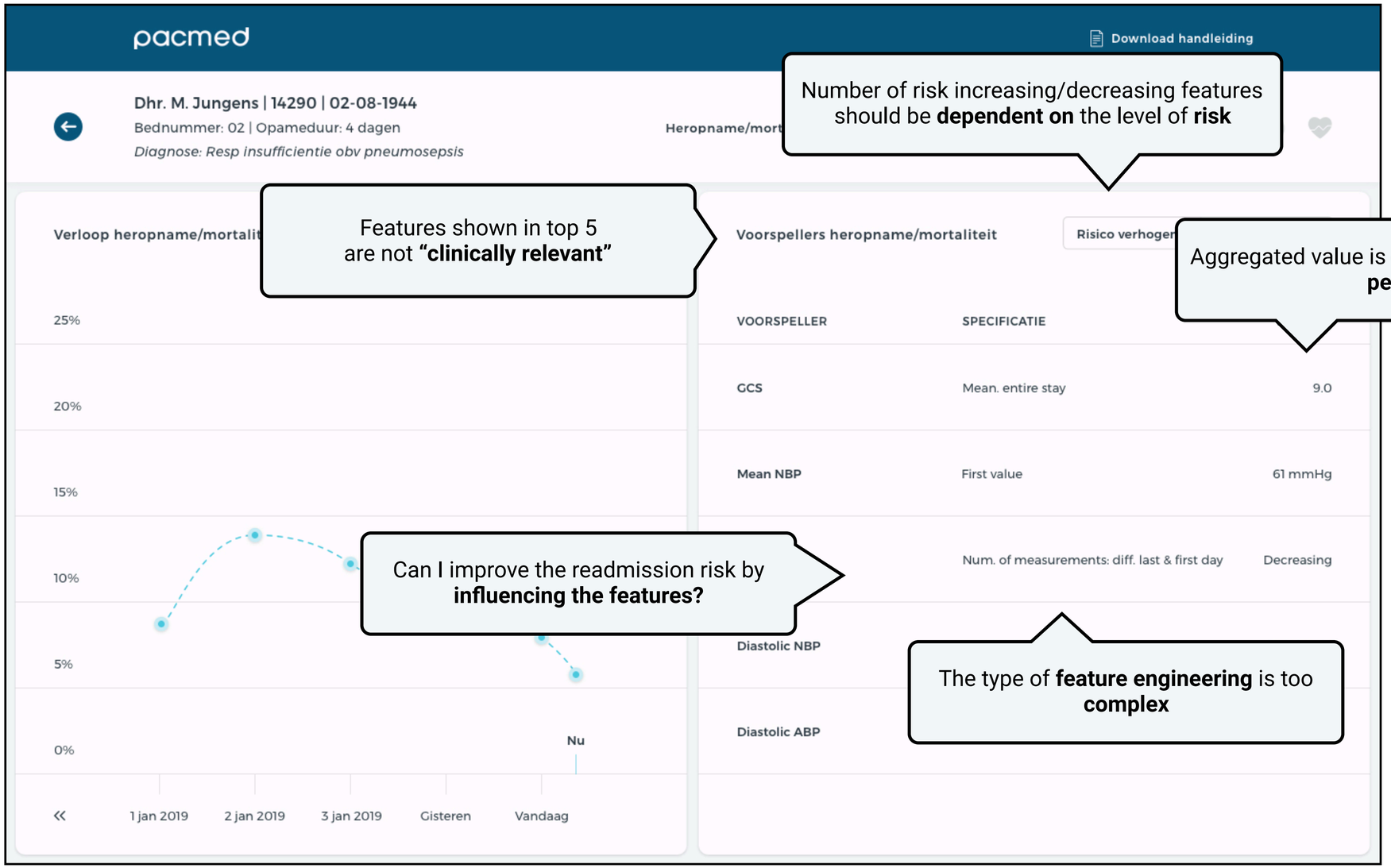
Let’s dive into this example



Let’s dive into this example



We learned a lot about what does not work in making ML interpretable



Attempt 2: SHAP-based interpretability layer with additional contextual information

pacmed V1.0.0

Download handleiding Feedback

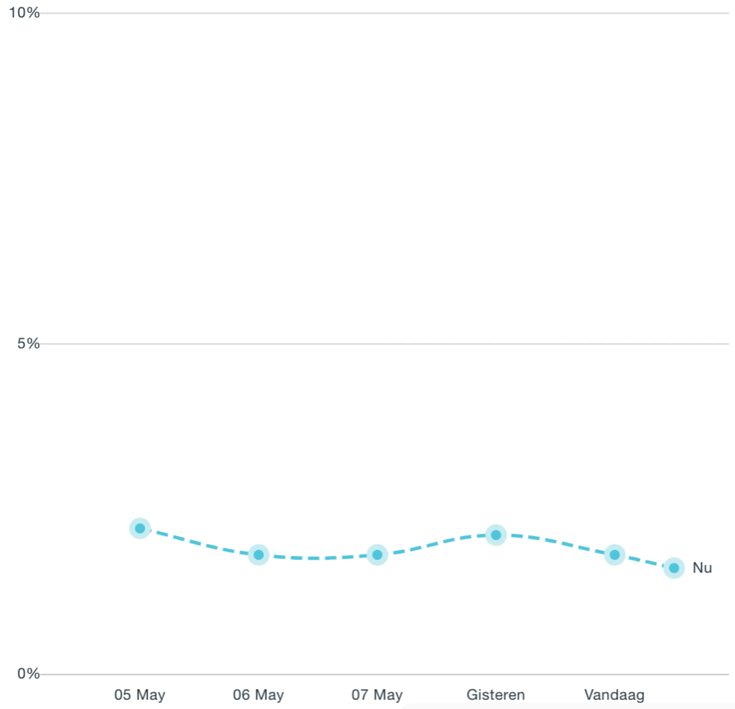
Janssen, J. Dhr. | 14250 | 1954-11-01
Bednummer: 01 | Opnameduur: 1 dagen
Diagnose: *Post-operatief CABG*

Heropname/ Mortaliteit risico **1.0%**

Ondersteuning

Verloop heropname/mortaliteit risico

Compleet | Recent



Top 10 belangrijkste voorspellers

- geassocieerd met een **hoger** dan gemiddeld risico op heropname/mortaliteit
- geassocieerd met een **lager** dan gemiddeld risico op heropname/mortaliteit

Let op: deze gegevens zijn slechts bedoeld om een beeld te geven van voorspellende kenmerken. Verbetering van getoonde waardes leidt niet tot verbetering van de model output.

Voorspeller		Waarde van patiënt	Gemiddelde ontslagen patiënten
HERKOMST AFDELING	🟢	Cardiologie/OK	55.39% had deze herkomst
CREATINE KINASE-MUSCLE/BRAIN (CK-MB) Is het gemeten in de afgelopen 24 uur?	🟢	Ja	gemeten bij 52.78%
CARDIALE OUTPUT Laatste waarde (over de gehele opname)	🟢	7 l/min	laatstgemeten waarde van 5.45 l/min
ADEMHALINGSFREQUENTIE MONITOR Gemiddelde over de laatste 24 uur	🟢	19 /min	gemiddelde van 17.6 /min
ZUURSTOF SATURATIE Maximum waarde over de gehele opname	🔴	98 %	maximum van 99.568 %
PH Laatste waarde (over de gehele opname)	🟢	7.44	laatstgemeten waarde van 7.417
CARDIALE OUTPUT Is het gemeten in de eerste 24 uur?	🟢	Ja	gemeten bij 35.27%

How did we improve it?

Top 10 belangrijkste voorspellers

- geassocieerd met een **hoger** dan gemiddeld risico op heropname/mortaliteit
- geassocieerd met een **lager** dan gemiddeld risico op heropname/mortaliteit

Let op: deze gegevens zijn slechts bedoeld om een beeld te geven van voorspellende kenmerken. Verbetering van getoonde waardes leidt niet tot verbetering van de model output.

Voorspeller		Waarde van patiënt	Gemiddelde ontslagen patiënten
HERKOMST AFDELING	●	Cardiologie/OK	55.39% had deze herkomst
CREATINE KINASE-MUSCLE/BRAIN (CK-MB) Is het gemeten in de afgelopen 24 uur?	●	Ja	gemeten bij 52.78%
CARDIALE OUTPUT Laatste waarde (over de gehele opname)	●	7 l/min	laatstgemeten waarde van 5.45 l/min
ADEMHALINGSFREQUENTIE MONITOR Gemiddelde over de laatste 24 uur	●	19 /min	gemiddelde van 17.6 /min
ZUURSTOF SATURATIE Maximum waarde over de gehele opname	●	98 %	maximum van 99.568 %
PH Laatste waarde (over de gehele opname)	●	7.44	laatstgemeten waarde van 7.417
CARDIALE OUTPUT Is het gemeten in de eerste 24 uur?	●	Ja	gemeten bij 35.27%

Disclaimer on **causality**

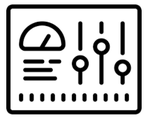
Remove clinically **irrelevant features**

Removed **too complex** feature engineering

Top 10 features **ordered on SHAP** value

Added **additional context** to the feature values

Next steps in interpretability we are working on



Towards an **interpretability** layer 3.0

Dealing with **uncertainty** in the predictions

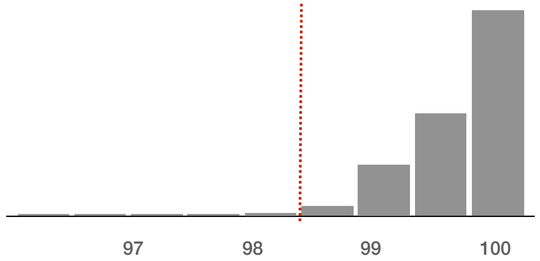
Testing and measuring added value of interpretability panel

Interpretability layer 3.0 provides additional information for the most important features

Predictor	Value for this patient	Average value discharged patients
HERKOMST AFDELING	Cardiologie/OK	55.39% had deze herkomst
CREATINE KINASE-MUSCLE/BRAIN (CK-MB) Is het gemeten in de afgelopen 24 uur?	Ja	gemeten bij 52.78%
CARDIALE OUTPUT Laatste waarde (over de gehele opname)	7 l/min	laatstgemeten waarde van 5.45 l/min
ADEMHALINGSFREQUENTIE MONITOR Gemiddelde over de laatste 24 uur	19 /min	gemiddelde van 17.6 /min
ZUURSTOF SATURATIE Maximum waarde over de gehele opname	98 %	maximum van 99.568 %
PH Laatste waarde (over de gehele opname)	7.44	laatstgemeten waarde van 7.417
CARDIALE OUTPUT Is het gemeten in de eerste 24 uur?	Ja	gemeten bij 35.27%

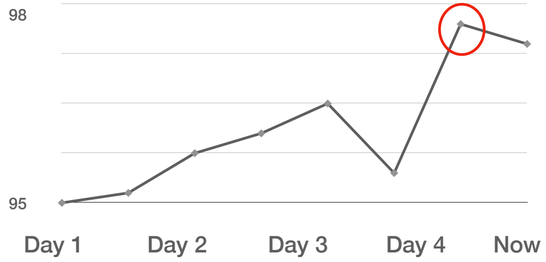
Distribution of values

Oxygen saturation, maximum entire admission

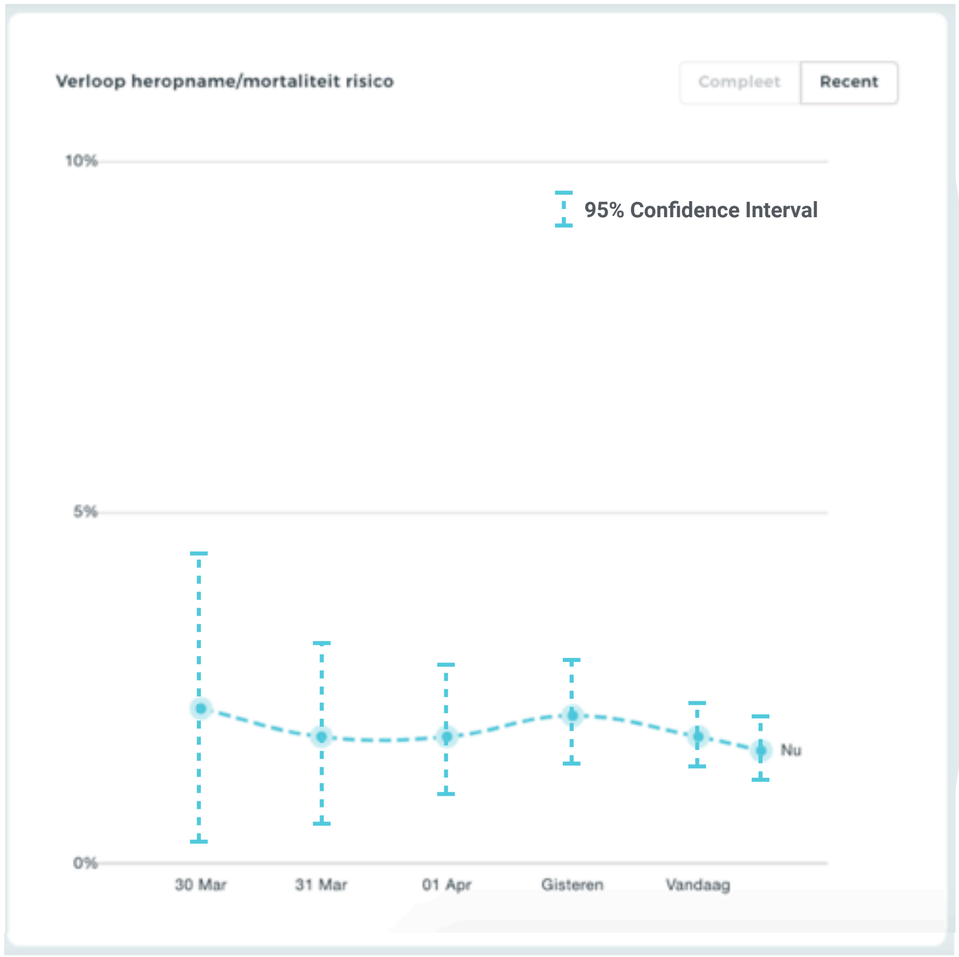


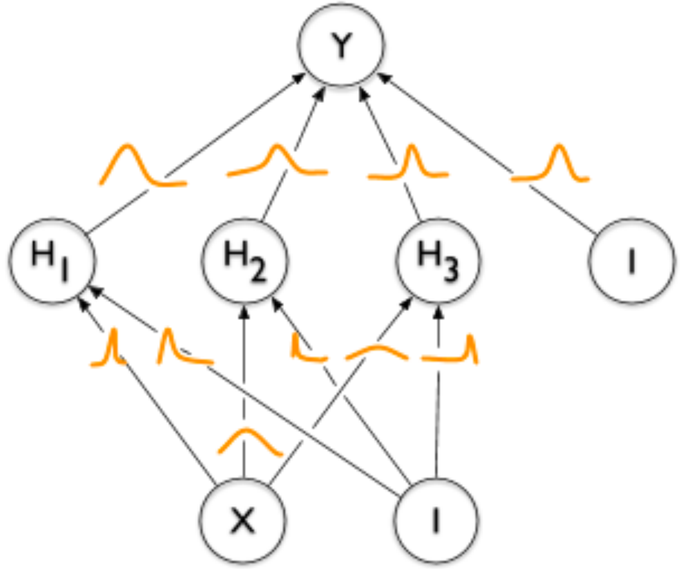
Development over time

Oxygen saturation



Adding uncertainty to the predictions can help doctor asses how to use the model

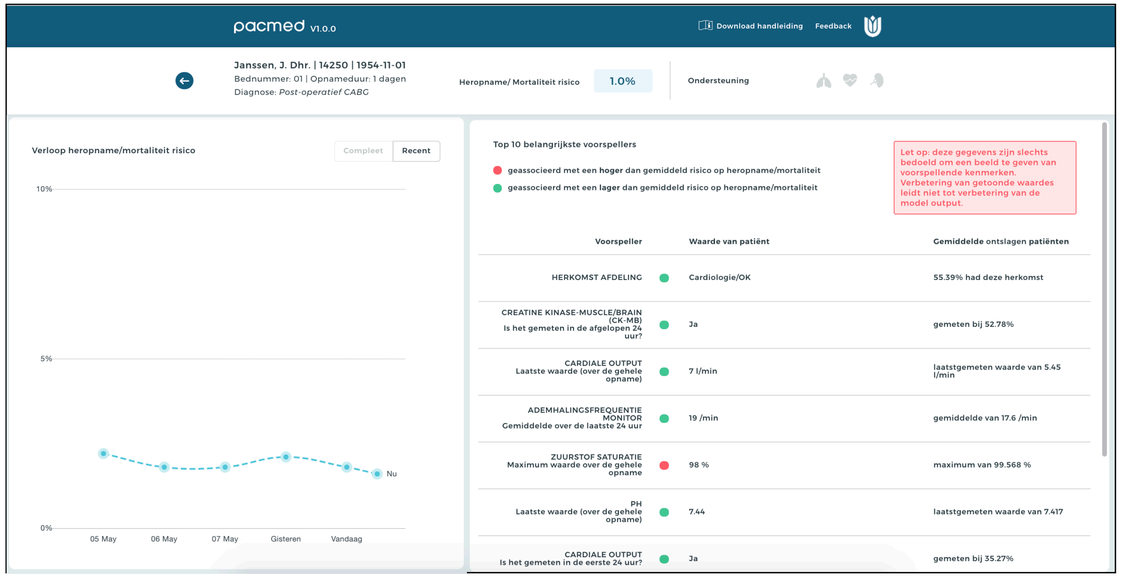




How to ‘scientifically’ test the added value of the interpretability panel?

1.0%

vs.



What is Interpretable? Using ML to Design Interpretable Decision-Support Systems

Owen Lahav[1], Nicholas Mastronarde[2], and Mihaela van der Schaar[1,3,4]

[1]University of Oxford, [2]University at Buffalo, [3]University of California Los Angeles (UCLA),

[4]Alan Turing Institute

We should be tackling interpretability for medical machine learning on different levels!

Transparency of the methodology and scope of the model

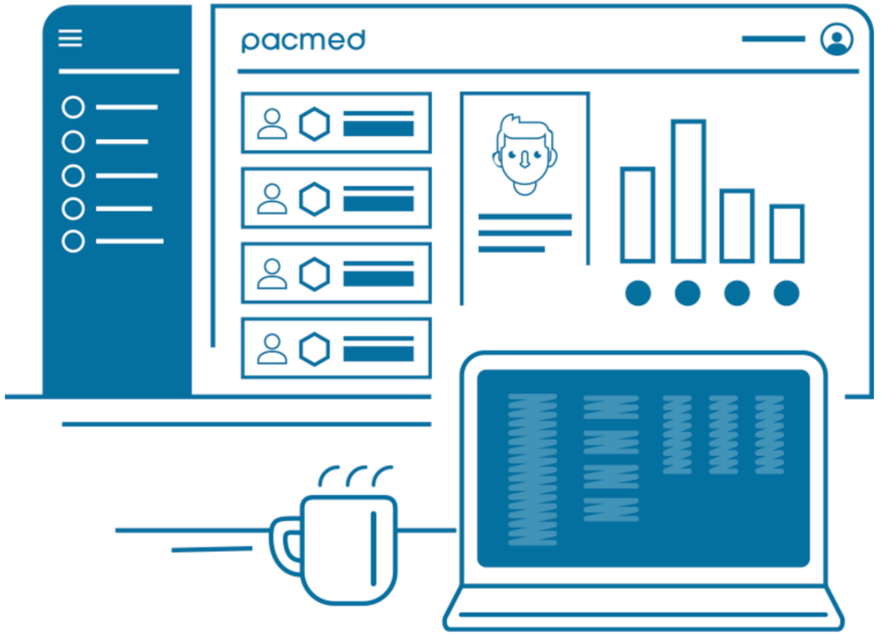
Global **understandability** of the workings of the **model**

Specific **interpretability** of the model’s **predictions**

pacmed

ECP

Platform voor de
InformatieSamenleving



Thank you!

david.ruhe@gmail.com