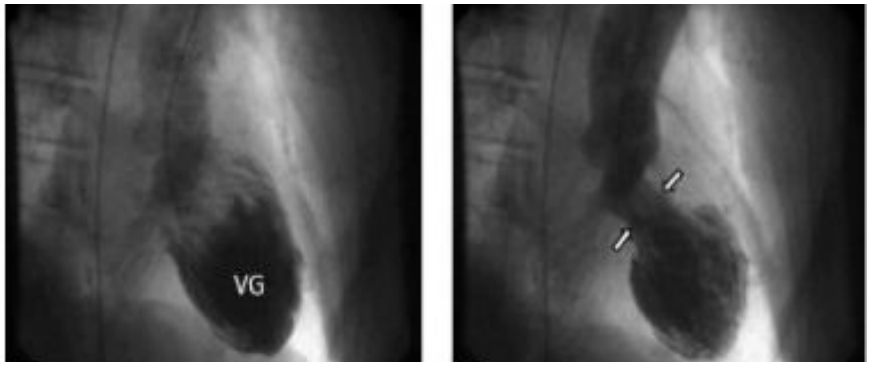
Takotsubo Syndrome: Prediction of hospitalization outcomes

Python Machine Learning Project
July 10th 2021



Daisuke KUWABARA
Nesrine BENANTEUR

OUTLINE

Introduction: Reminders of basic heart physiology to understand the context

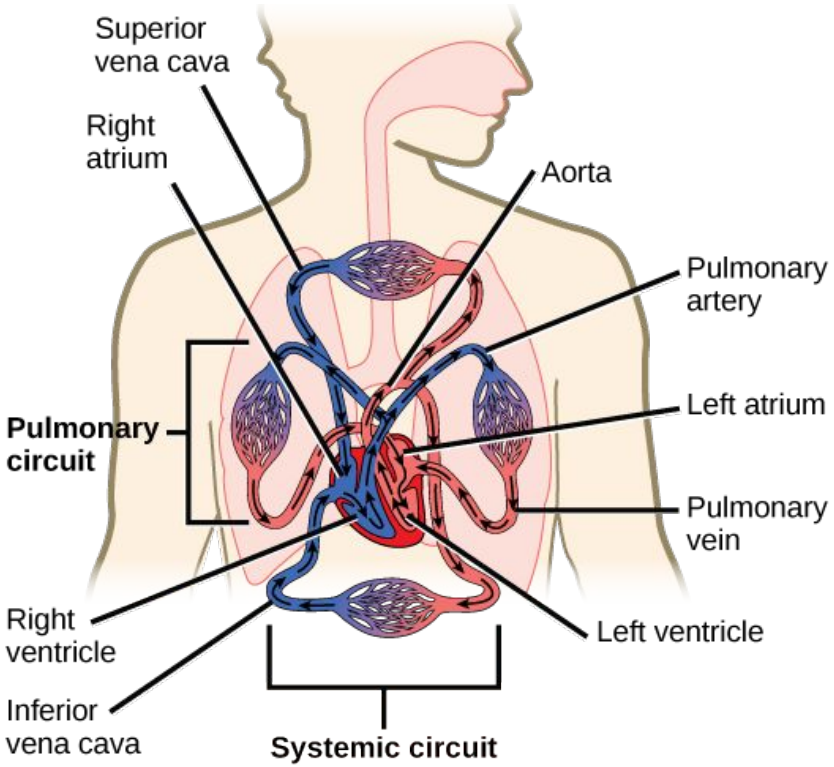
Takotsubo Syndrome

Description of the cohorts and the variables used for prediction

Algorithms predictions

Conclusion: Perspectives and improvements

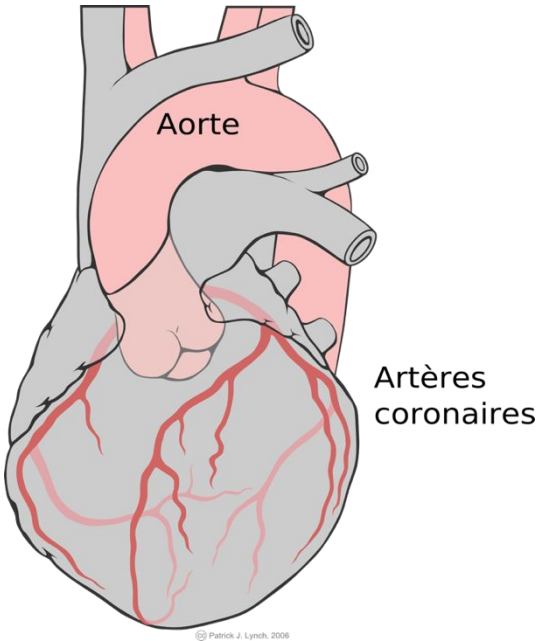
GENERAL HEART PHYISOLOGY TO UNDERSTAND THE CONTEXT



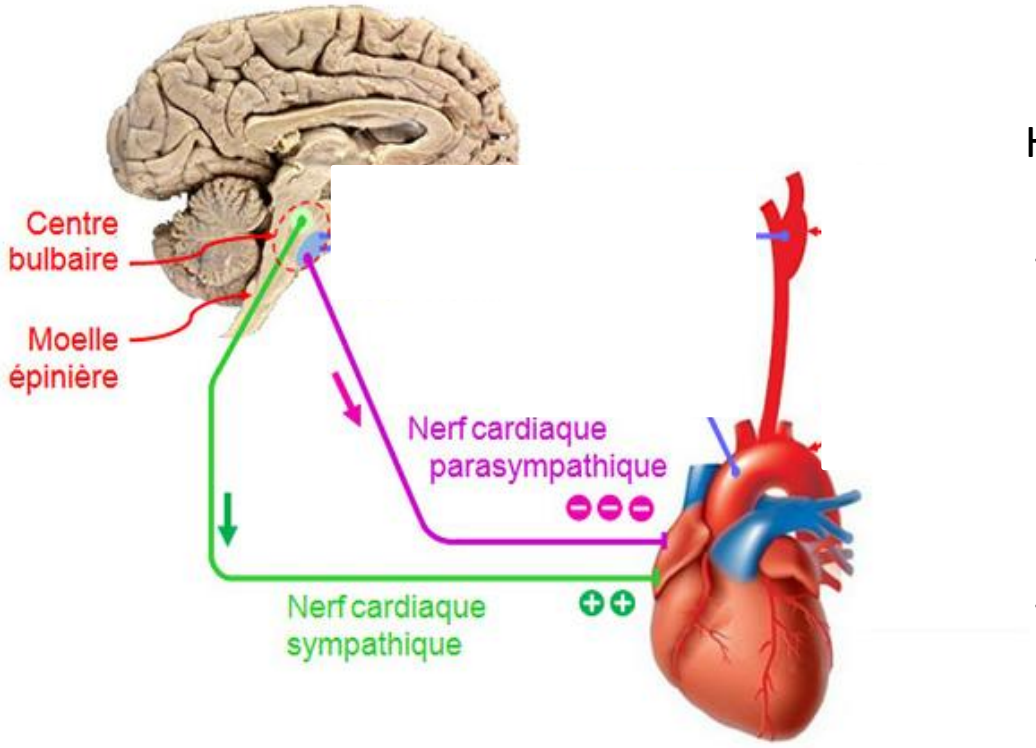
Circulatory system: network consisting of blood, blood vessels, and the heart.

This network supplies tissues in the body with oxygen and other nutrients, transports hormones, and removes unnecessary waste products.

The heart is made of specialized cardiac muscle tissue called **myocardium**.
It is made of specific cells called **cardiomyocytes**.



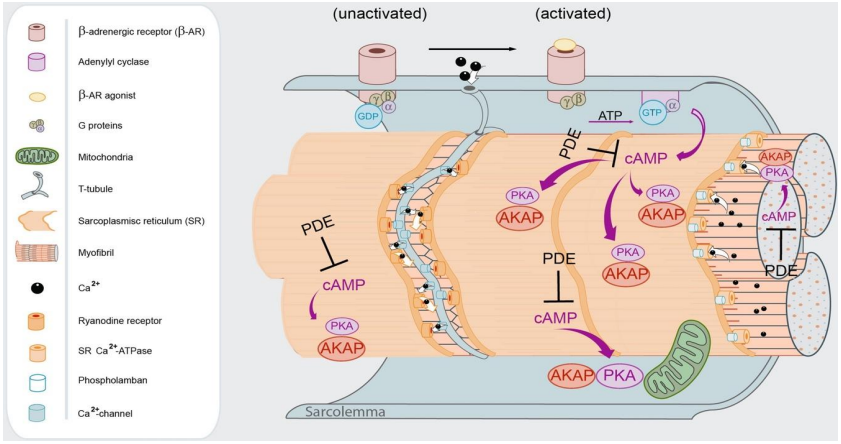
GENERAL HEART PHYISOLOGY TO UNDERSTAND THE CONTEXT



How does the heart beat?

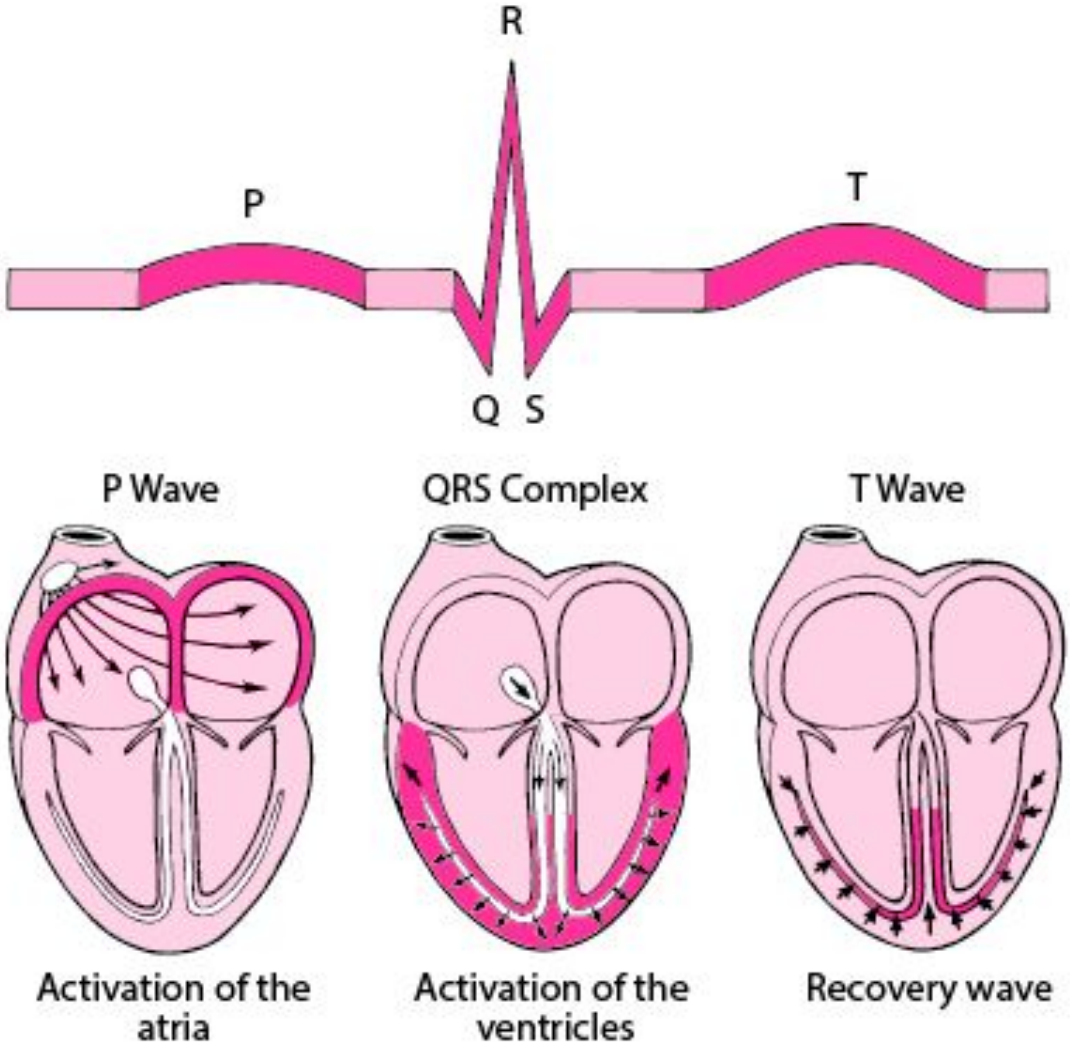
- Nerve conduction: sympathic cardiac nerve (increases the heartbeat) or parasympathic cardiac nerve (decreases the heartbeat).

- Biochemicals: catecholamines (adrenaline, noradrenaline) are released to increases the contraction of the cardiomyocytes.

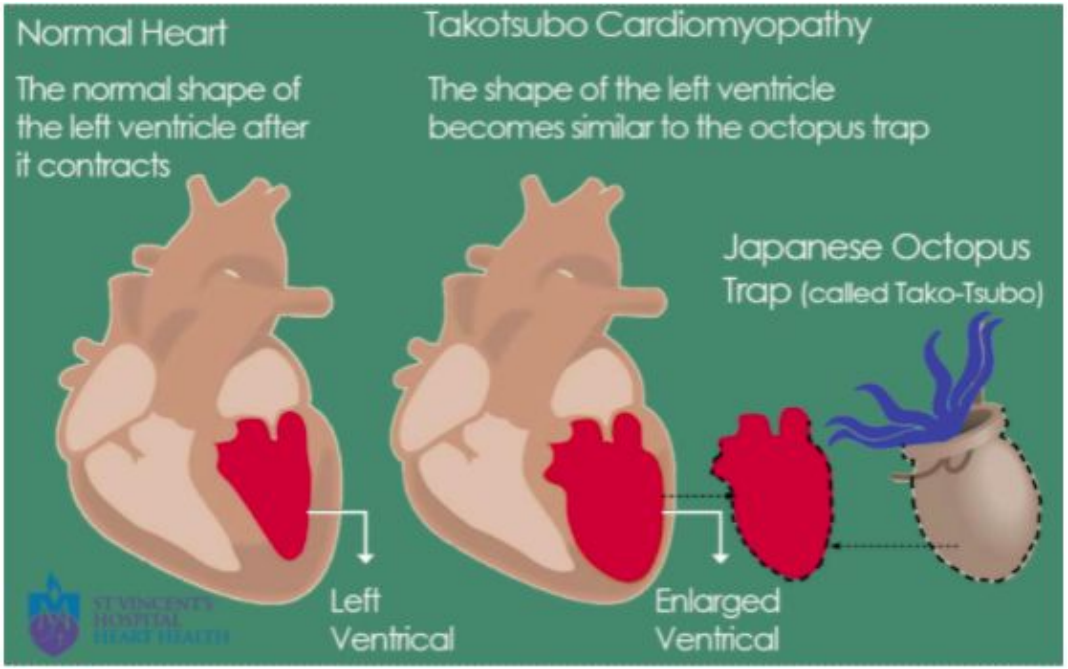


GENERAL HEART PHYISOLOGY TO UNDERSTAND THE CONTEXT

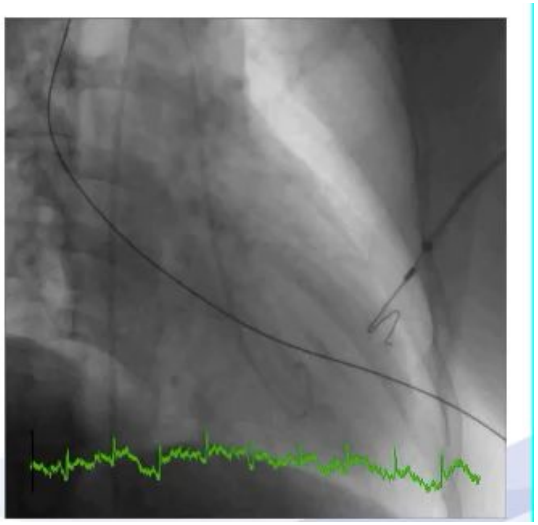
- The heart rate can be analyzed through an exam called ECG.

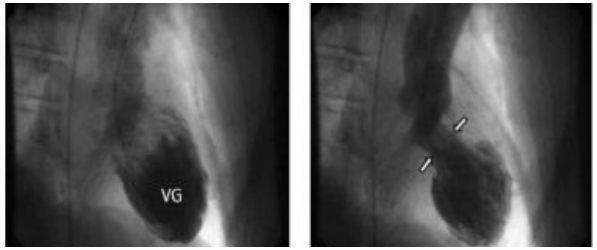


TAKOTSUBO SYNDROME



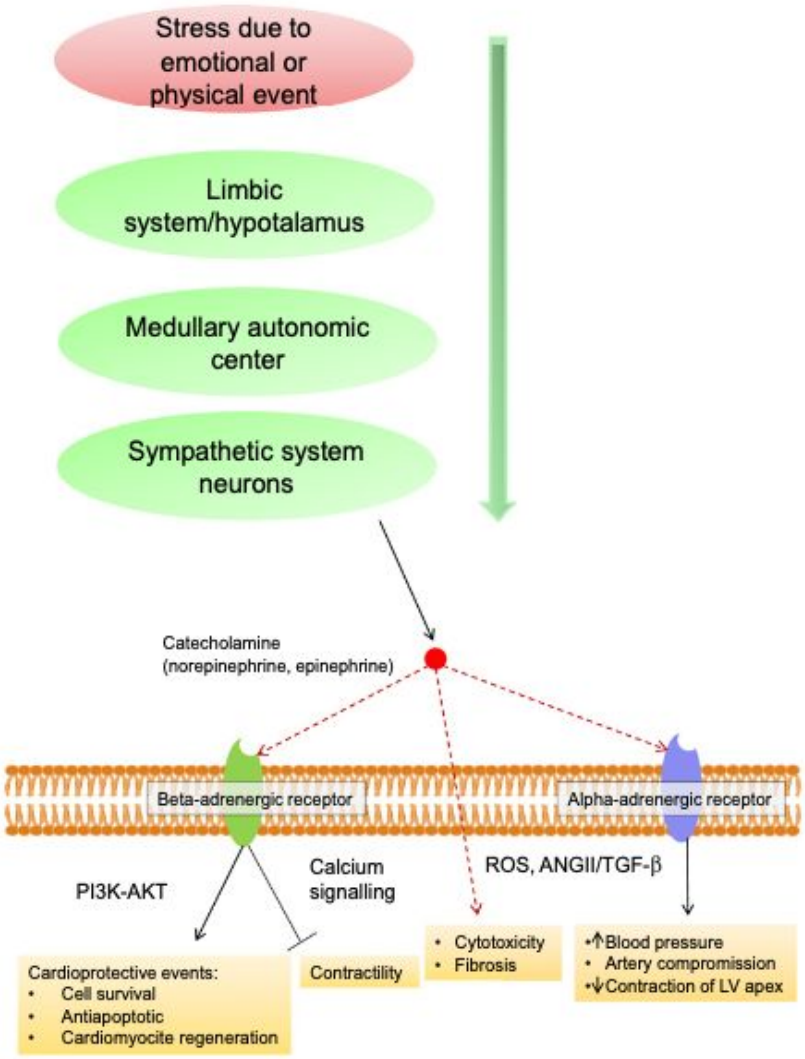
Takotsubo cardiomyopathy is a transient weakening of the left ventricle, the heart's main pumping chamber.
Usually: result of severe emotional or physical stress.





TAKOTSUBO SYNDROME

Possible causes:



Lots of hypothesis today, but articles seem to all incriminate a catecholamine toxicity.

High prevalence in middle-aged women seem to show that estrogen Deficiency due to menopause might predispose them to the disease.

TAKOTSUBO SYNDROME

What happens during a Takotsubo event?

Left ventricle pumps blood less efficiently → Amount of blood going out of the heart decreases significantly (LVEF↘) → Organs receive less blood and suffer from the lack of oxygen and nutrients.

Why chest pains?

- Coronary arteries receive a lot less blood than usual.

⟶ Cardiomyocytes aren't fed in oxygen and nutrients, and die.

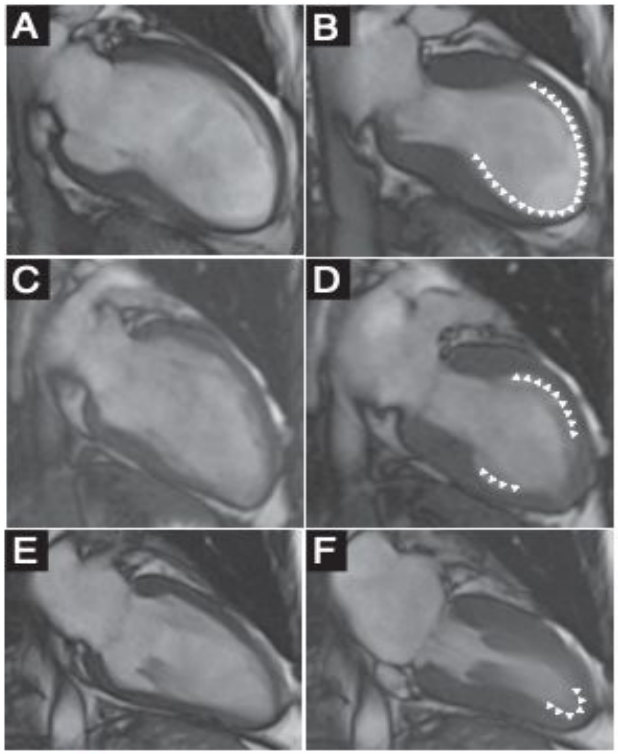
Consequences?

- Cardiogenic shock: the blood pumped into the whole body cannot meet the other organs' needs, provoking damages to the liver, kidneys from lack of oxygen, which can be permanent.
- Rhythmic abnormalities
- Thrombus due to residual blood in the left ventricle.
- Death

TAKOTSUBO SYNDROME

Criterion for Takotsubo diagnosis:
(Mayo Clinic, 2004)

1. Transient hypokinesis, akinesis, or dyskinesis of the left ventricular mid-segments with or without apical involvement; the regional wall motion abnormalities extend beyond a single epicardial vascular distribution; a stressful trigger is often, but not always present.



Diversity of left ventricle contraction patterns

A, C, E: Diastole – heart relaxing and filling up with blood.
B, D, F: Systole – heart contraction

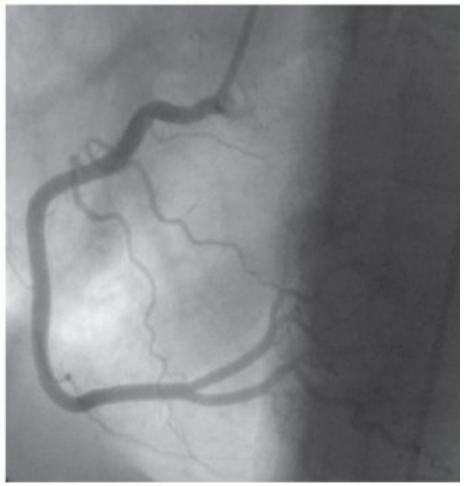
TAKOTSUBO SYNDROME

Criterion for Takotsubo diagnosis:
(Mayo Clinic, 2004)

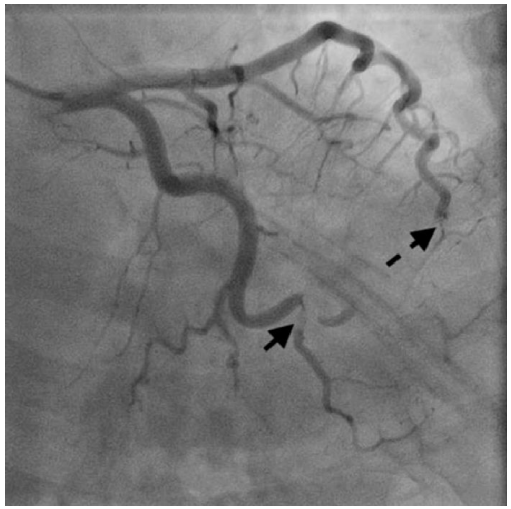
1. Transient hypokinesis, akinesis, or dyskinesis of the left ventricular mid-segments with or without apical involvement; the regional wall motion abnormalities extend beyond a single epicardial vascular distribution; a stressful trigger is often, but not always present.

2. Absence of obstructive coronary disease or angiographic evidence of acute plaque rupture.

Normal coronary arteries on the left side of the heart



Normal coronary arteries on the right side of the heart



Coronary arteries obstructed

TAKOTSUBO SYNDROME

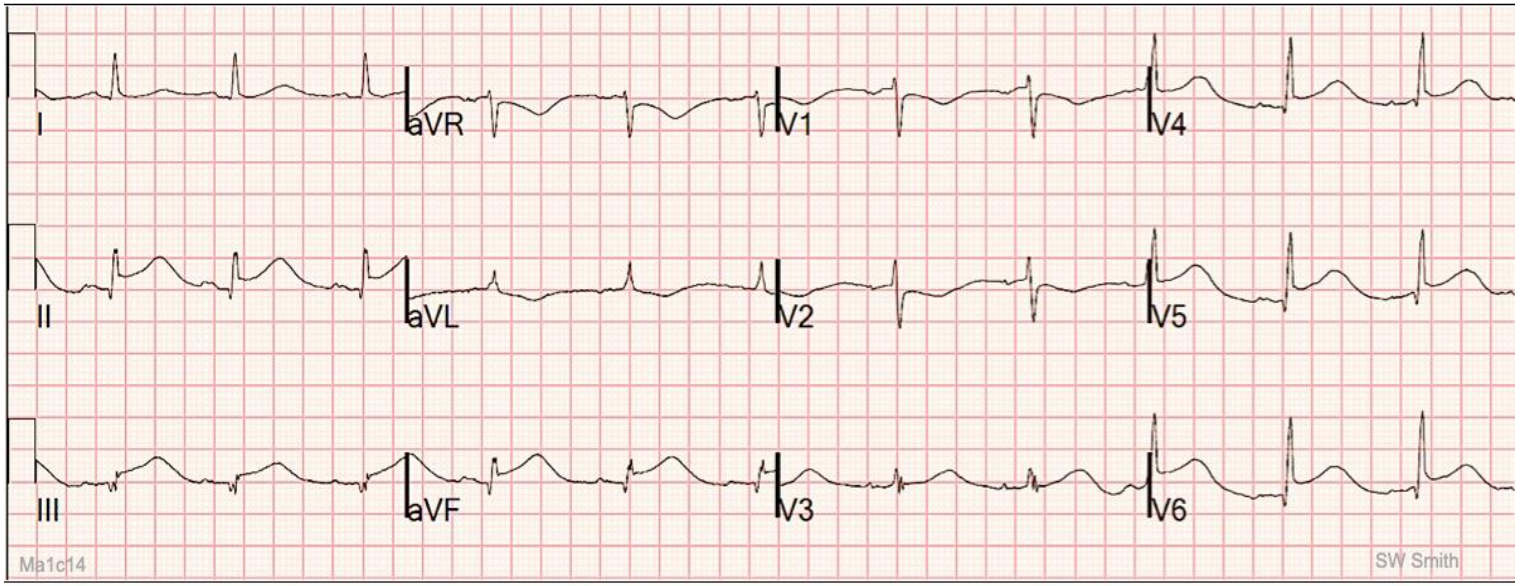
Criterion for Takotsubo diagnosis:
(Mayo Clinic, 2004)

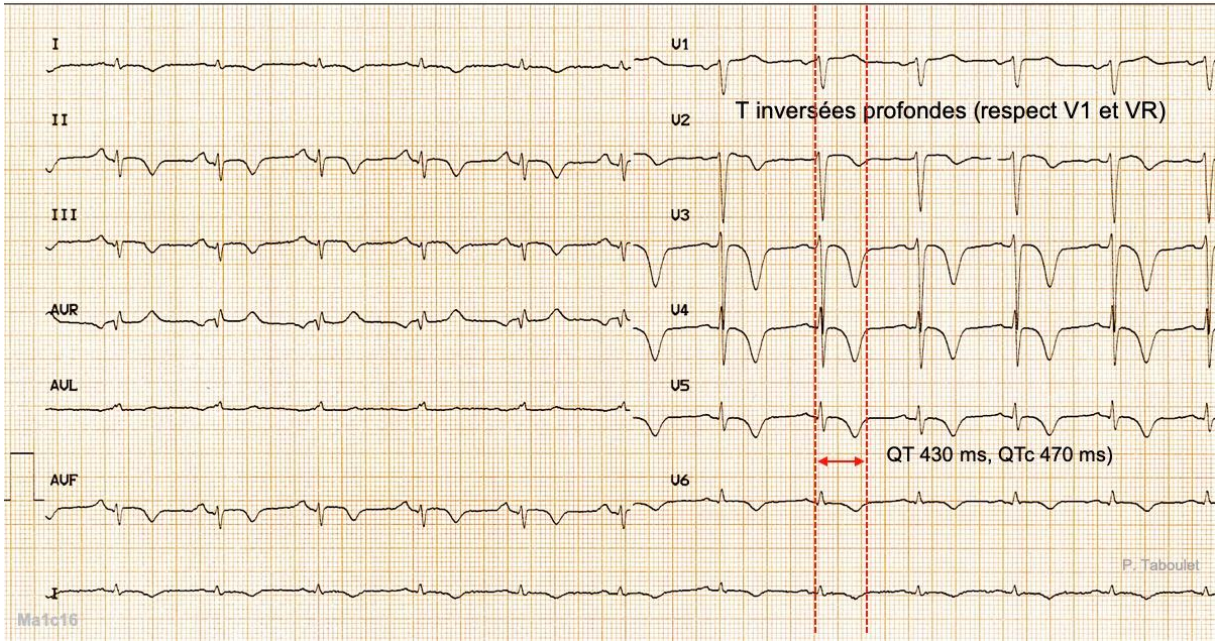
1. Transient hypokinesis, akinesis, or dyskinesis of the left ventricular mid-segments with or without apical involvement; the regional wall motion abnormalities extend beyond a single epicardial vascular distribution; a stressful trigger is often, but not always present.

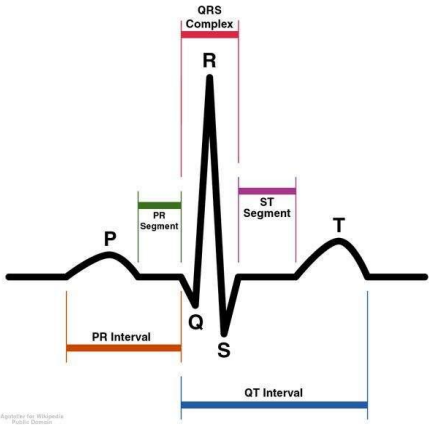
2. Absence of obstructive coronary disease or angiographic evidence of acute plaque rupture.

3. New electrocardiographic abnormalities (either ST-segment elevation and/or T-wave inversion) or modest elevation in cardiac troponin.

TAKOTSUBO SYNDROME







TAKOTSUBO SYNDROME

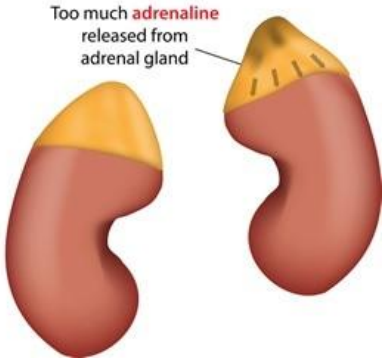
Criterion for Takotsubo diagnosis:
(Mayo Clinic, 2004)

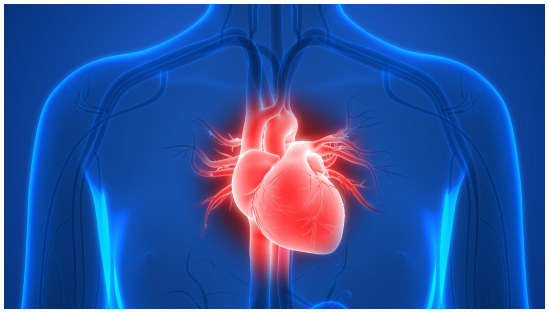
1. Transient hypokinesis, akinesis, or dyskinesis of the left ventricular mid-segments with or without apical involvement; the regional wall motion abnormalities extend beyond a single epicardial vascular distribution; a stressful trigger is often, but not always present.

2. Absence of obstructive coronary disease or angiographic evidence of acute plaque rupture.

3. New electrocardiographic abnormalities (either ST-segment elevation and/or T-wave inversion) or modest elevation in cardiac troponin.

4. Absence of: a. Pheochromocytoma b. Myocarditis

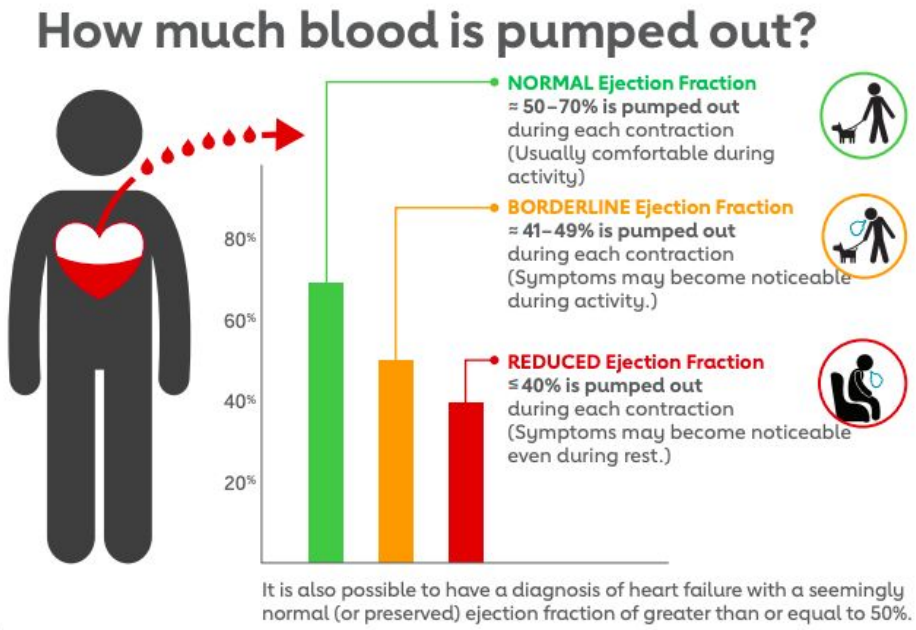




TAKOTSUBO SYNDROME: PROBLEMATIC

▷ **Project goals:**

- Predict the in-hospital complications
- Predict the kind of heart failure to anticipate the medical care needed.



DECRIPTION OF THE COHORT AND VARIABLES USED FOR PREDICTION

Who?

Takotsubo patients between 2015 and 2021 from the CHU de Toulouse.

What?

Variables selected by reading research articles on the short and long-term prognosis or diagnosis of the disease.

How?

Reading the patients records and extracting the data patient by patient.

DDN	âge	Date hospit initiale	Homme	poids	taille	IMC (kg/ ci	ATCD dépression/	ATCD psychiatrique	patho neurologiques	HTA	Dyslipidém	Tabac	Diabète	IRC	AVC/AIT	ATCD Canc	Can
3/5/1955	66	5/3/2015	1	65	1,78	20,52	0	0	1	0	0	1	0	0	0	0	
1/6/1964	57	3/2/2016	0	58	1,6	22,66	0	0	0	1	0	1	0	0	0	0	
30/6/1940	81	10/2/2016	0	50	1,53	21,36	0	0	0	1	0	0	1	0	0	0	
20/10/1962	58	25/2/2016	1	60	1,78	18,94	0	0	1	0	0	1	0	0	0	0	
19/8/1939	81	1/3/2016	0	70	1,56	28,76	0	0	0	1	0	0	0	0	0	1	
6/11/1929	91	28/2/2016	0	57	1,54	24,03	0	0	1	1	0	0	0	0	0	0	
15/2/1935	86	6/4/2016	0	65			0	0	0	1	0	0	1	0	0	0	
20/4/1944	77	20/4/2016	0	80	1,56	32,87	0	0	1	1	0	0	1	0	1	1	
29/5/1991	30	24/4/2016	0	63	1,58	25,24	0	0	0	0	0	0	0	0	0	0	
3/9/1958	62	2/5/2016	1	55	1,68	19,49	0	1	0	0	1	1	1	0	0	0	
11/5/1928	93	18/5/2016	0	55	1,55	22,89	0	0	0	0	0	1	0	0	0	0	
12/3/1949	72	2/6/2016	0	80	1,55	33,30	0	0	0	1	1	0	0	0	0	0	
11/8/1939	81	3/6/2016	0	98	1,69	34,31	0	0	0	1	1	1	1	0	0	0	
12/10/1934	86	13/7/2016	0	67	1,7	23,18	0	0	1	1	0	0	0	0	1	1	
27/6/1949	72	23/7/2016	0	65	1,69	22,76	0	0	0	1	0	0	0	0	0	0	
12/3/1931	90	23/7/2016	1	73	1,7	25,26	0	0	0	0	0	0	0	0	0	0	
27/7/1955	65	26/7/2016	0	82	1,67	29,40	1	0	0	1	0	1	0	0	0	0	
30/7/1955	65	17/8/2016	0	58	1,63	21,83	0	0	0	1	0	1	1	0	0	0	
26/5/1933	88	20/9/2016	0	57	1,55	23,73	0	0	0	1	0	0	1	0	0	0	
31/12/1941	79	27/9/2016	0	74	1,5	32,89	0	0	0	1	1	0	0	0	0	0	
27/6/1952	69	5/10/2016	0	58	1,6	22,66	1	0	0	1	0	0	0	0	0	0	
1/7/1935	86	12/10/2016	0	60	1,58	24,03	0	0	0	1	1	0	0	0	0	0	
19/1/1944	77	19/10/2016	0	55	1,55	22,89	0	0	0	1	1	0	0	0	0	0	
20/2/1943	78	13/11/2016	0	70	1,76	22,60	0	0	0	0	0	0	1	0	0	0	
31/1/1939	82	29/11/2016	0	68	1,64	25,28	1	0	1	1	0	0	0	0	1	0	
15/1/1949	72	28/12/2016	1	70	1,73	23,39	0	0	0	0	0	1	0	0	0	0	
27/12/1944	76	4/1/2017	0	49	1,6	19,14	0	0	0	0	0	1	0	0	0	0	
3/10/1942	78	20/1/2017	0	60	1,54	25,30	1	0	0	0	0	0	0	0	1	0	
3/5/1973	48	26/1/2017	1	84	1,86	24,28	0	0	0	0	0	0	0	0	0	0	
20/9/1998	22	27/1/2017	0	59	1,6	23,05	0	0	0	0	0	0	0	0	0	0	
5/12/1946	74	8/2/2017	0	75	1,58	30,04	0	0	0	1	1	0	0	0	0	1	
23/7/1931	89	27/2/2017	0	63	1,53	26,91	0	0	0	1	0	0	0	0	0	0	
16/2/1949	72	27/2/2017	0	67	1,59	26,50	0	0	0	0	1	1	0	0	0	1	
31/12/1949	71	14/5/2017	0	64	1,69	22,41	0	0	1	0	1	1	0	0	0	0	
11/6/1959	62	11/6/2017	0	65	1,68	23,03	1	0	0	0	1	0	0	0	0	0	
19/4/1924	97	14/6/2017	0	53	1,6	20,70	0	0	0	1	0	0	0	0	0	0	
8/2/1930	91	20/6/2017	0	60	1,57	24,34	0	0	0	0	0	0	0	0	0	1	

DECRIPTION OF THE COHORT AND VARIABLES USED FOR PREDICTION

DEMOGRAPHICS

Age, BMI, Sex.

STRESS FACTOR

- Emotional
- Physical

MEDICAL HISTORY

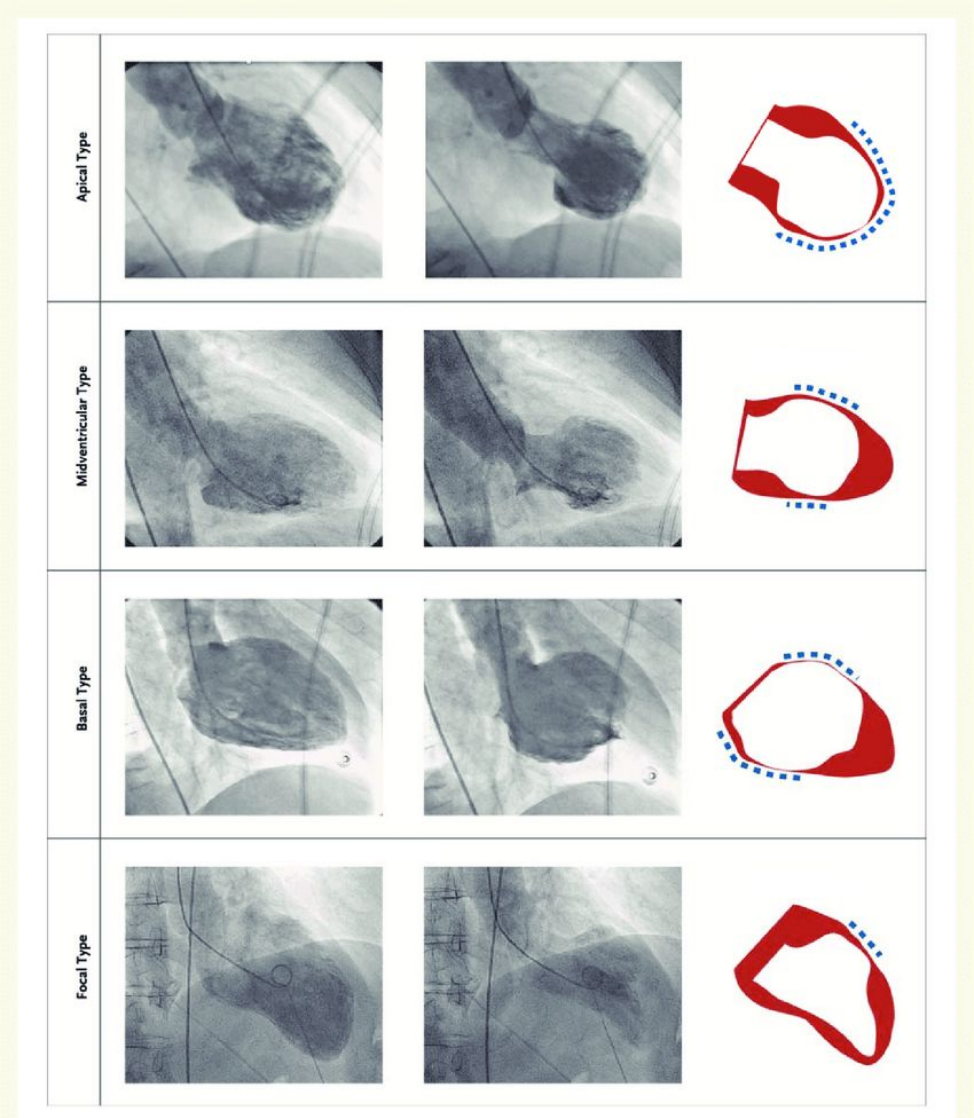
- Depression or Anxiety
- Psychiatric disorders
- Neurological diseases
- Hypertension
- Dyslipidemia
- Smoking
- Diabetes
- Chronic Renal Failure
- Stroke or Transient Ischemic Attack (TIA)
- Cancer history or active Cancer
- Chronic Obstructive Pulmonary Disease (COPD) or asthma
- Alcoholism
- Cardiac diseases history

TREATMENTS BEFORE/AFTER EVENT

- Beta Blockers
- Angiotensin Converting Enzyme Inhibitor
- Angiotensin II receptor blockers
- Aspirin
- Antiplatelets drugs
- Oral anticoagulation drugs
- Statins
- Anti-depressants and anxiolytics

DECRIPTION OF THE COHORT AND VARIABLES USED FOR PREDICTION

ANATOMY



Apical type

Medioventricular type

Basal type

Focal type

DECRIPTION OF THE COHORT AND VARIABLES USED FOR PREDICTION

HAEMODYNAMICS

- Left Ventricle Ejection Fraction (LVEF)
  - entry LVEF
  - out LVEF (to predict)
- ECG abnormalities (ST +, long QT, T-)

BIOMARKERS

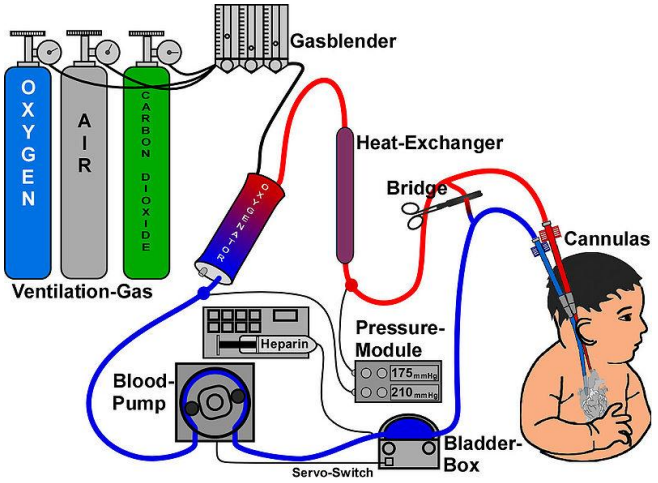
- Troponin T (entry and peak)
- NT pro-BNP
- CRP

CORONAROGRAPHY

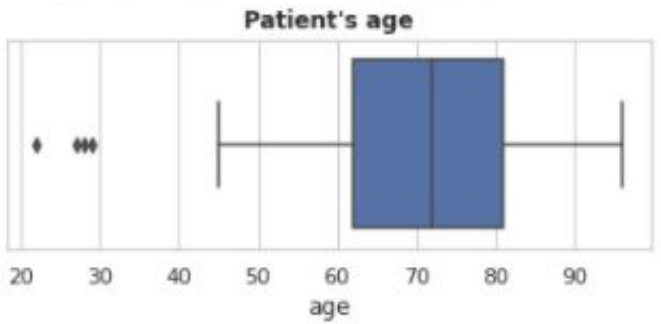
- Harm
- Healthy

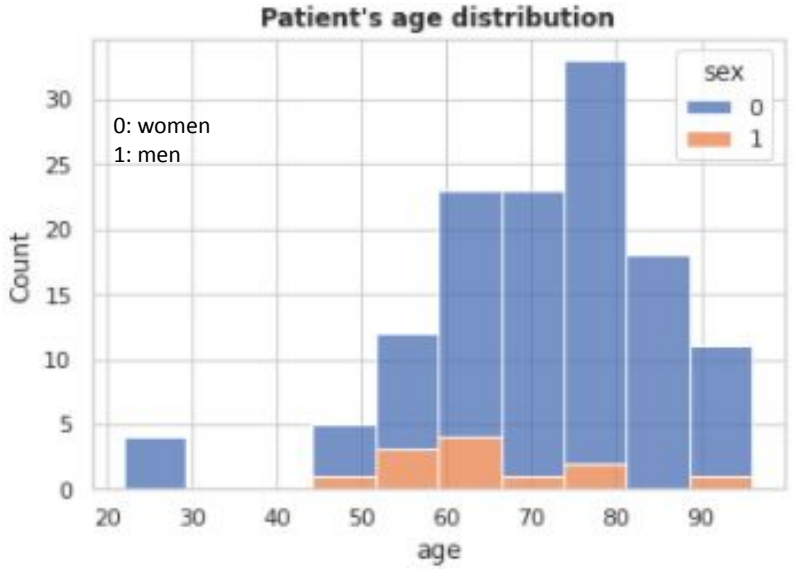
IN-HOSPITAL COMPLICATIONS (to predict)

- Heart Failure
- Right Ventricle Harm
- Ventricular arrhythmia
- Left ventricle thrombus
- Cardiogenic shock
- ECMO (Extracorporeal membrane oxygenation)
- Death



DATA EXPLORATION



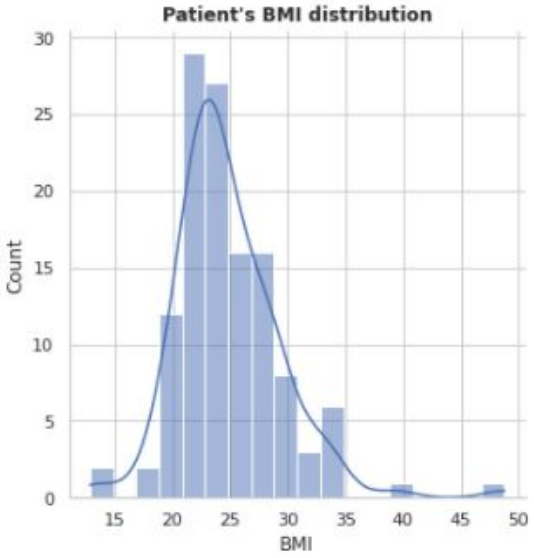


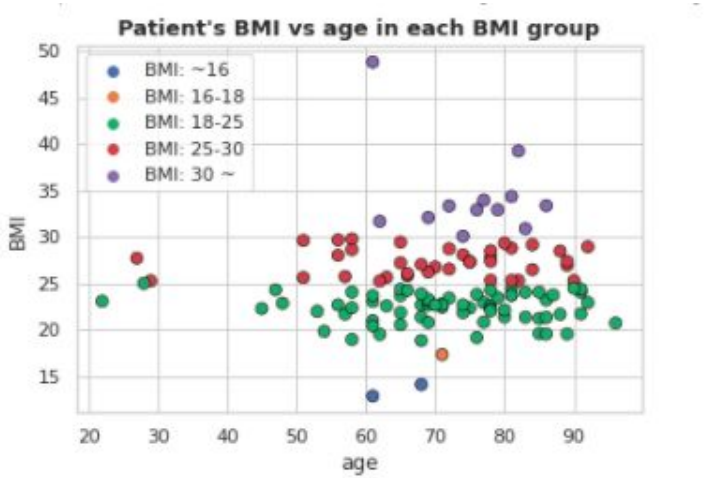
Here above you can see our cohort's age distribution.
Most of our patients are women, and the average age is around 71 years old.

DATA EXPLORATION

BMI Category

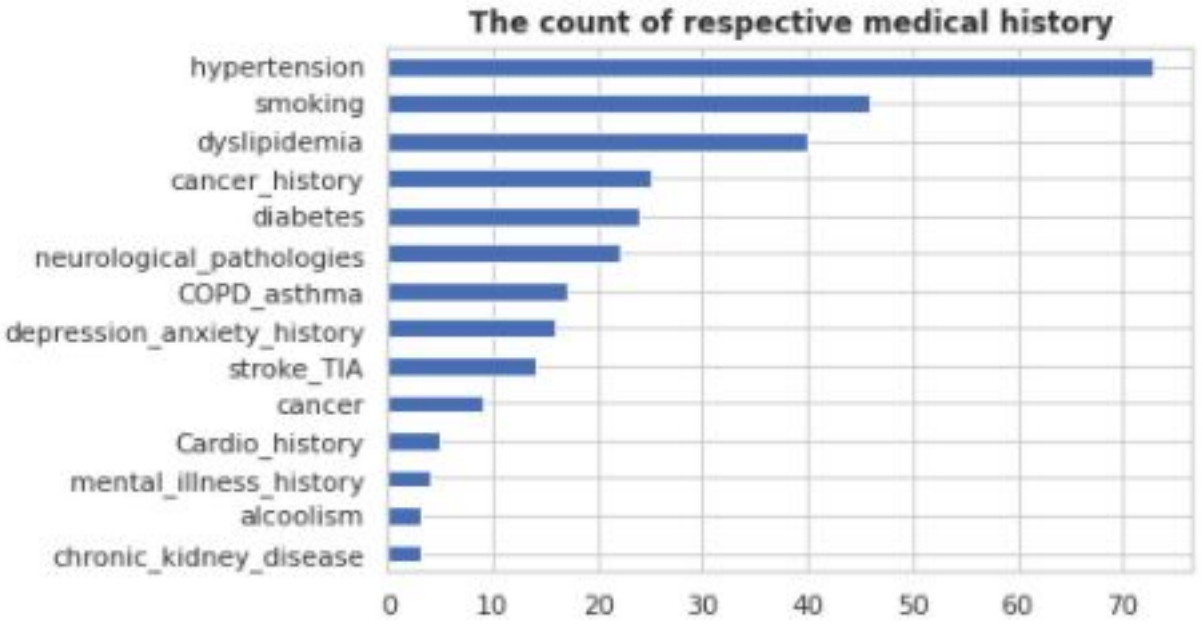
Value	Category	Groups
~ 16	denutrition	1
16-18	underweight	2
18-25	normal	3
25-30	overweight	4
30 ~	obesity	5





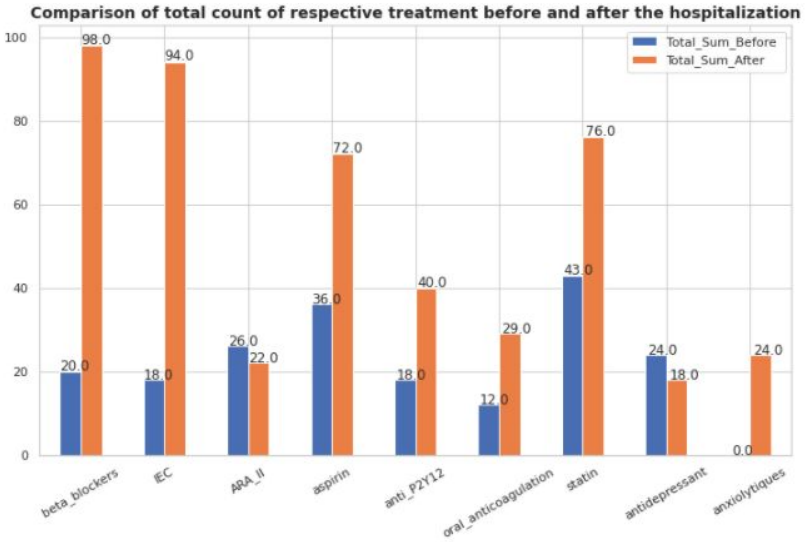
Here above you can see our cohort's BMI distribution.
Most of our patients have a normal BMI.

DATA EXPLORATION



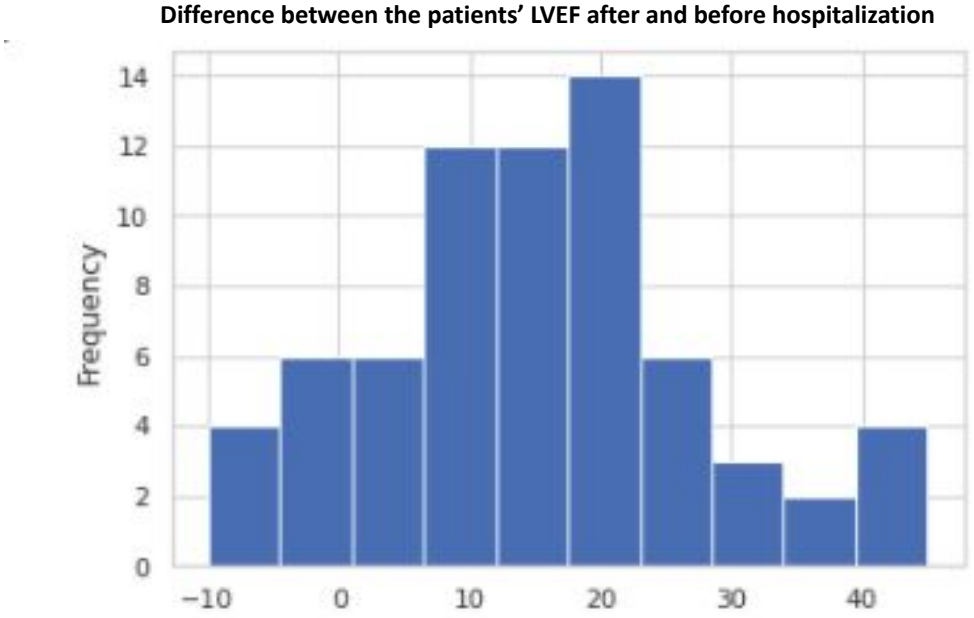
Here above you can see our cohort's medical history.
Most of our patients have a history of hypertension and smoking.
We can also find history of dyslipidemia, cancer, diabetes, neurological pathologies (mostly degenerative diseases), COPD or asthma, depression or anxiety, stroke or TIA, mental illnesses (excluding depression and anxiety), alcoholism, chronic kidney disease, cardiovascular disease and alcoholism.

DATA EXPLORATION



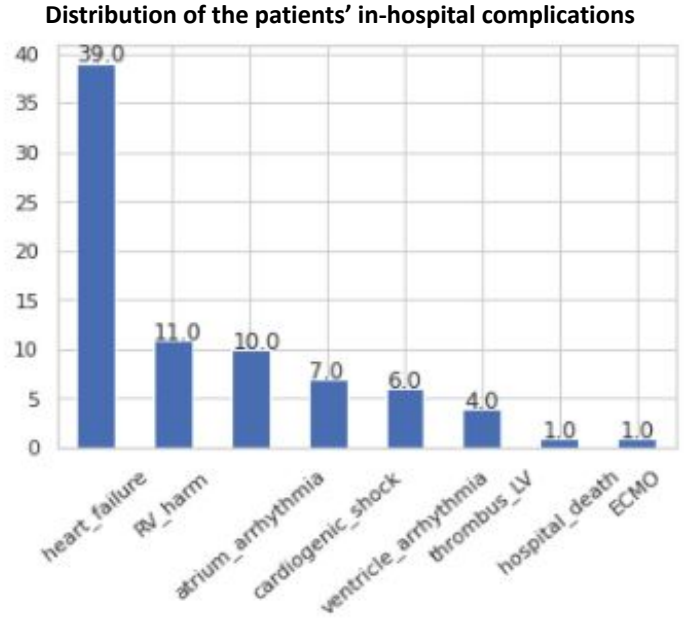
Here above you can find our cohort's treatments, before and after the Takotsubo event.
- As you can see, there a significant rise in the prescription of cardioprotective molecules like beta blockers, Angiotensin Converting Enzyme Inhibitor which lower the heartbeat and the blood pressure, respectively. We wan see the same tendency with treatments like aspirin, anti P2Y12, and oral anti-coagulation drugs which prevents blood clots, a very prevalent complication found in Takotsubo patients.
- The slight drop Angiotensin II receptor blockers might be explained by the rise in the use of Angiotensin Converting Enzyme Inhibitor, which basically has the same effect as Angiotensin II receptor blockers.
- The rise in the prescription of statins might be due to the discovery of coronary disease in patients on which we performed a coronarography.
- The slight drop in antidepressant prescription and rise in anxiolytics might be due to an adjustment In treatments: indeed, the effect of new anti-depressant takes week to be observable, while anxiolytics have an immediate effect.

DATA EXPLORATION



Here above you can find our cohort's LVEF evolution's distribution.
The x axis represents the difference between the LVEF after and before hospitalization (in percentage).
As you can see, most of our patients can see their LVEF increasing after their stay at the hospital.
The negative part represents the proportion of patients for which the LVEF decreased after hospitalization. This might be a sign of severe heart failure, and the patients are usually followed up very seriously after hospitalization.

DATA EXPLORATION



Here above you can find our cohort's in-hospital complications distribution.
Heart failure is the most prevalent complication here, while extreme treatment like the use of ECMO, or death are less likely to happen, which is the tendency observed in other cohorts studied in different publications.

ALGORITHM PREDICTION

MISSING DATA

by mice package R

visualizing and more

PYCARET

OPTUNA

SHAP

MICE (Multivariate Imputation by chained equations)

Multiple Imputation by Chained Equations is a robust, informative method of dealing with missing data in datasets. The procedure “fils in” missing data through an iterative series of predictive models.

Data Leakage:

MICE is particularly useful if missing values are associated with the target variable in a way that introduces leakage. For instance, let’s say you wanted to model customer retention at the time of sign up. A certain variable is collected at sign up or 1 month after sign up. The absence of that variable is a data leak, since it tells you that the customer did not retain for 1 month.

Confidence Intervals:

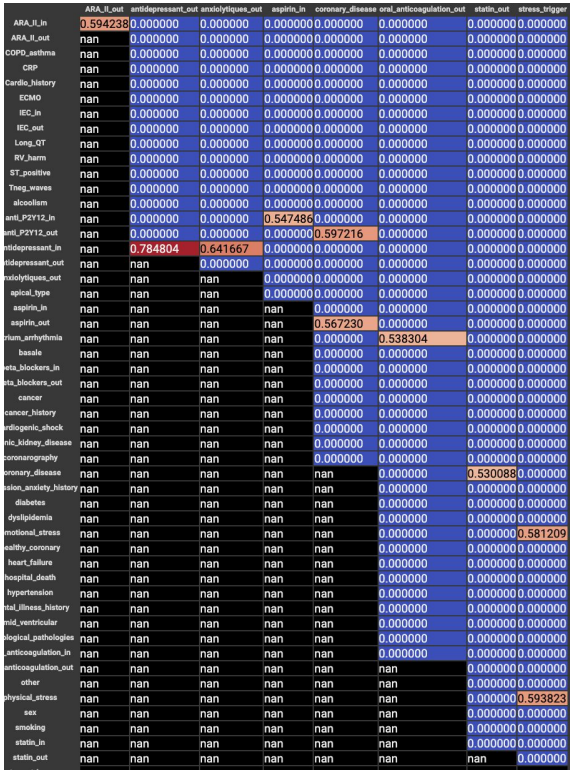
MICE can be used to impute missing values, however it is important to keep in mind that these imputed values are a prediction. Creating multiple datasets with different imputed values allows you to do two types of inference:

- Imputed Value Distribution: A profile can be built for each imputed value, allowing you to make statements about the likely distribution of that value.
- Model Prediction Distribution: With multiple datasets, you can build multiple models and create a distribution of predictions for each sample. Those samples with imputed values which were not able to be imputed with much confidence would have a larger variance in their predictions.

Checking correlation before and after imputation gives an insight of how imputation changes the each distribution of variable and how these variables correlates to each other.

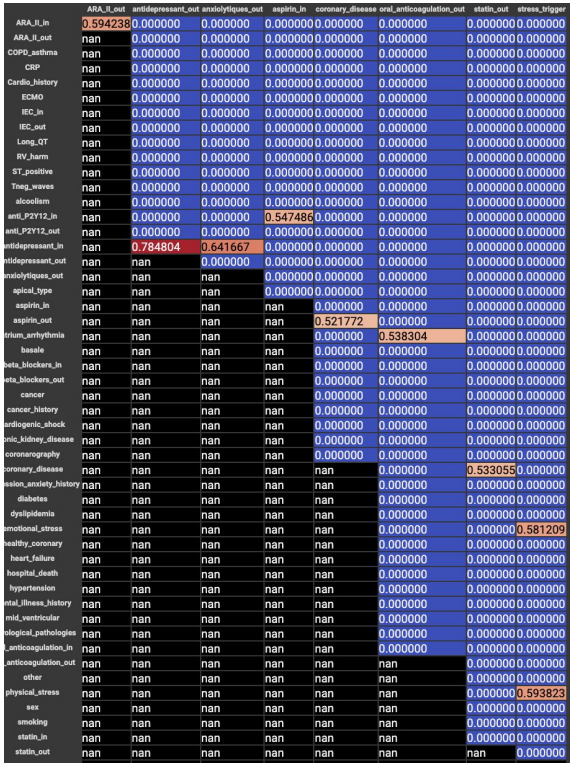
MACHINE LEARNING AND PREDICTION

Correlation Before Imputation

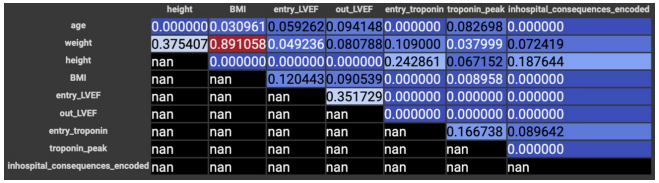


	ARA_II_out	antidepressant_out	anxiolytiques_out	aspirin_in	coronary_disease	oral_anticoagulation_out	statin_out	stress_trigger
ARA_II_in	0.594238	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ARA_II_out	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
COPD_asthma	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
CRP	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Cardio_history	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ECMO	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
IEC_in	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
IEC_out	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Long_QT	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
RV_harm	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ST_positive	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Tneg_waves	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
alcoolism	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
anti_P2Y12_in	nan	0.000000	0.000000	0.547486	0.000000	0.000000	0.000000	0.000000
anti_P2Y12_out	nan	0.000000	0.000000	0.000000	0.597216	0.000000	0.000000	0.000000
ntidepressant_in	nan	0.784804	0.641667	0.000000	0.000000	0.000000	0.000000	0.000000
ntidepressant_out	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
nxiolytiques_out	nan	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000
apical_type	nan	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000
aspirin_in	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
aspirin_out	nan	nan	nan	nan	0.567230	0.000000	0.000000	0.000000
rium_arrhythmia	nan	nan	nan	nan	0.000000	0.538304	0.000000	0.000000
basale	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
eta_blockers_in	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
ta_blockers_out	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
cancer	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
cancer_history	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
rdiogenic_shock	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
nic_kidney_disease	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
coronarography	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
oronary_disease	nan	nan	nan	nan	nan	0.000000	0.530088	0.000000
ssion_anxiety_history	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
diabetes	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
dyslipidemia	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
motional_stress	nan	nan	nan	nan	nan	0.000000	0.000000	0.581209
ealthy_coronary	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
heart_failure	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
hospital_death	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
hypertension	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
tal_illness_history	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
mid_ventricular	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
ological_pathologies	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
anticoagulation_in	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
anticoagulation_out	nan	nan	nan	nan	nan	nan	0.000000	0.000000
other	nan	nan	nan	nan	nan	nan	0.000000	0.000000
hysical_stress	nan	nan	nan	nan	nan	nan	0.000000	0.593823
sex	nan	nan	nan	nan	nan	nan	0.000000	0.000000
smoking	nan	nan	nan	nan	nan	nan	0.000000	0.000000
statin_in	nan	nan	nan	nan	nan	nan	0.000000	0.000000
statin_out	nan	nan	nan	nan	nan	nan	nan	0.000000

Correlation After Imputation



	ARA_II_out	antidepressant_out	anxiolytiques_out	aspirin_in	coronary_disease	oral_anticoagulation_out	statin_out	stress_trigger
ARA_II_in	0.594238	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ARA_II_out	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
COPD_asthma	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
CRP	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Cardio_history	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ECMO	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
IEC_in	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
IEC_out	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Long_QT	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
RV_harm	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ST_positive	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Tneg_waves	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
alcoolism	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
anti_P2Y12_in	nan	0.000000	0.000000	0.547486	0.000000	0.000000	0.000000	0.000000
anti_P2Y12_out	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ntidepressant_in	nan	0.784804	0.641667	0.000000	0.000000	0.000000	0.000000	0.000000
ntidepressant_out	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
nxiolytiques_out	nan	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000
apical_type	nan	nan	nan	0.000000	0.000000	0.000000	0.000000	0.000000
aspirin_in	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
aspirin_out	nan	nan	nan	nan	0.521772	0.000000	0.000000	0.000000
rium_arrhythmia	nan	nan	nan	nan	0.000000	0.538304	0.000000	0.000000
basale	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
eta_blockers_in	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
ta_blockers_out	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
cancer	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
cancer_history	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
rdiogenic_shock	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
nic_kidney_disease	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
coronarography	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
oronary_disease	nan	nan	nan	nan	nan	0.000000	0.533055	0.000000
ssion_anxiety_history	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
diabetes	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
dyslipidemia	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
motional_stress	nan	nan	nan	nan	nan	0.000000	0.000000	0.581209
ealthy_coronary	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
heart_failure	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
hospital_death	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
hypertension	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
tal_illness_history	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
mid_ventricular	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
ological_pathologies	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
anticoagulation_in	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000
anticoagulation_out	nan	nan	nan	nan	nan	nan	0.000000	0.000000
other	nan	nan	nan	nan	nan	nan	0.000000	0.000000
physical_stress	nan	nan	nan	nan	nan	nan	0.000000	0.593823
sex	nan	nan	nan	nan	nan	nan	0.000000	0.000000
smoking	nan	nan	nan	nan	nan	nan	0.000000	0.000000
statin_in	nan	nan	nan	nan	nan	nan	0.000000	0.000000
statin_out	nan	nan	nan	nan	nan	nan	nan	0.000000



	height	BMI	entry_LVEF	out_LVEF	entry_troponin	troponin_peak	inhospital_consequences_encoded
age	0.000000	0.030961	0.059262	0.094148	0.000000	0.082698	0.000000
weight	0.375407	0.891058	0.049236	0.080788	0.109000	0.037999	0.072419
height	nan	0.000000	0.000000	0.000000	0.242861	0.067152	0.187644
BMI	nan	nan	0.120443	0.090539	0.000000	0.008958	0.000000
entry_LVEF	nan	nan	nan	0.351729	0.000000	0.000000	0.000000
out_LVEF	nan	nan	nan	nan	0.000000	0.000000	0.000000
entry_troponin	nan	nan	nan	nan	nan	0.166738	0.089642
troponin_peak	nan	nan	nan	nan	nan	nan	0.000000
inhospital_consequences_encoded	nan	nan	nan	nan	nan	nan	nan

	height	BMI	BMI_range	entry_LVEF	out_LVEF	entry_LVEF_category	entry_troponin	troponin_peak	NT_proBNP	inhospital_consequences_encoded
age	0.000000	0.039105	0.062840	0.059262	0.018047	0.000000	0.000000	0.083115	0.000000	0.000000
weight	0.368175	0.874133	0.769652	0.043880	0.027360	0.055342	0.100495	0.033372	0.000000	0.070210
height	nan	0.000000	0.000000	0.000000	0.005326	0.000000	0.229430	0.066585	0.000000	0.190722
BMI	nan	nan	0.885144	0.104190	0.039195	0.059107	0.000000	0.004905	0.000000	0.000000
BMI_range	nan	nan	nan	0.090361	0.012280	0.036509	0.019177	0.052233	0.000000	0.000000
entry_LVEF	nan	nan	nan	nan	0.240180	0.898215	0.000000	0.000000	0.000000	0.000000
out_LVEF	nan	nan	nan	nan	nan	0.164910	0.000000	0.000000	0.000000	0.000000
entry_LVEF_category	nan	nan	nan	nan	nan	nan	0.000000	0.000000	0.000000	0.000000
entry_troponin	nan	nan	nan	nan	nan	nan	nan	0.169407	0.000000	0.082838
troponin_peak	nan	nan	nan	nan	nan	nan	nan	nan	0.121622	0.000000
NT_proBNP	nan	nan	nan	nan	nan	nan	nan	nan	nan	0.130669
inhospital_consequences_encoded	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan

Observation

We can observe there is no significant difference of correlation before and after imputation

MACHINE LEARNING AND PREDICTION

Heart Failure:

- Target
  - Heart Failure
- Strategy
  - Binary Classification
- Algorithm
  - PyCaret(AutoML)
  - LightGBM with Optuna
    - LightGBM is a gradient boosting framework that uses tree based learning algorithms. It is designed to be distributed and efficient with speed, memory usage, accuracy, parallel computing, large-scale data[1].
    - Optuna is an automatic hyperparameter optimization software framework, particularly designed for machine learning[2].
- Evaluation Metrics:
  - Accuracy
  - Precision
    - The fraction of relevant instances among the retrieved instances
  - **Recall**
    - The fraction of relevant instances that were retrieved
  - AUC
  - SHAP: Explainability of the model
    - SHAP(SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions.

1.https://lightgbm.readthedocs.io/en/latest/
2. https://github.com/optuna/optuna
3. https://github.com/slundberg/shap

MACHINE LEARNING AND PREDICTION

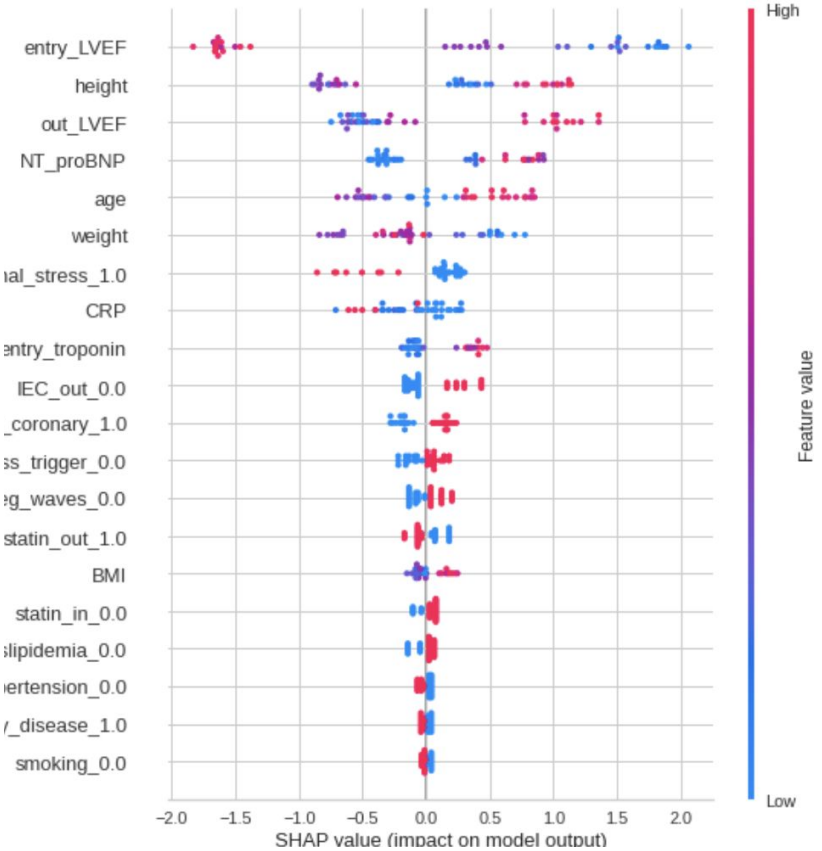
Heart Failure:

Algorithm/Metrics	Accuracy	Precision	Recall	AUC
PyCaret: Ada Boost Classifier	81.11%	66.67%	**68.00%**	85.54%
PyCaret: Extra Trees Classifier	78.89%	85%	32%	81.69%
PyCaret Ridge Classifier	76.67%	62.78%	60.00%	—
LightGBM with Optuna	72.27%	—	—	—

Observation

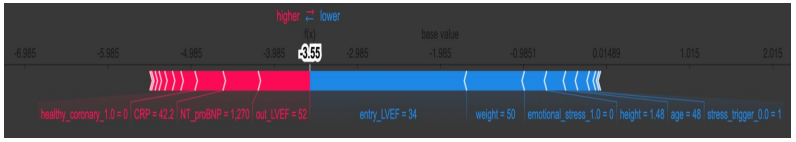
PyCaret's Ada Boost shows the highest Recall rate, which is the main evaluation in this context. High recall shows less False Negative, that is to say, the algorithm wrongly diagnose a patient not as Heart Failure.

SHAP Value:



This plot shows the largest contribution to the model is entry_LVEF. Blue dots shows the low value of that variable while red shows the high value, which means entry_LVEF has a negative correlation with the output of heart failure.

Interestingly, height has the second largest contribution to whether patients have Heart Failure or not. This plot explains taller people tends to have Heart Failure comparatively more often than shorter people.



This plot shows each variable contribution for an individual patient. We can observe the same tendency as above plot.

MACHINE LEARNING AND PREDICTION

Inhospital Complications:

- Target
  - Inhospital Compiications(Heart Failure, RV harm, Atrium Arrhythmia, Cardiogenic Shock, Ventricle Arrhythmia, Thrombus LV, Hospital Death, ECMO)

- Strategy
  - Multi-Label Classification

- Algorithm
  - Decision Tree Classifier
  - Multilabel K Nearest Neighbours
  - Multilabel K Nearest Neighbours with Grid Search

- Evaluation Metrics:
  - Exact Match Ratio(EMR)
    - The Exact Match Ration extends the concept the accuracy from the single-label classification problem to a multi-label classification problem
    - One of the drawbacks of using EMR is it does not account for partially correct labels
  - Hamming Loss
    - Hamming Loss computes the proportion of incorrectly predicted labels to the total number of labels
    - For a multi-label classification, we compute the number of False Positive and False Negative per instance and average them over the total number of training instances

MACHINE LEARNING AND PREDICTION

Inhospital Complications:

Algorithm/Metrics	Exact Match Ratio(EMR)	Hamming Loss
Decision Tree Classifier	46.15%	10.26%
Multilabel K Nearest Neighbours	48.72%	8.97%
Multilabel K Nearest Neighbours with Grid Search(k:22, s:0.5)	**61.54%**	**7.37%**

Observation

Decision Tree and KNN(K Nearest Neighbours) shows nearly same result at the beginning. With grid search, KNN improves its performance by more than 10% for EMR and less than 1% for Hamming Loss

```
1 from sklearn.model_selection import GridSearchCV
2
3 parameters = {'k': range(1,30), 's': [0.5, 0.7, 1.0]}
4 score = 'accuracy'
5
6 clf = GridSearchCV(MLkNN(), parameters, scoring=score)
7 clf.fit(X_train.values, y_train.values)
8
9 print (clf.best_params_, clf.best_score_)
```

```
{'k': 22, 's': 0.5} 0.6222222222222222
```

Implementation of Multi KNN with Grid Search

CONLUSION AND PERSPECTIVES

- Difficulty of data collection: precision.
- Clinical features mostly – how about genetics, epigenetics, proteomics… ?
- Difficulty in data processing: dropping data when we already have just a few?
- Make the algorithm more precise by distinguishing the physical/emotional stresses.
- Contradiction of adopting the latest technologies.

BIBLIOGRAPHY AND SOURCES

ARTICLES :

Templin, C., Ghadri, J. R., Diekmann, J., Napp, L. C., Bataiosu, D. R., Jaguszewski, M., Cammann, V. L., Sarcon, A., Geyer, V., Neumann, C. A., Seifert, B., Hellermann, J., Schwyzer, M., Eisenhardt, K., Jenewein, J., Franke, J., Katus, H. A., Burgdorf, C., Schunkert, H., … Lüscher, T. F. (2015). Clinical Features and Outcomes of Takotsubo (Stress) Cardiomyopathy. New England Journal of Medicine, 373(10), 929–938.

Bloom, M. W., Greenberg, B., Jaarsma, T., Januzzi, J. L., Lam, C. S. P., Maggioni, A. P., Trochu, J.-N., & Butler, J. (2017). Heart failure with reduced ejection fraction. Nature Reviews Disease Primers, 3(1).

Sharkey, S. W., Windenburg, D. C., Lesser, J. R., Maron, M. S., Hauser, R. G., Lesser, J. N., Haas, T. S., Hodges, J. S., & Maron, B. J. (2010). Natural History and Expansive Clinical Profile of Stress (Tako-Tsubo) Cardiomyopathy. Journal of the American College of Cardiology, 55(4), 333–341.

Ono, R., & Falcão, L. M. (2016). Takotsubo cardiomyopathy systematic review: Pathophysiologic process, clinical presentation and diagnostic approach to Takotsubo cardiomyopathy. International Journal of Cardiology, 209, 196–205.

Ghadri, J. R., Kato, K., Cammann, V. L., Gili, S., Jurisic, S., Di Vece, D., Candreva, A., Ding, K. J., Micek, J., Szawan, K. A., Bacchi, B., Bianchi, R., Levinson, R. A., Wischnewsky, M., Seifert, B., Schlossbauer, S. A., Citro, R., Bossone, E., Münzel, T., … Templin, C. (2018). Long-Term Prognosis of Patients With Takotsubo Syndrome. Journal of the American College of Cardiology, 72(8), 874–882.

Veillet-Chowdhury, M., Hassan, S. F., & Stergiopoulos, K. (2014). Takotsubo cardiomyopathy: A review. Acute Cardiac Care, 16(1), 15–22.

Limongelli, G., D'Alessandro, R., Masarone, D., Maddaloni, V., Vriz, O., Minisini, R., Citro, R., Calabrò, P., Russo, M. G., Calabrò, R., Pacileo, G., Bossone, E., & Elliott, P. M. (2013). Takotsubo Cardiomyopathy. Heart Failure Clinics, 9(2), 207–216.

Nguyen, T. H., Neil, C. J., Sverdlov, A. L., Mahadavan, G., Chirkov, Y. Y., Kucia, A. M., Stansborough, J., Beltrame, J. F., Selvanayagam, J. B., Zeitz, C. J., Struthers, A. D., Frenneaux, M. P., & Horowitz, J. D. (2011). N-Terminal Pro-Brain Natriuretic Protein Levels in Takotsubo Cardiomyopathy. The American Journal of Cardiology, 108(9), 1316–1321.

Scantlebury, D. C., & Prasad, A. (2014). Diagnosis of Takotsubo Cardiomyopathy. Circulation Journal, 78(9), 2129–2139.

WEBSITES

https://www.health.harvard.edu/heart-health/takotsubo-cardiomyopathy-broken-heart-syndrome
http://www.remede.org/librairie-medicale/pdf/e9782224030766.pdf
https://www.svhhearthealth.com.au/conditions/takotsubo-cardiomyopathy
https://www.heartfoundation.org.nz/your-heart/heart-tests/coronary-angiography
https://www.vidal.fr/actualites/22737-syndrome-du-c-ur-brise-de-takotsubo-publication-d-un-consensus-d-experts-mais-des-questions-demeurent.html
https://www.takotsubo-registry.com/takotsubo-score.html
http://www.mednuc.net/wp-content/uploads/2017/05/Takotsubo-et-ph%C3%A9ochromocytome.pdf
https://www.uofmhealth.org/health-library/tx4091abc