

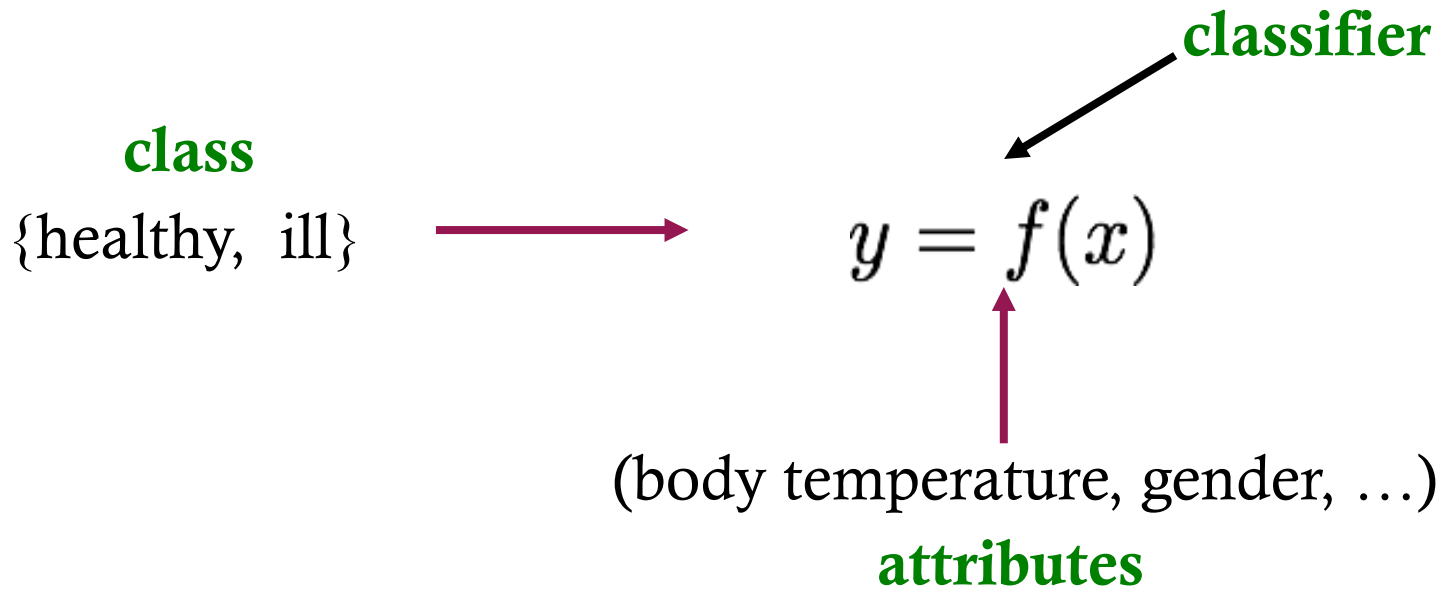
# Explainable and Ethical Machine Learning with applications to healthcare



**Stockholm  
University**

**Panagiotis Papapetrou, PhD  
Professor, Stockholm University**

# Explainability in Machine Learning



Properties of a good classifier include:

- predictive performance
- **explainability** (e.g., why is a person healthy?)

# The EXTREMUM team

SU



**Panagiotis Papapetrou**



**Lars Asker**



**Stanley Greenstein**

**5 PhDs**

**1 Postdoc**

**1 Developer**

**KTH**



**Cristian Rojas**

**RISE**



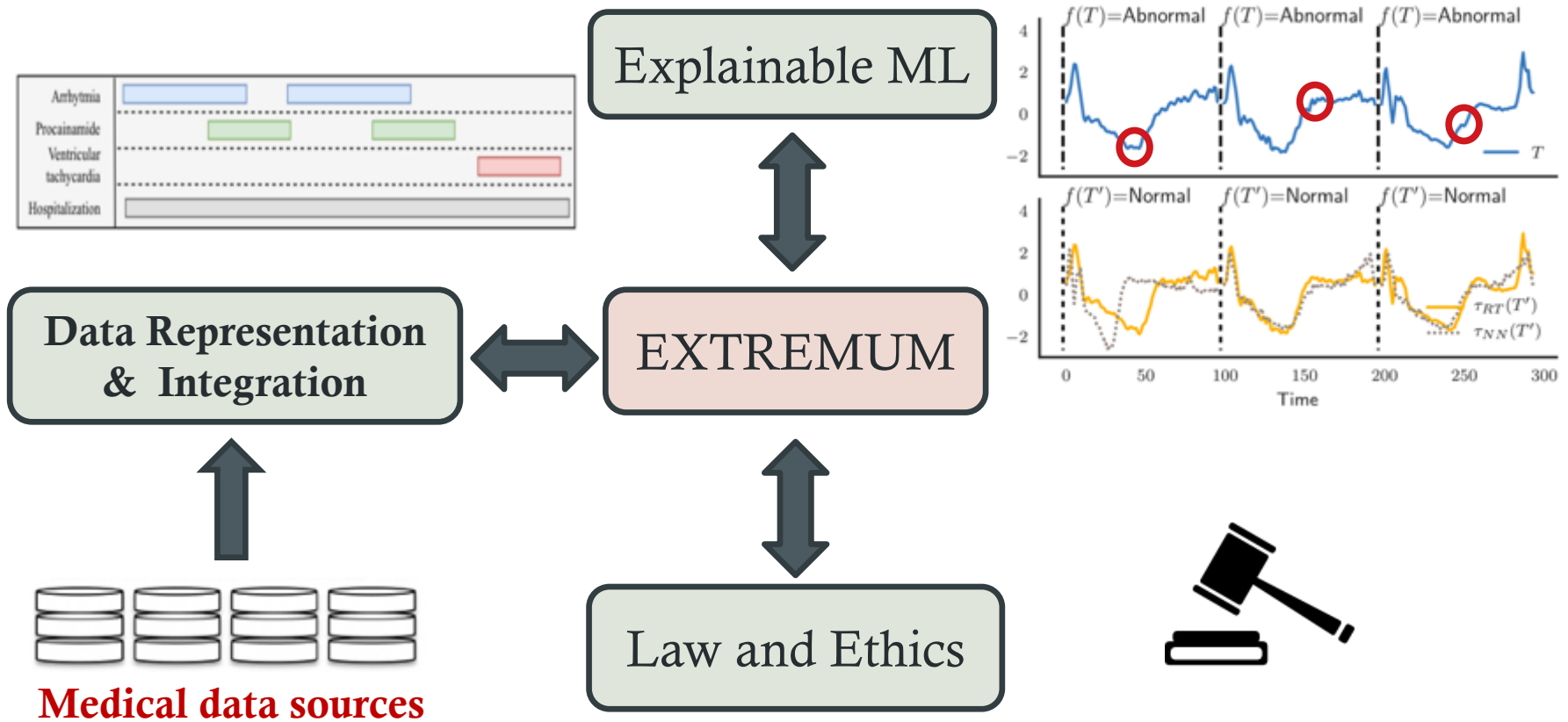
**Rami Mochaourab**

**Externals**

**Karolinska Institute**

Funded by **Digital Futures**  
**2020-2023**

# The EXTREMUM framework

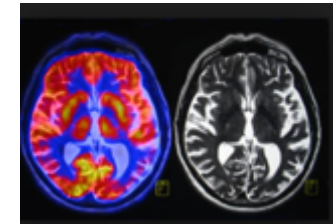


# Electronic Health Records (EHRs)

- **Diagnoses (ICD)**
- **Medications (ATC)**
- **Procedures (CPT)**
- **Blood tests**
- **More complex data sources**
  - clinical notes
  - medical images
  - MRIs
  - ultra-sounds
  - ECGs
  - ...

I25.110

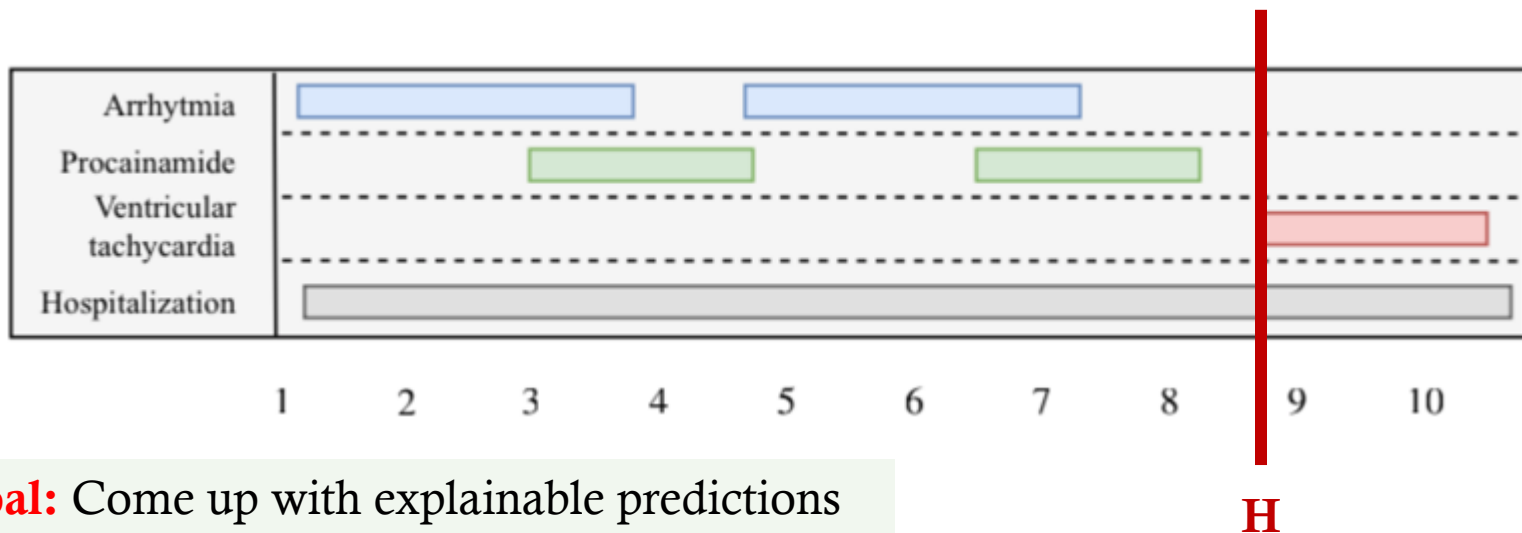
A01AD05





# Defining temporal abstractions

- EHR: patient record
- Many complex variables (static, temporal, text, images)
- An event of interest **H**: e.g., an Adverse Drug Event (ADE), Heart Failure



# Pillar II: Explainable Machine Learning

- Trade-offs between explainability + accuracy
- Ability to **understand the predictions** + **act** to prevent undesirable outcomes without compromising predictive performance



black box classifier



The patient will die from HF in **2 days!**

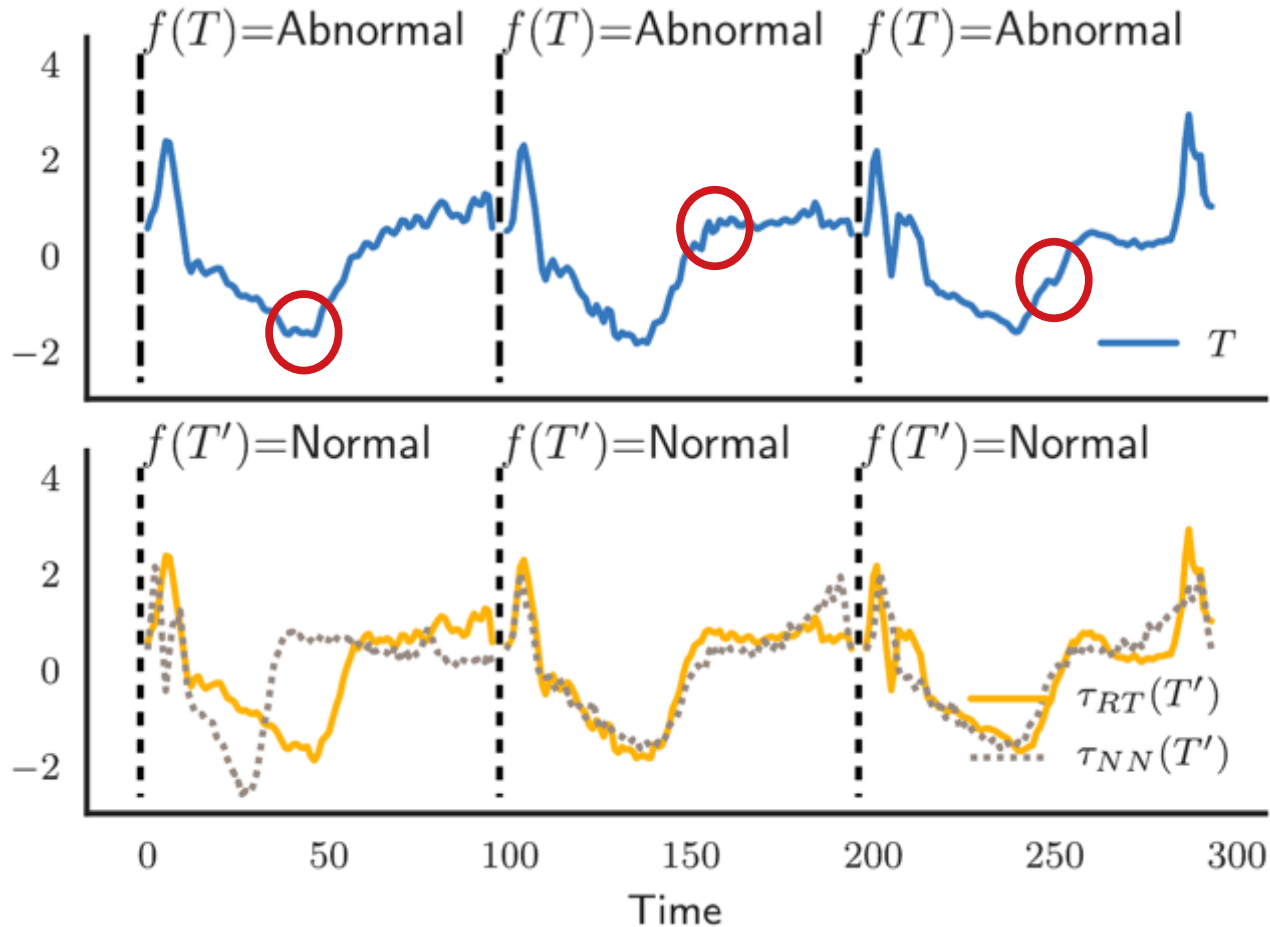


**Explainable** and  
**Transparent ML**

Now what?  
Please tell me  
**why?**

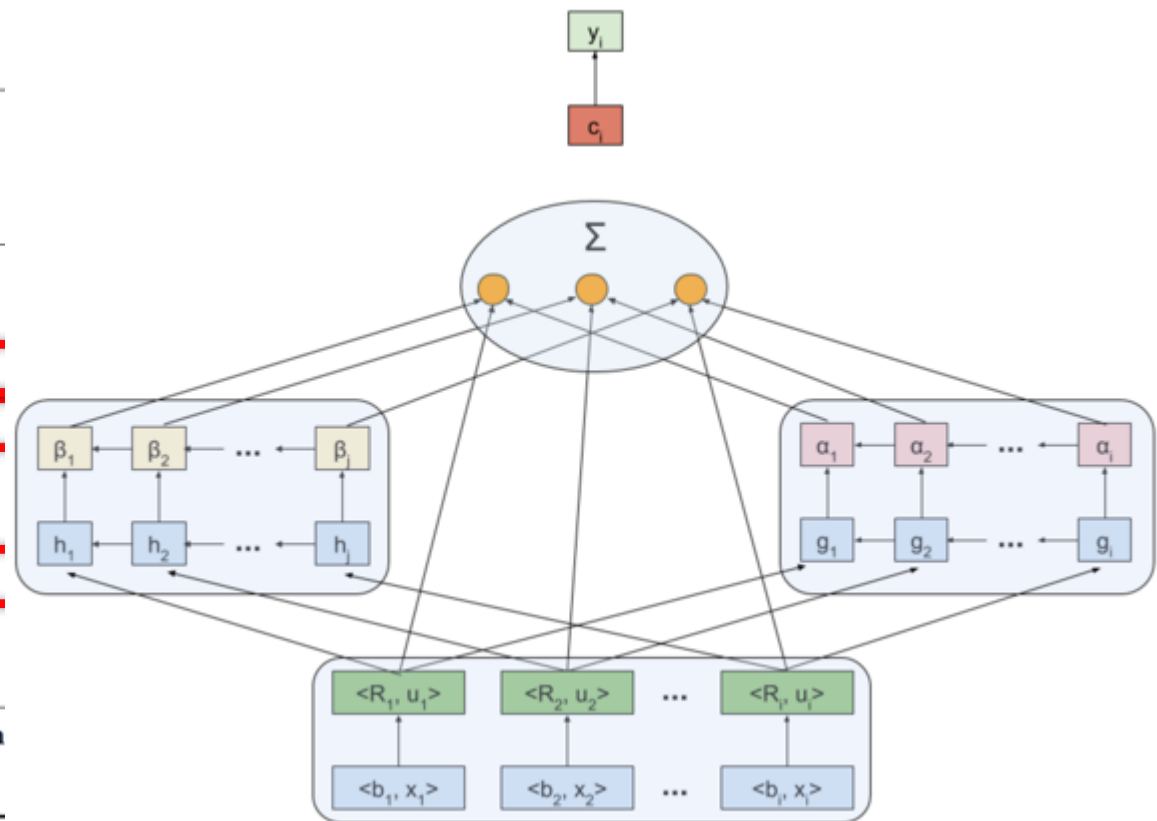


# Time series explainability

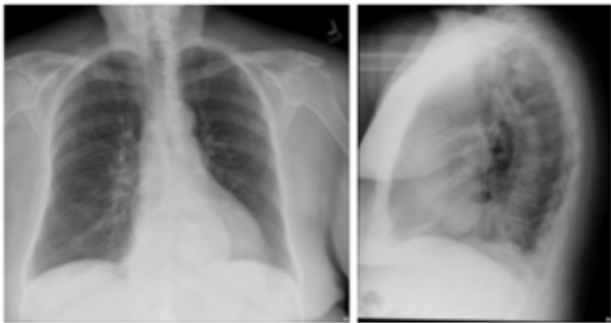


# Attention-based ADE prediction

Code	Description	Score
<b>Visit 1</b>		
L50	Urticaria	
<b>Visit 2</b>		
R42	Dizziness and giddiness	
A02 <sup>†</sup>	Drugs for acid related disorders	
<b>Visit 3</b>		
L50	Urticaria	
R06 <sup>†</sup>	Antihistamines for systemic use	
H02 <sup>†</sup>	Corticosteroids for systemic use	
<b>Visit 4</b>		
L50	Urticaria	
R06 <sup>†</sup>	Antihistamines for systemic use	
C01 <sup>†</sup>	Cardiac therapy	
H02 <sup>†</sup>	Corticosteroids for systemic use	
<b>Prediction</b>		
T784	Adverse effects: allergy	

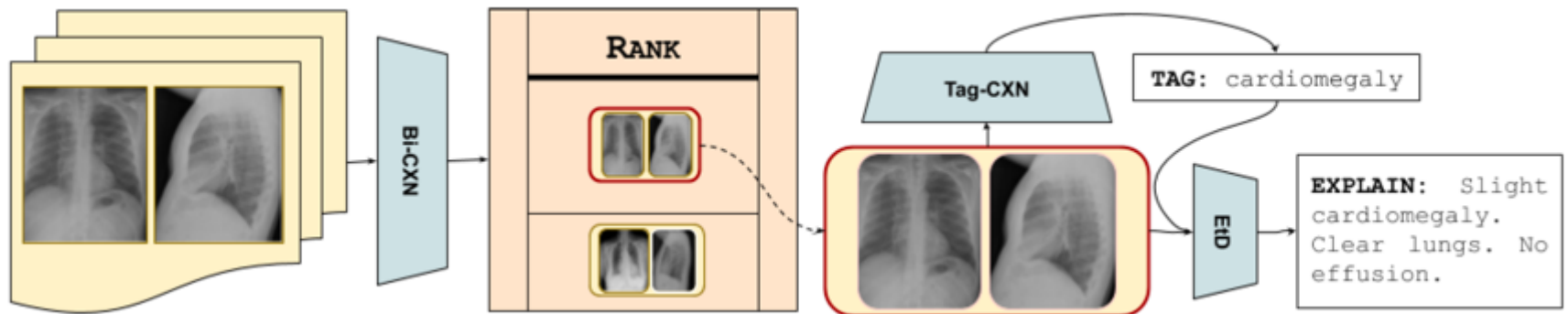
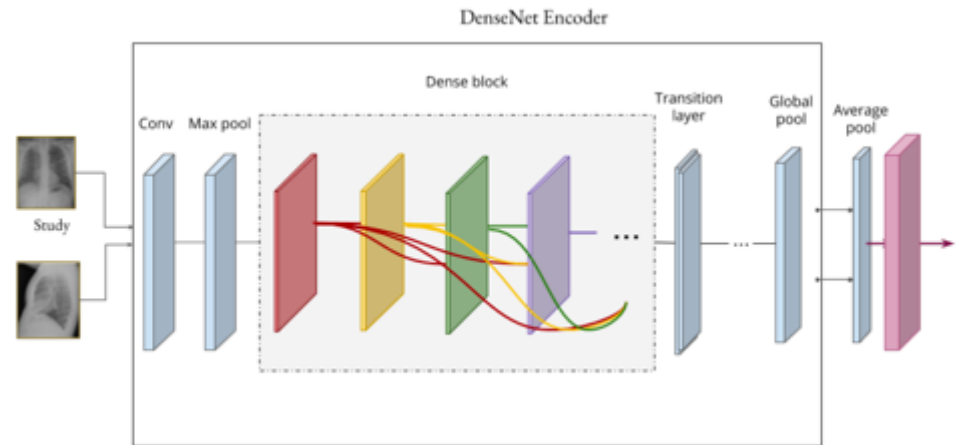


# X-Ray Ranking and Explanatory Diagnostic Tagging



**FINDINGS:** The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

**TAGS:** Atelectases, Cervical Arthritis, Atelectasis, Spondylarthritis. Thoracic Spondylosis.



# Pillar III: Legal and Ethical Compliance

- Legal requirements for explanations (GDPR)
- **Bias detection** and **mitigation** in the training data:
  - discriminatory variables
  - complex relation patterns
- Formulation of a **legal framework**
  - ability to check the **legal compliance** of ML algorithms
  - ability to **identify** and **remove** bias



# Demonstrator beta 1.0

- Tool created using **Django** and **Python**
- Facilitate its scalability with future data science applications developed in this programming language

## Thanks to

Luis Quintero (SU)

Sugandh Sinha (RISE)

Demo scenario	Data types	Classification models	Explainability techniques
<b>Cardiovascular disease identif.</b>	Tabular data: binary, categorical	Random forests	Actionable Feature Tweaking a.k.a Counterfactual Explanations
<b>Time series tweaking via shapelet transformations</b>	Time series: univariate and multivariate	Random shapelet forests SVM K-NN	Explanation guided by prototypes LIME on DFT features Global tweaking: k-NN, SVM Explain [T1]
<b>Medical X-ray ranking and captioning</b>	Images: x-ray medical images	Ranking: BI-CXN Tagging: TAG-CXN Captioning: LSTM-ETD	X-Ray Ranking and Explanatory Diagnostic Tagging

# Demonstrator beta 1.0

**INTERPRETABLE | MACHINE LEARNING**

MENU

>Explainable and Ethical Machine Learning for Knowledge Discovery from Medical Data Sources

**CORONARY HEART DISEASE**

**GO TO DEMO**

Clinical Decision Support System to estimate risk of Coronary Heart Disease in Ten-Years  
Implements Feature Tweaking algorithm on Random Forest Classifiers

**X-RAY ANALYSIS**

**GO TO DEMO**

Analysis of Cancer on Chest X-Ray Images  
Implements Feature Tweaking algorithm on Random Forest Classifiers

Myocardial Infarction Classification

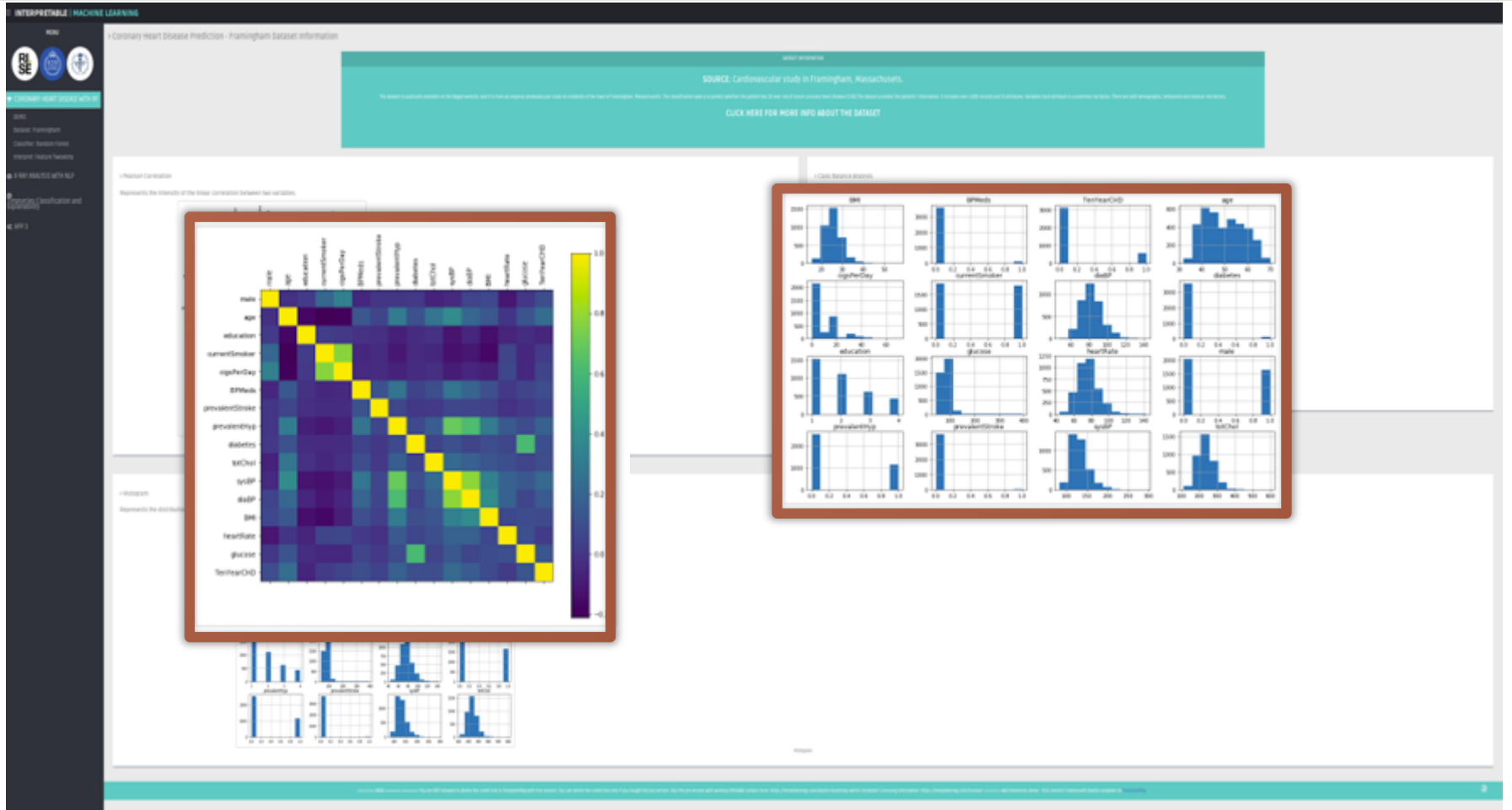
**CORONARY HEART DISEASE**

**GO TO DEMO**

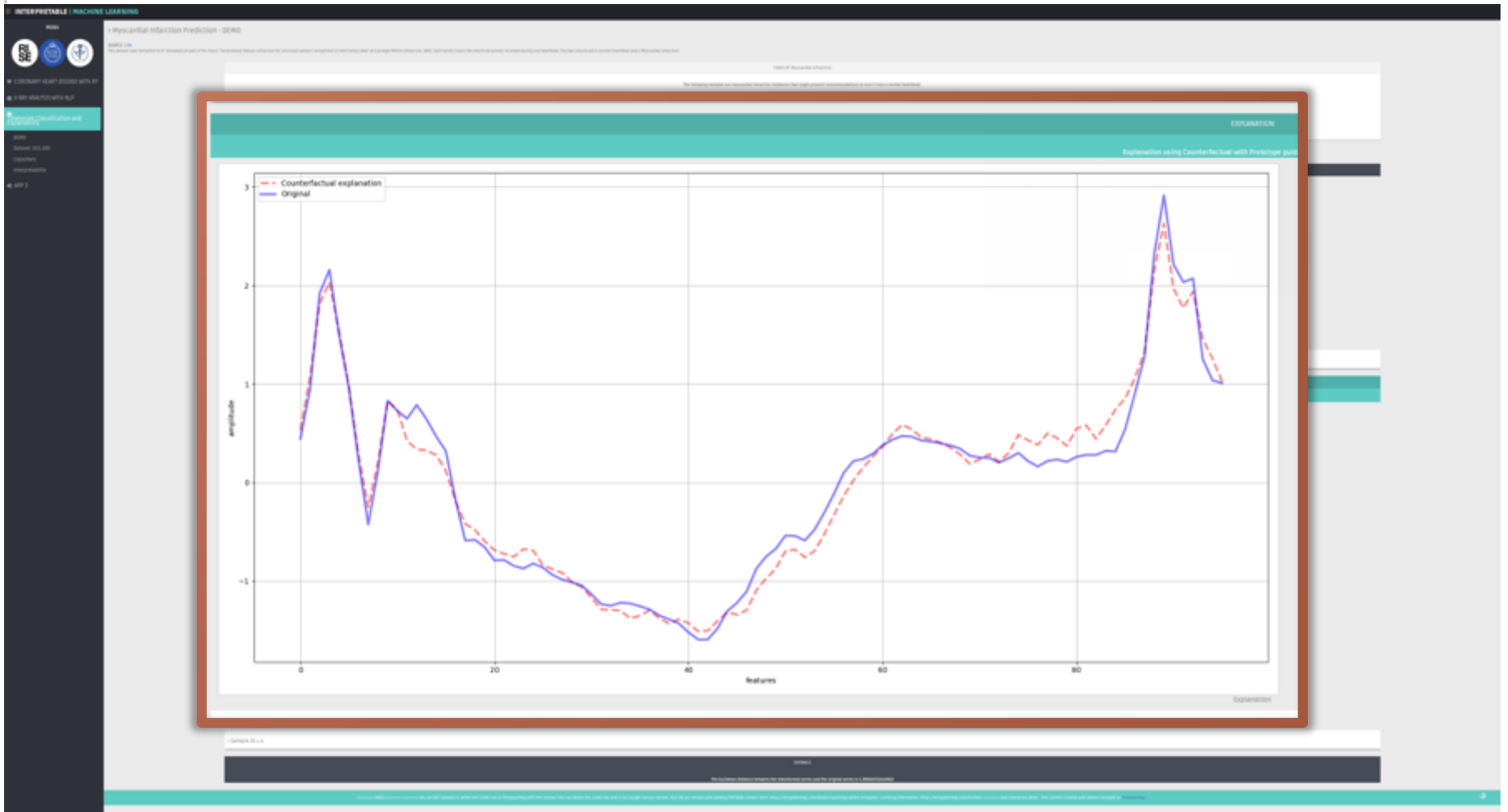
Clinical Decision Support System to estimate risk of Coronary Heart Disease in Ten-Years  
Implements Feature Tweaking algorithm on Random Forest Classifiers

in Progress

# Data exploration



# Time series counterfactuals





# Thank you!

Questions?

## The data science group at SU

<https://dsv.su.se/en/research/research-areas/datascience/>

## The EXTREMUM project

<https://dsv.su.se/en/research/research-areas/datascience/extremum-explainable-and-ethical-machine-learning-for-knowledge-discovery-from-medical-data-sources-1.442728>