

# VIPR: An Interactive Tool for Meaningful Visualization of High-Dimensional Data



Donghan Wang [donghanw@cs.cmu.edu](mailto:donghanw@cs.cmu.edu)  
 Artur Dubrawski [awd@cs.cmu.edu](mailto:awd@cs.cmu.edu)



Madalina Fiterau  
[mfiterau@cs.stanford.edu](mailto:mfiterau@cs.stanford.edu)

## MOTIVATION

Our method targets applications where a **human operator** is involved in the decision. The process must be:

- Transparent
- Comprehensible

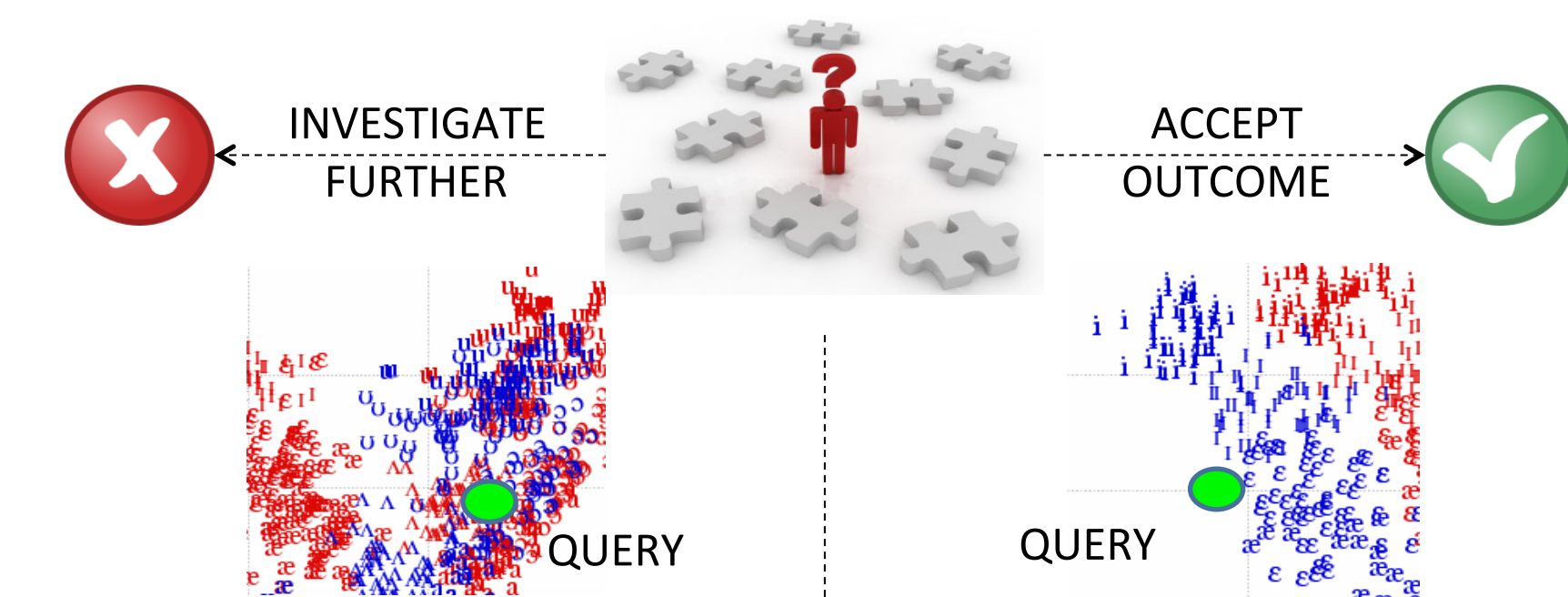
Thus, the problem of **finding subspaces** where data is classified with **high accuracy** but which also give operators **confidence** in the predictions.

**Informative Projection Ensemble (IPE)** methodology has proven effective in finding interpretable renderings of high-dimensional data that reveal hidden low-dimensional structures if they exist.

User is in control of the choice:

- Investigate Further – expensive
- Accept Outcome - assume responsibility

### Automated Decision Support Systems

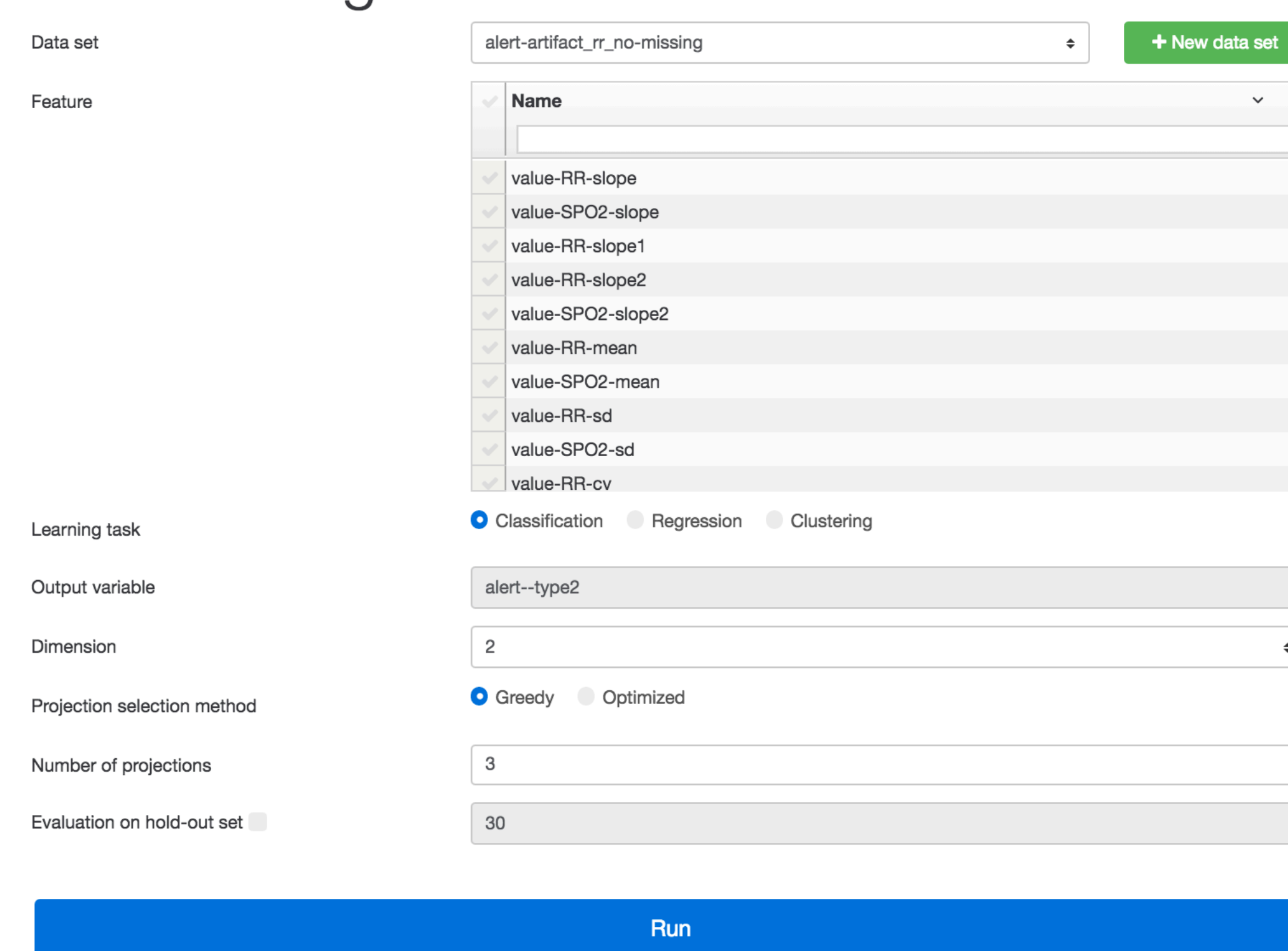


## THE VIPR INTERFACE

Visual toolkit for Informative Projection Recovery (VIPR)

1. Analysis tool for the following tasks: regression, classification, clustering
2. User-specified parameters
  - Features to be used
  - Number of submodels
  - Dimensionality of subspaces
  - Hypothesis class
  - Hold-out set evaluation
3. Manipulations of trained models
  - Add/remove features/samples
  - Compare models
  - Observe prediction on test samples
  - Provide feedback on labels

### VIPR Settings

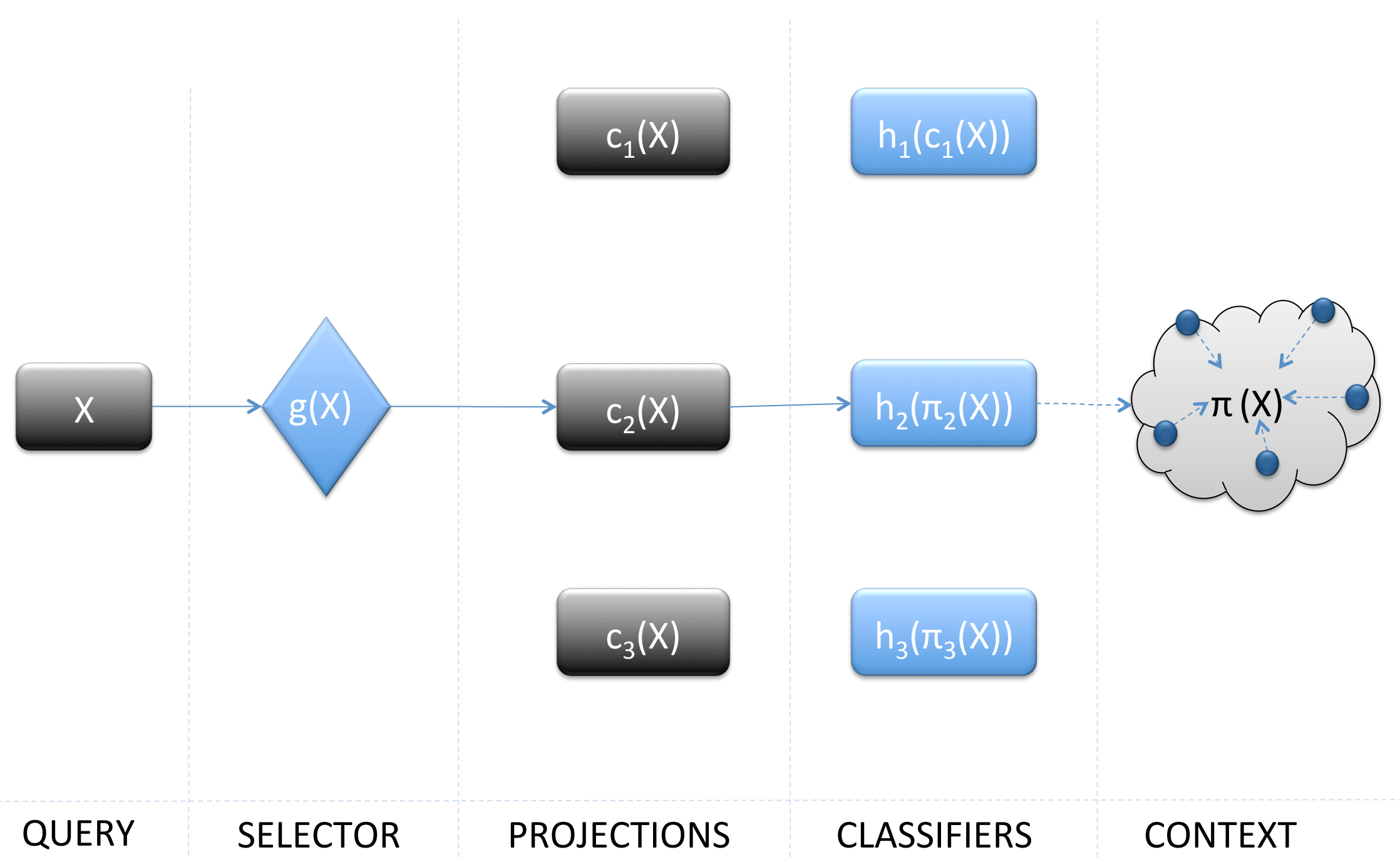


## EXPERIMENTAL RESULTS

Dataset	# Features	# Samples	# Classes	Method	Classifier	Selection	Optimization
Alert BP	147	96	2	IPE-Min	k-NN	Min loss	Two-stage
Alert RR	147	362	2	IPE-H	k-NN	Hyperrectangle	Greedy
Alert SPO2	147	259	2	IPE-Linear	SVM	Multiclass SVM	Two-stage
Chars74k	85	3410	62	IPE-H SVM	SVM	Hyperrectangle	Greedy
G50C	50	550	2				
Letter	16	16000	26				
MNIST	784	60000	10				
USPS	256	11000	10				

Dataset	k-NN	Feature Selection	IPE-Min	IPE-HC	SVM	Feature Selection	IPE-Linear	IPE-HC
		+ k-NN (Best)	k-NN	k-NN		+ SVM (Best)	SVM	SVM
Alert BP	0.7216	76.1	88.6	82.3	75	73.96	77.08	76.04
Alert RR	0.226	0.1253	0.1479	0.1645	0.063	0.1294	0.1391	15.06
Alert SPO2	0.0397	0.0557	0.0117	0.0157	0.0408	0.0648	0.0442	0.0531
Chars74k	31.61	25.34	35.78	35.54	27.07	30.76	35.92	33.72
G50C	87.27	92	94.18	93.45	95.09	94.18	95.64	94.36
Letter	95.25	92.61	95.33	95.1	97.07	91.66	97.1	94.86
MNIST	97.21	92.49	97.57	97.48	9.15	90.53	93.96	9.35
USPS	95.82	93.5	96.69	97.13	93.91	91.38	9.58	9.55
	0.0037	0.001	4.71E-04	7.17E-04	0.0056	0.0062	0.01	0.0014
	0.0036	0.0064	0.0052	0.0061	0.086	0.0964	0.057	0.62

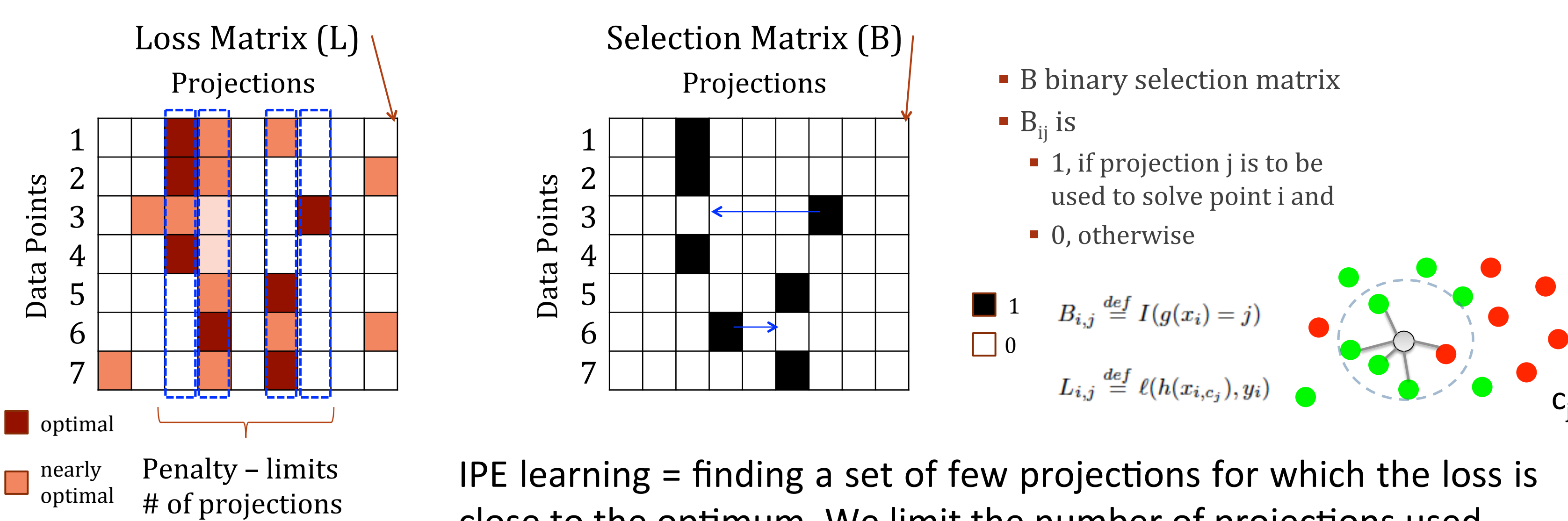
## INFORMATIVE PROJECTION ENSEMBLES



### User-System Interaction:

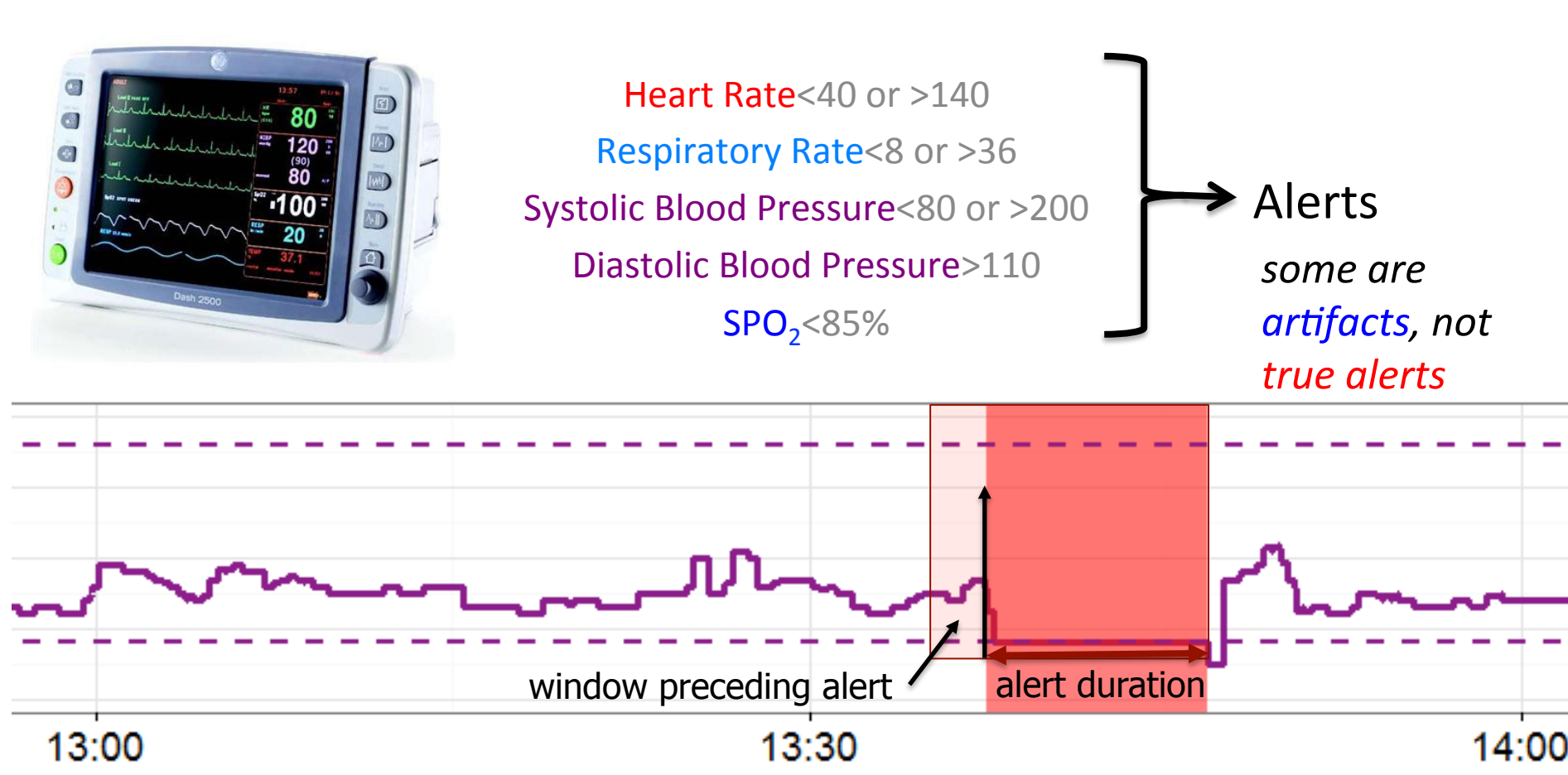
1. the user provides the system a **query** point;
2. system finds a **query-specific projection**
3. system displays result and **illustration** of how the label was obtained

## LEARNING THE ENSEMBLE

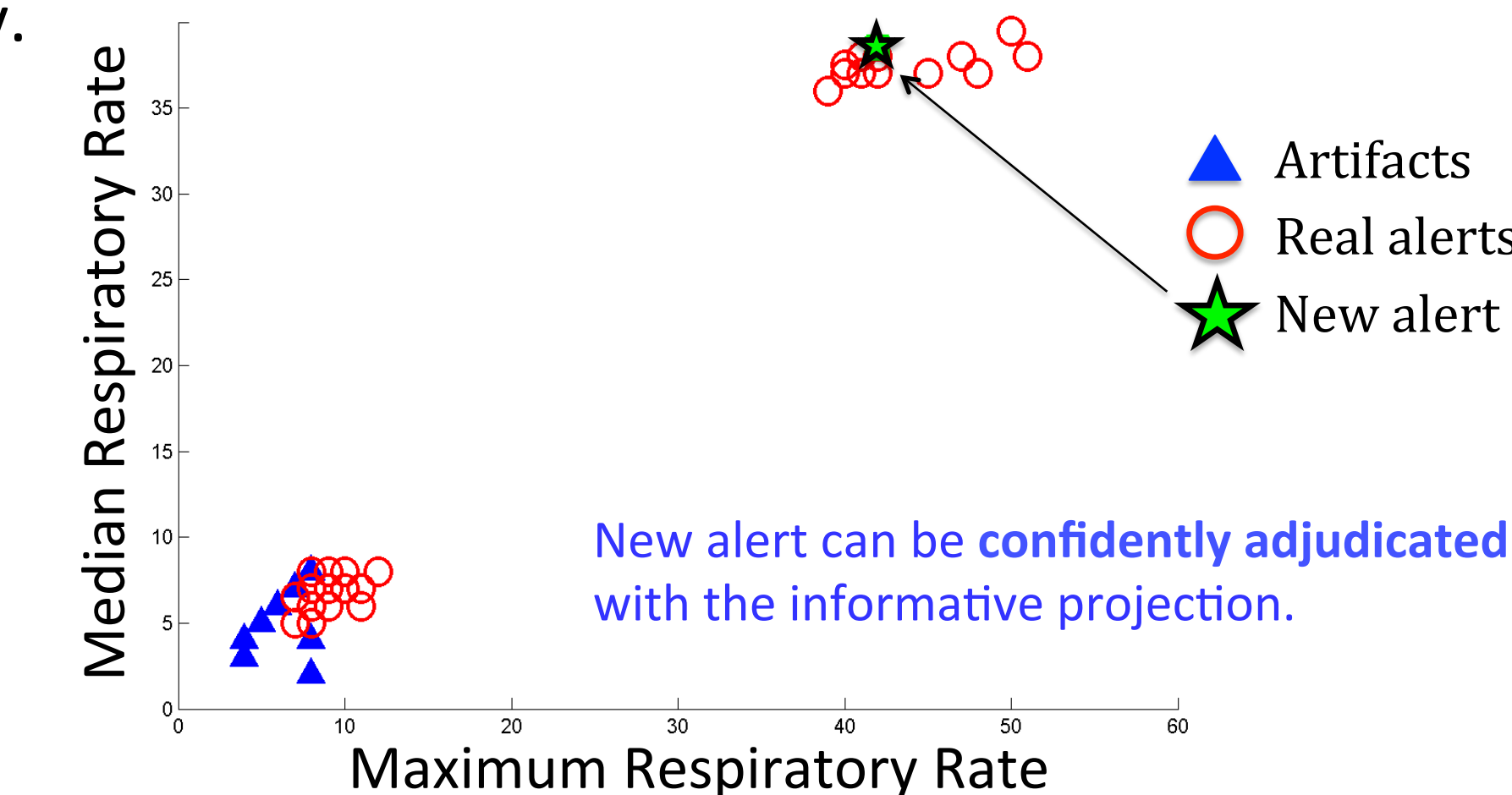
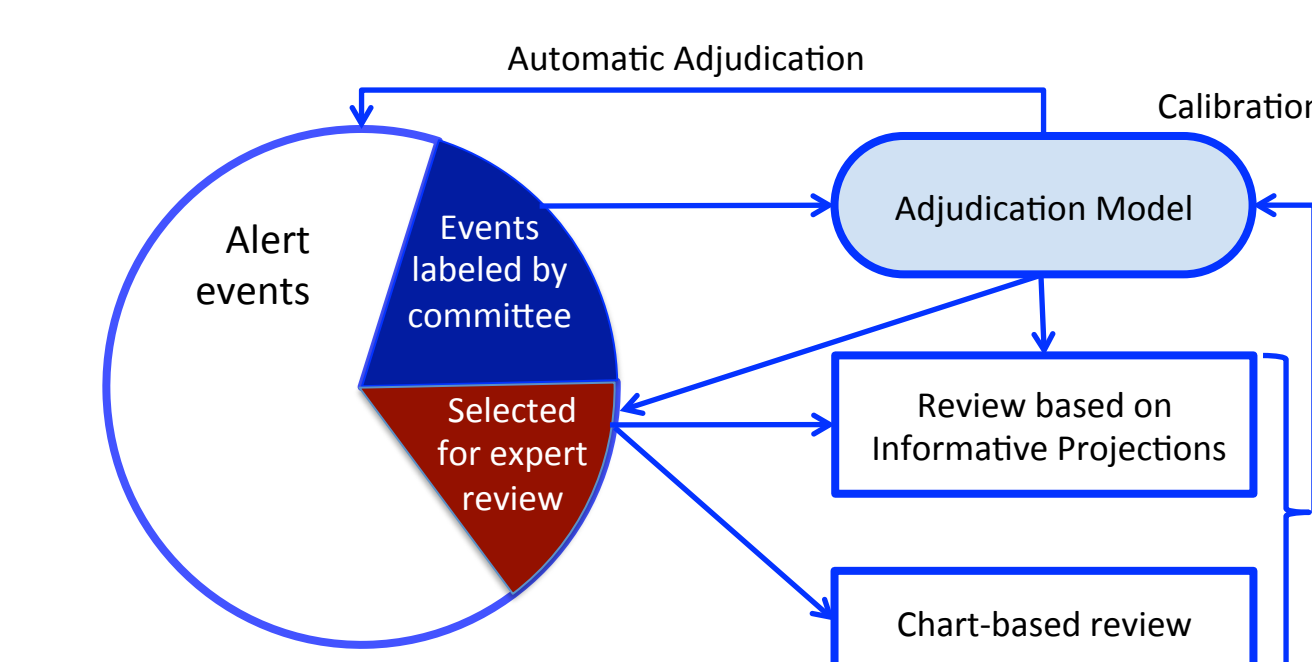


## EXAMPLES

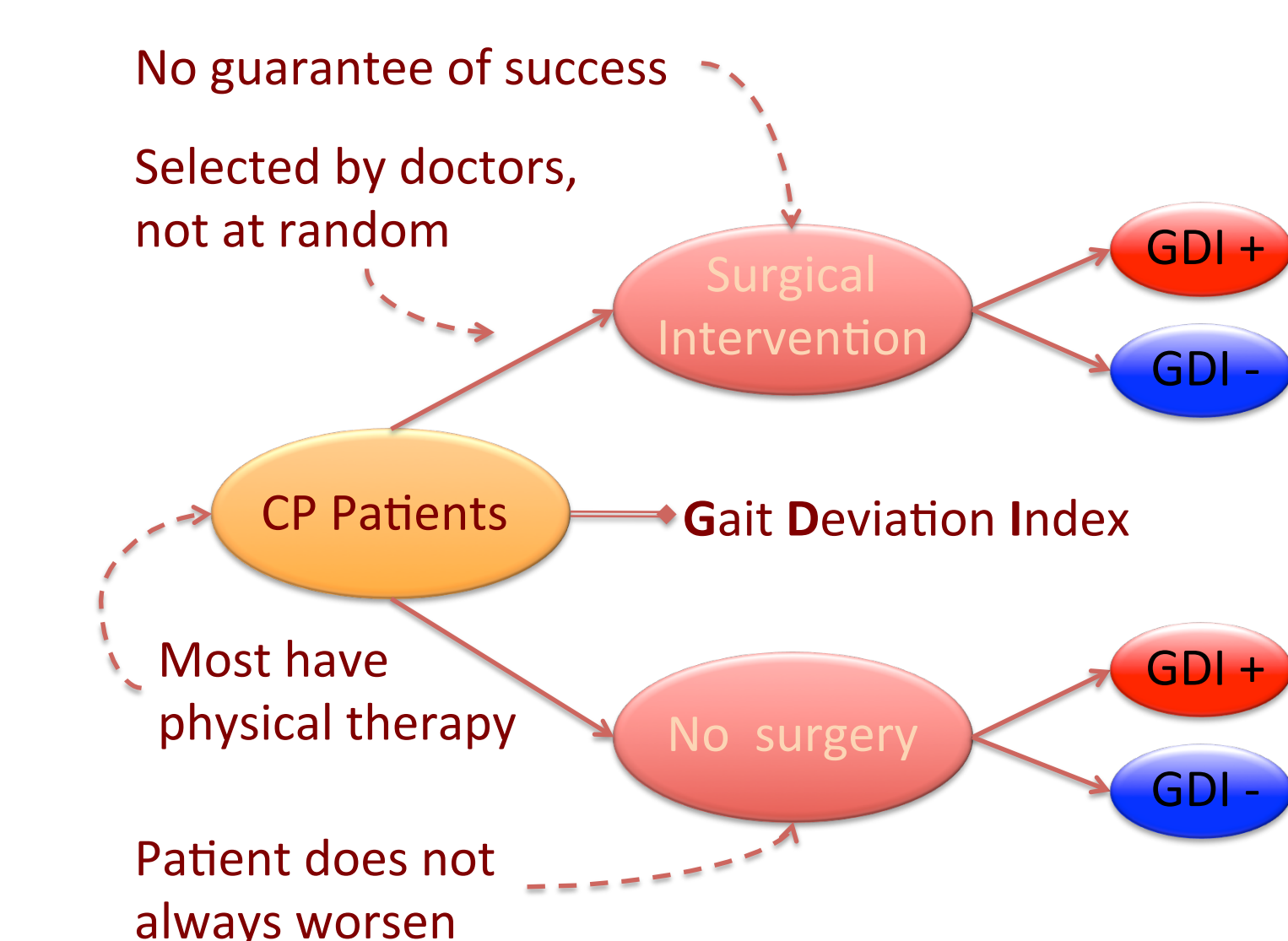
### CLINICAL ALERT ADJUDICATION



Objective: Reduce expert annotation effort through **semi-automatic** adjudication of VS alerts as **real or artifacts**, while maintaining high accuracy.



## CEREBRAL PALSY PROGRESSION



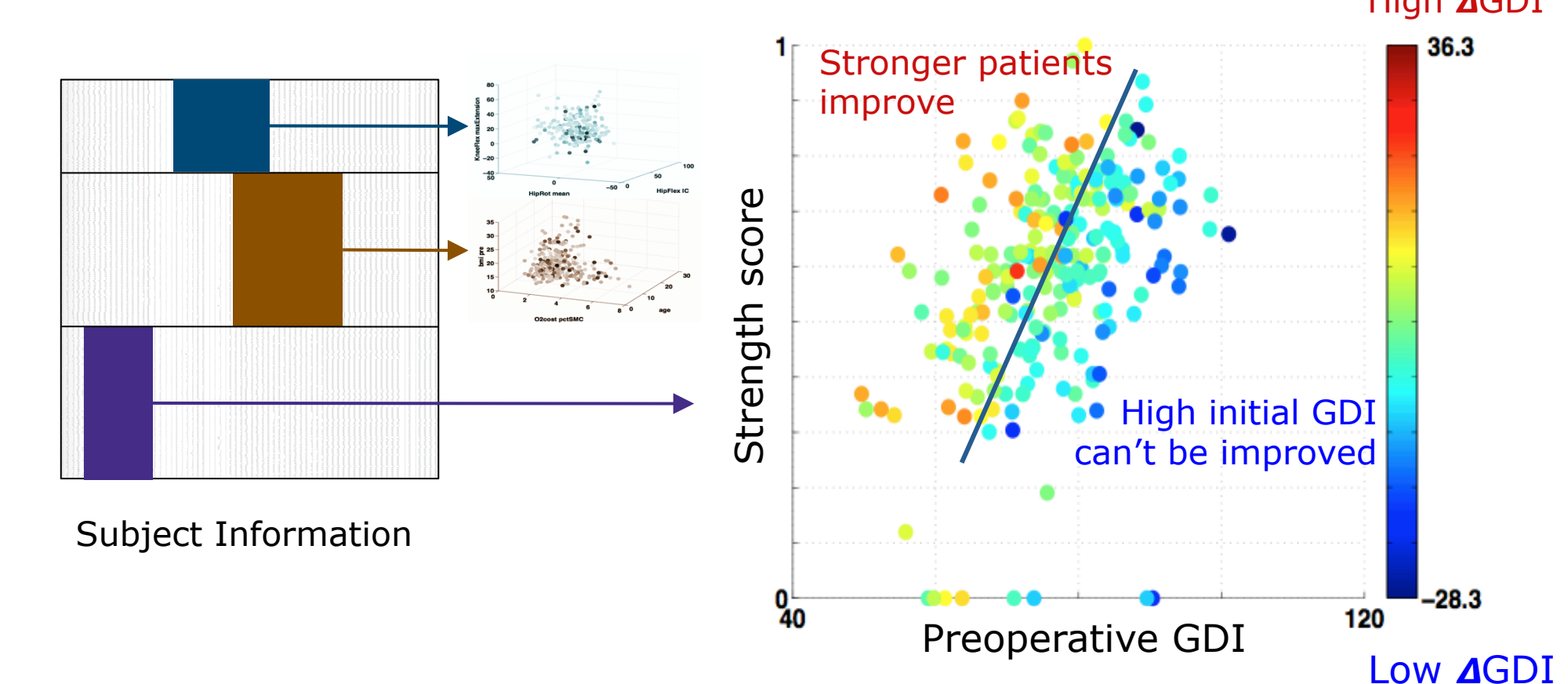
We aim to predict the improvement in **Gait Deviation Index (ΔGDI)** following

- Surgery, for the treatment group
- Alternative treatment for the control group (for instance, physical therapy)

Features used in prediction: age, BMI, joint angles, motor control, strength, walking efficiency (oxygen cost), initial GDI

We show a subset of preliminary results for Propensity Score group 3 (**PS3**)

### Predict ΔGDI (Gait Deviation Index) following surgery



The severity of the disease, in terms of **Gait Deviation Index**, is the **strongest predictor** of outcome

### Predict ΔGDI for the controls

