



Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

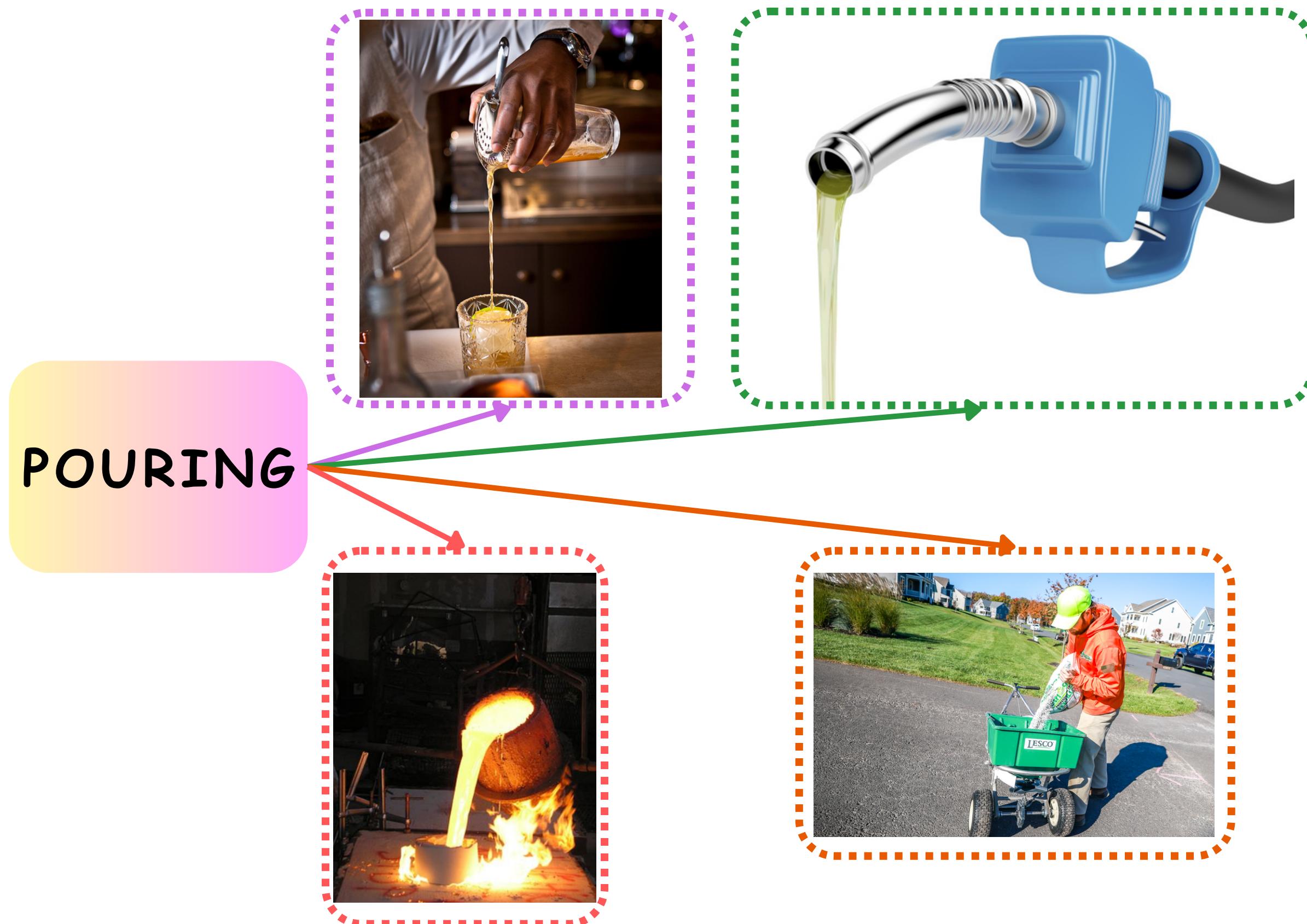


Multimodal Deep Anomaly Detection for Robot Pouring Task

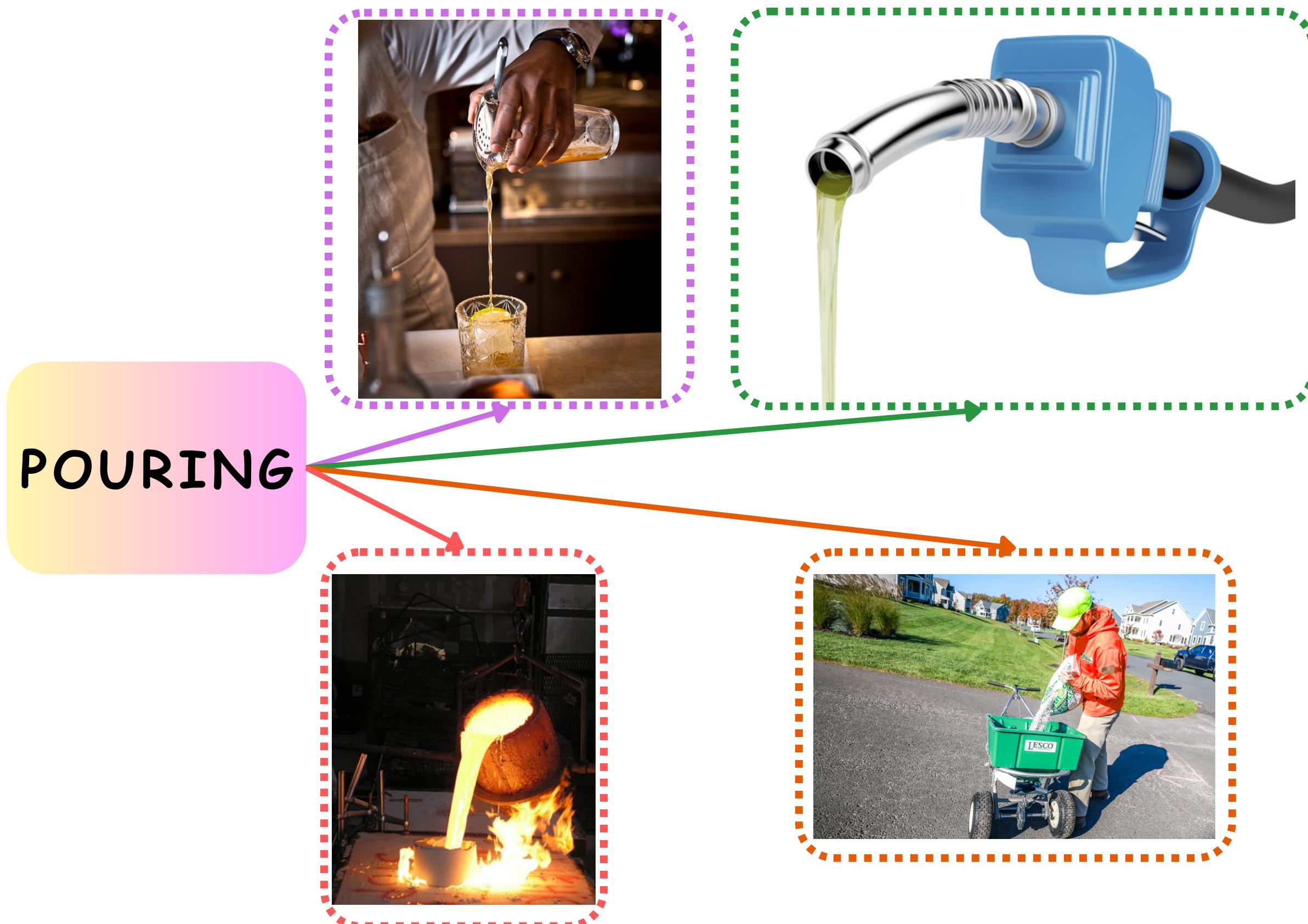
Adithya Narasimhaiah Sathish

Prof. Dr. Paul G. Plöger
Prof. Dr. Nico Hogeshwender
M.Sc. Santosh Thoduka

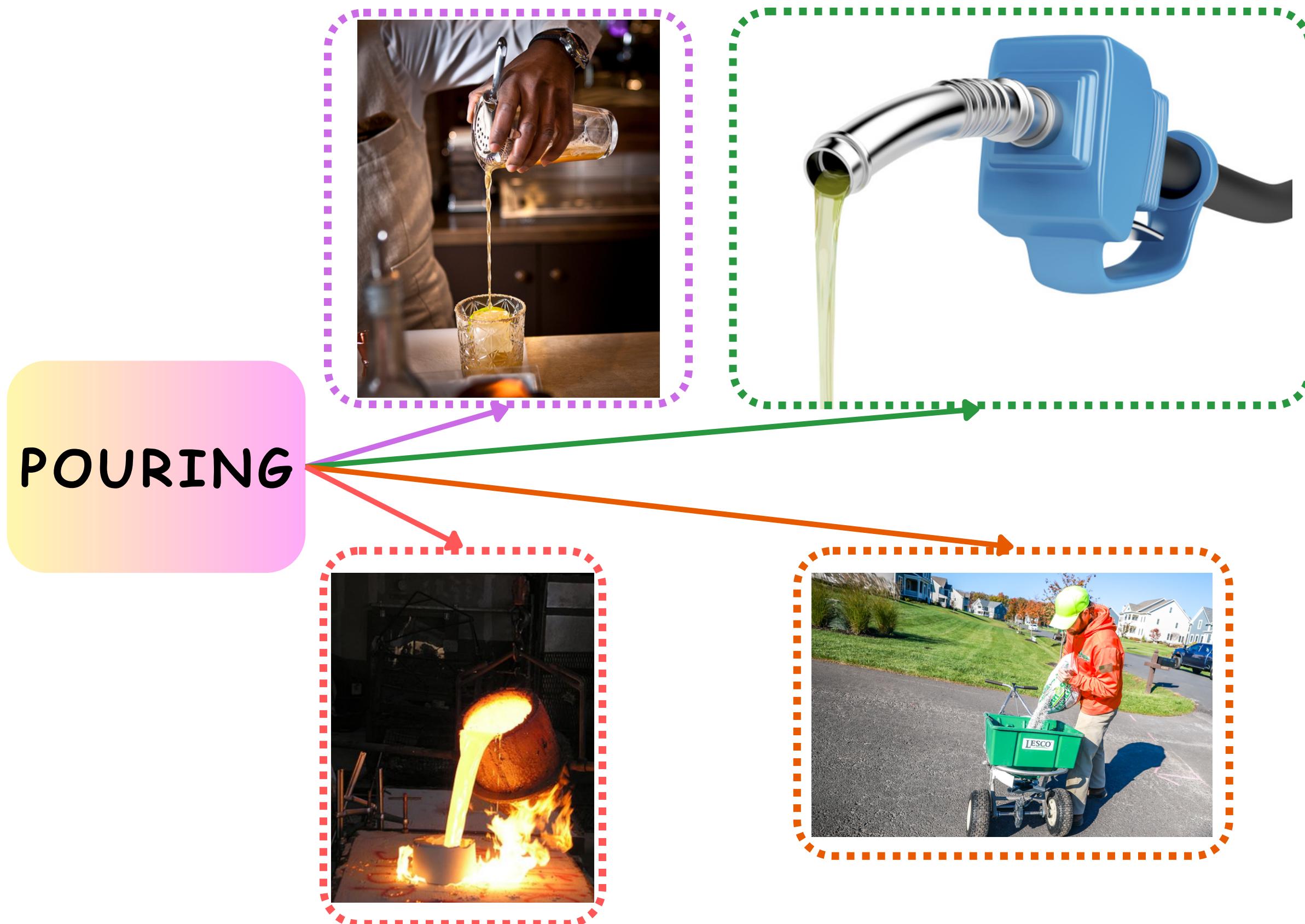
Motivation: Anomaly Detection for Pouring Task



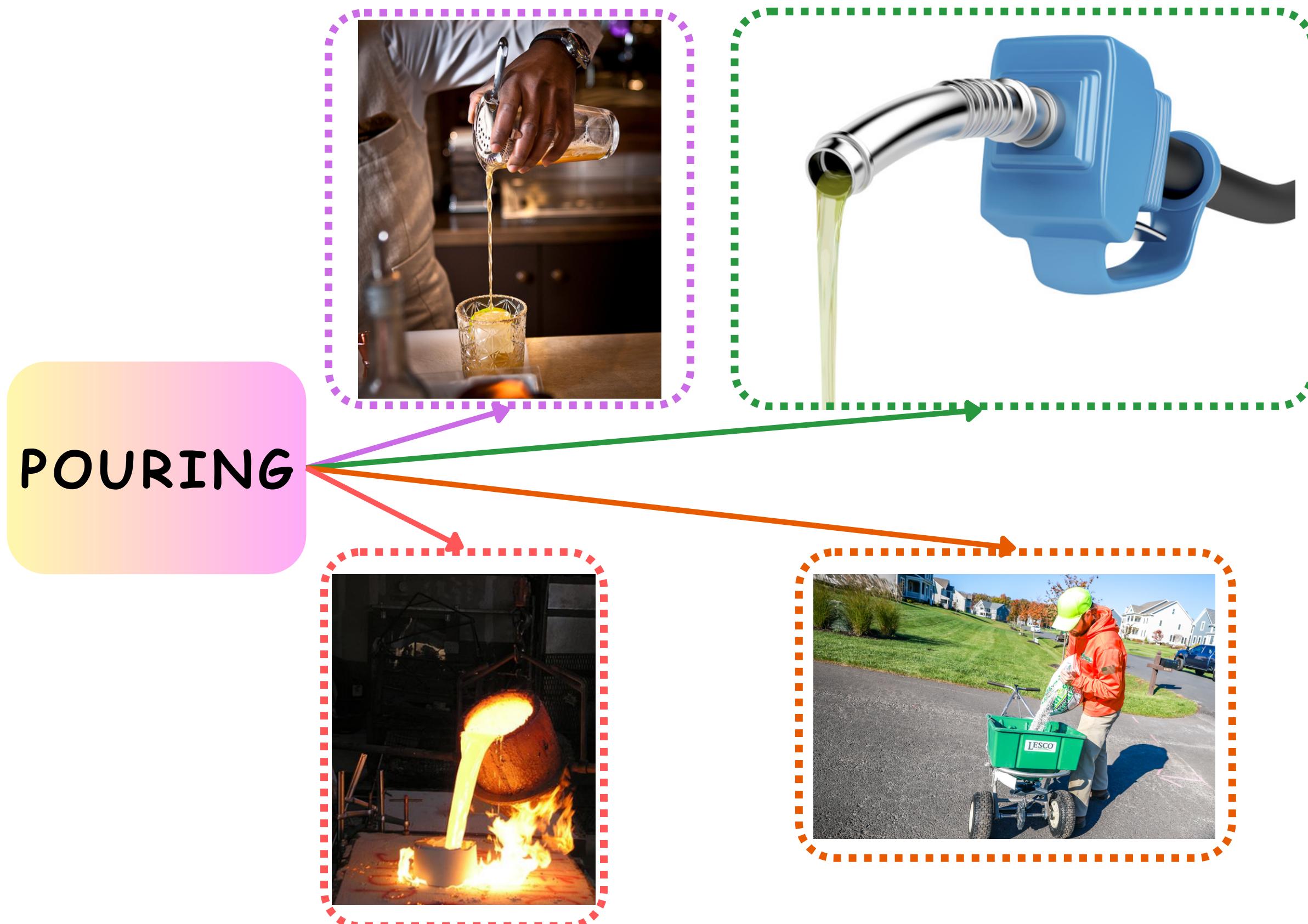
Motivation: Anomaly Detection for Pouring Task



Motivation: Anomaly Detection for Pouring Task



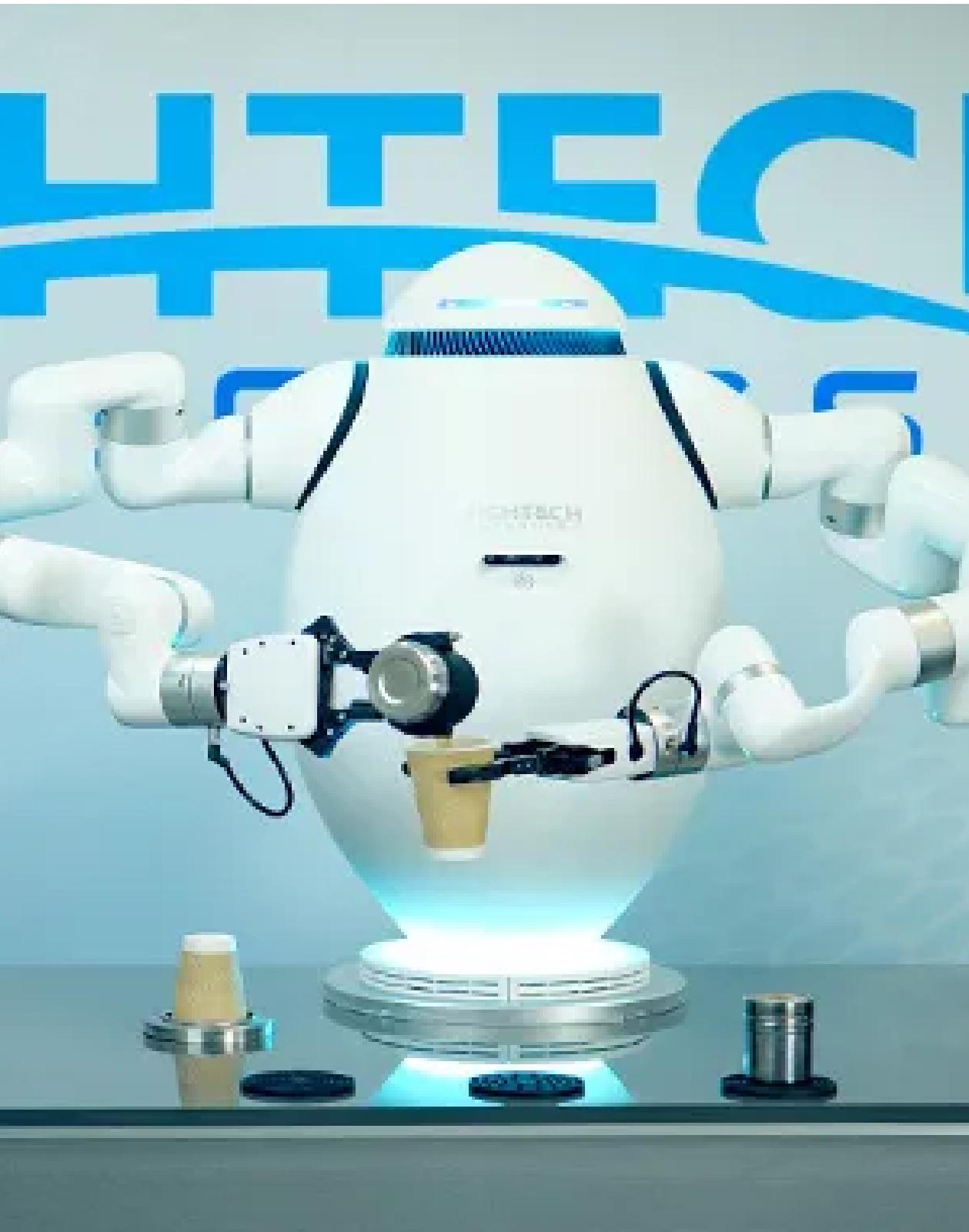
Motivation: Anomaly Detection for Pouring Task





Hochschule
Bonn-Rhein-Sieg

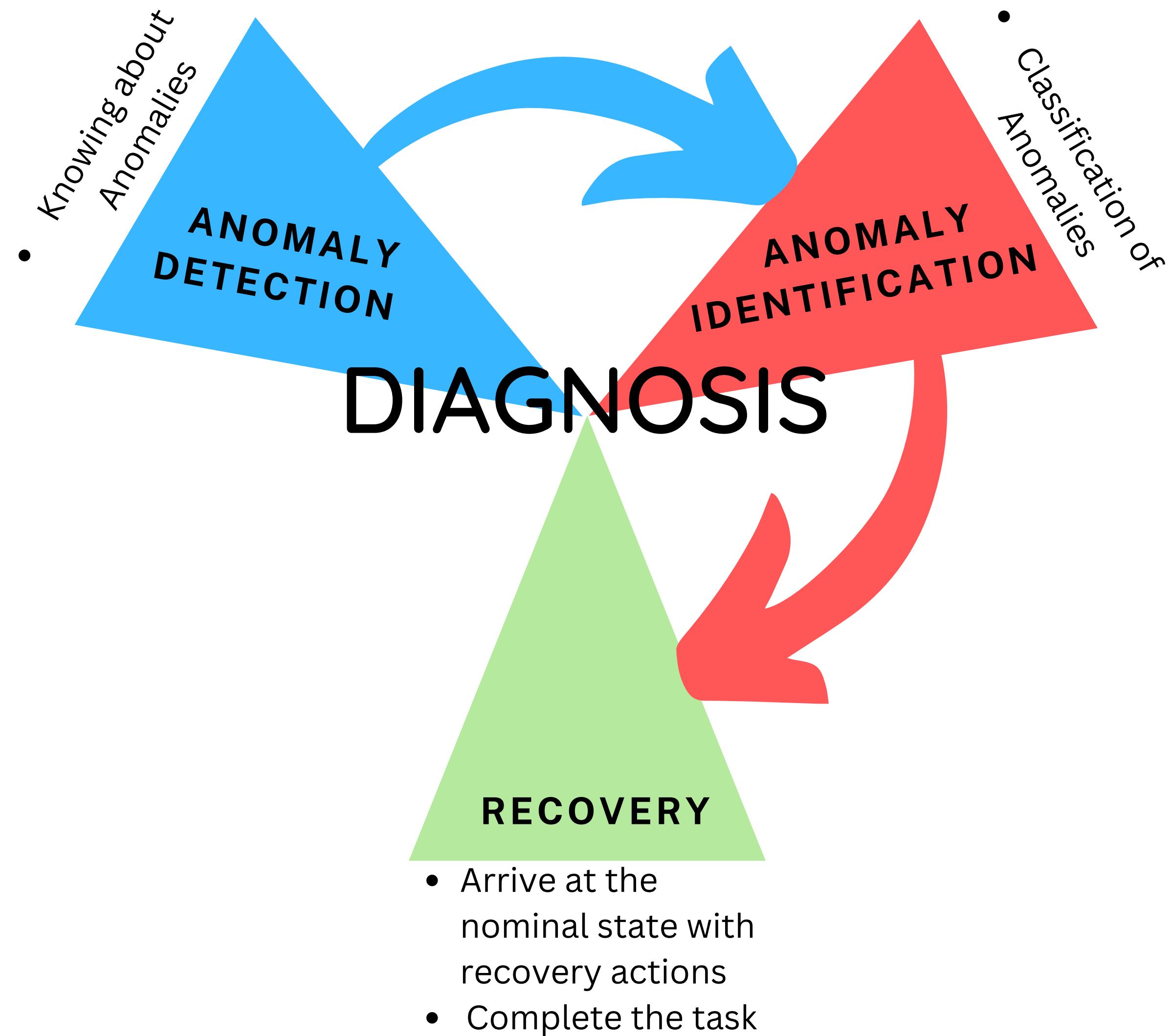




Assistive Robots

- Robots also suffer from anomalies
- Robots should be aware of Anomalies

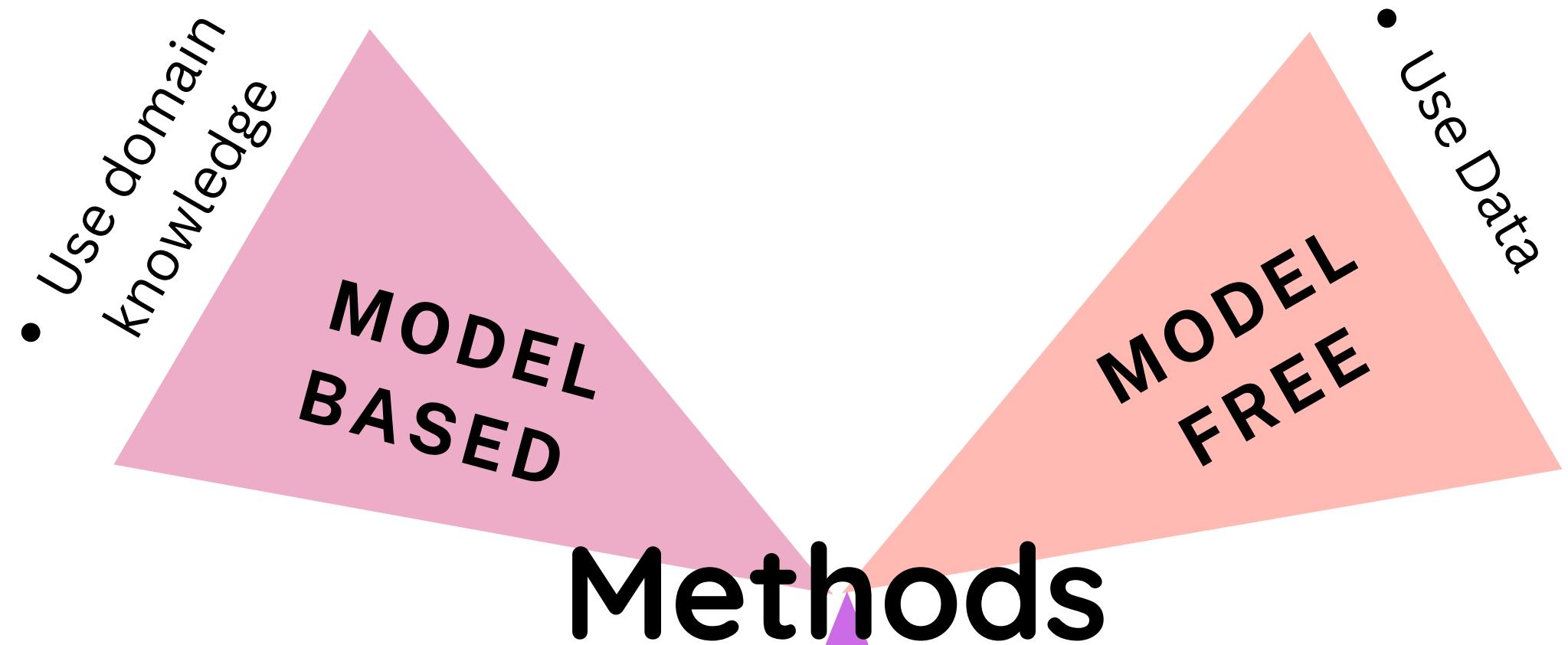
Copyright: Richtech Robotics





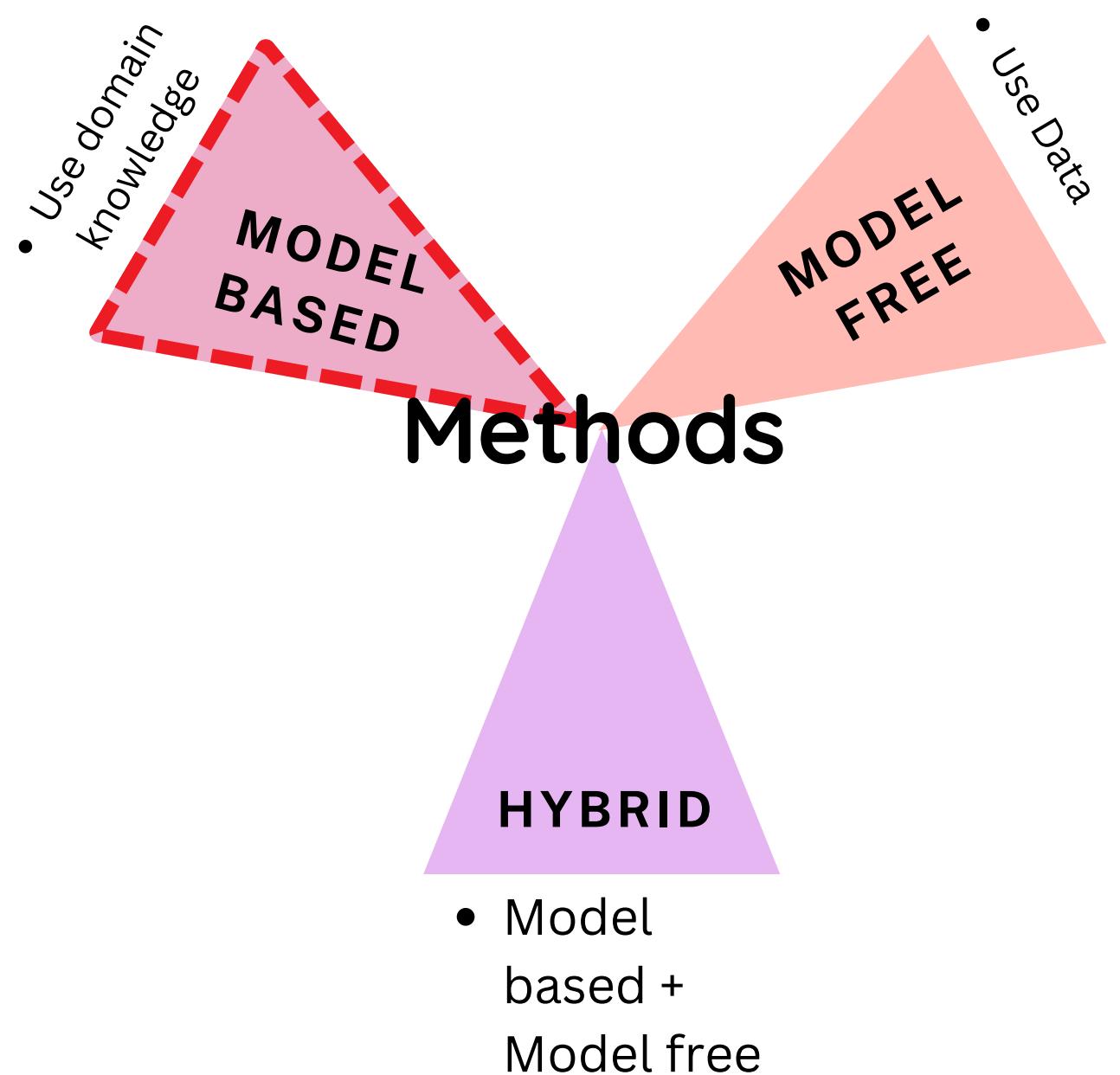
Hochschule
Bonn-Rhein-Sieg

Literature Review

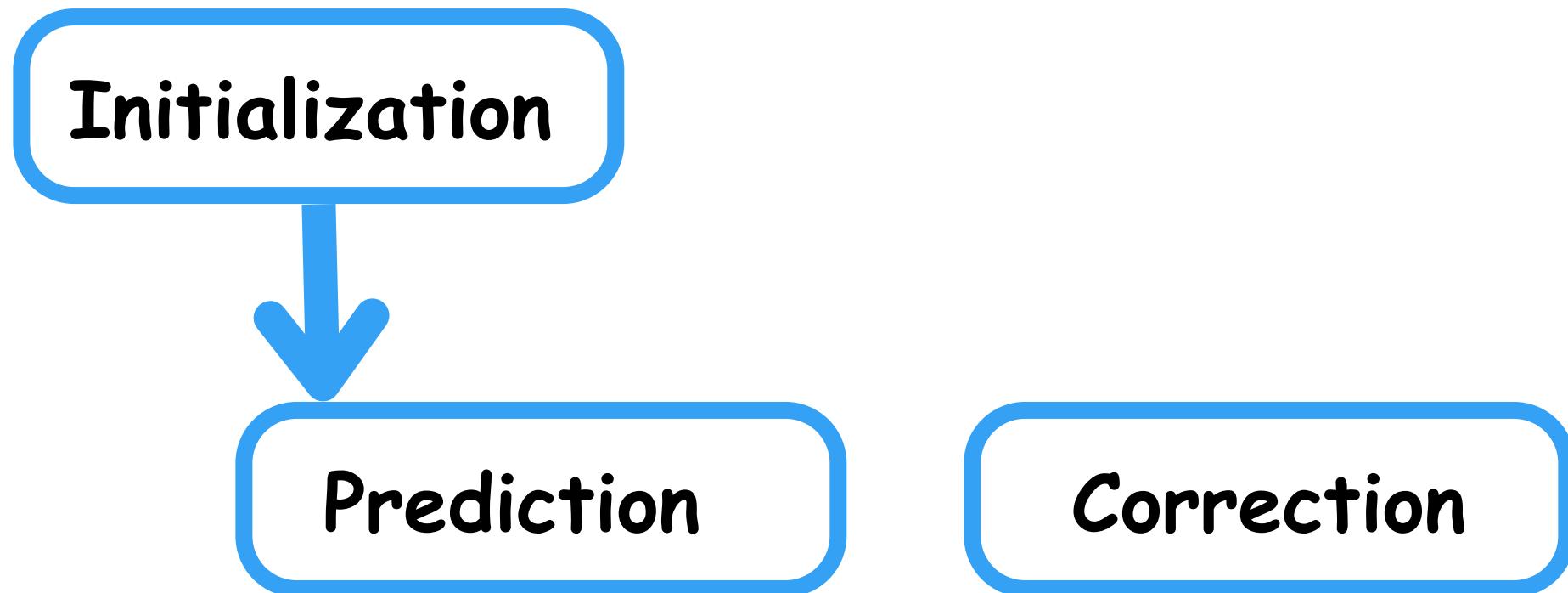
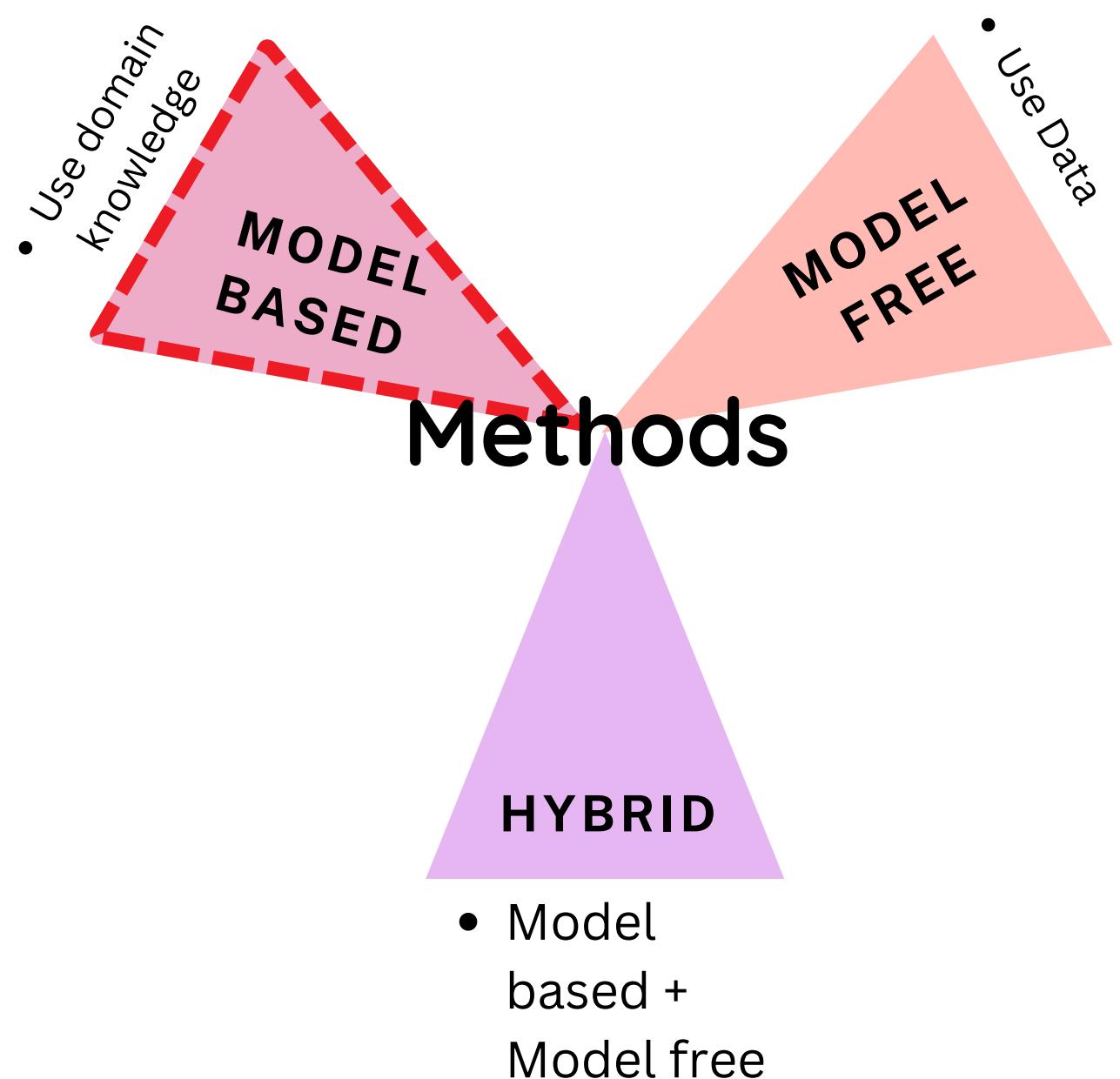


- Model based + Model free

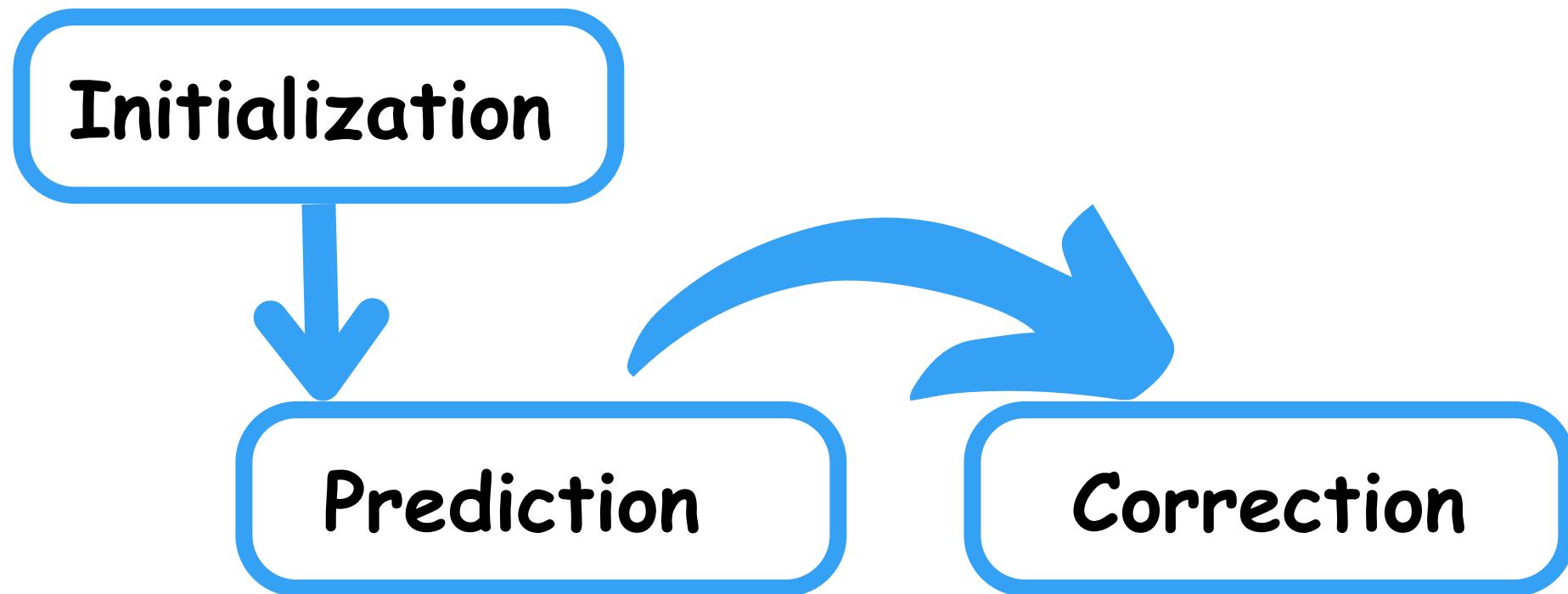
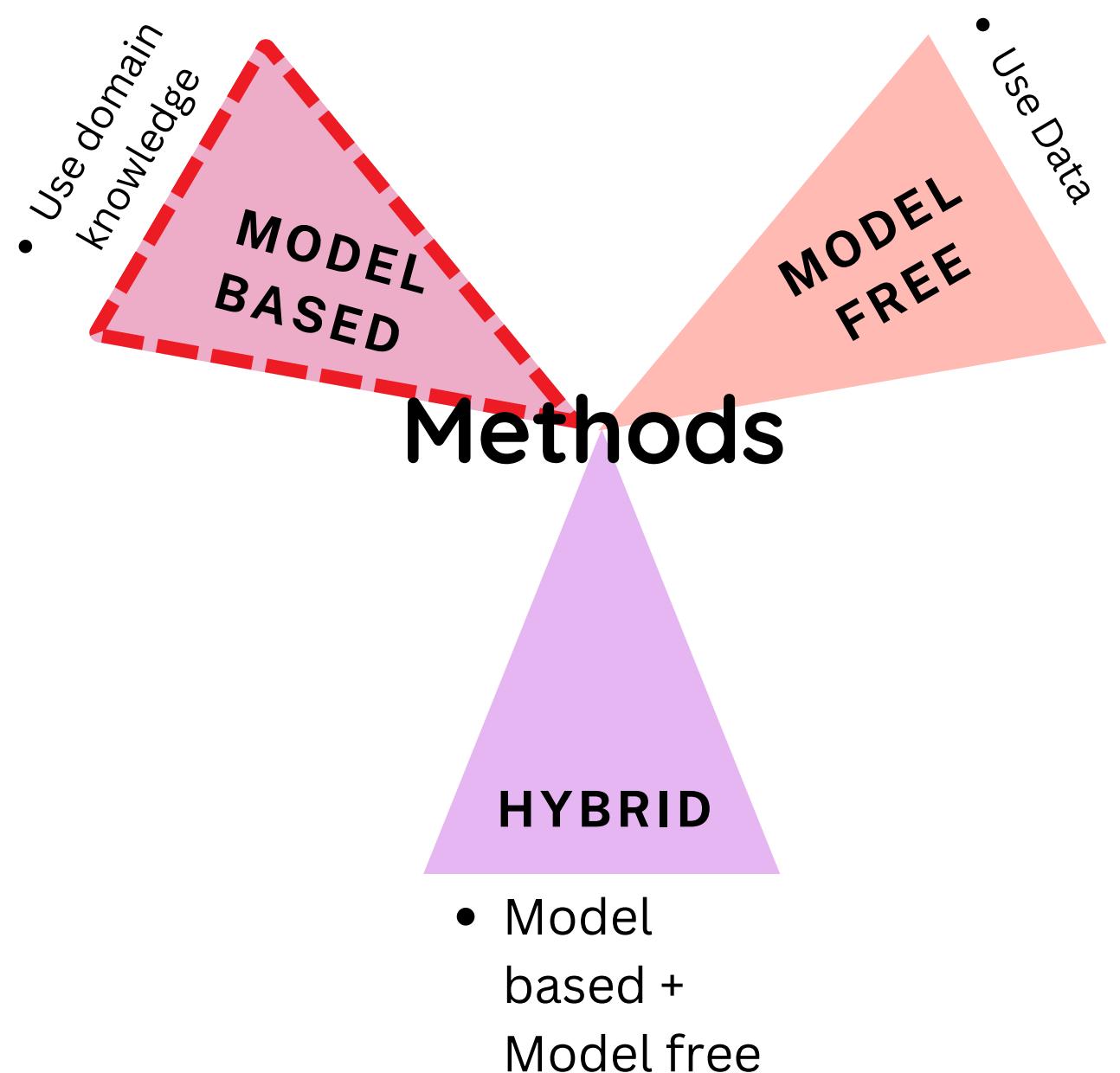
Initialization



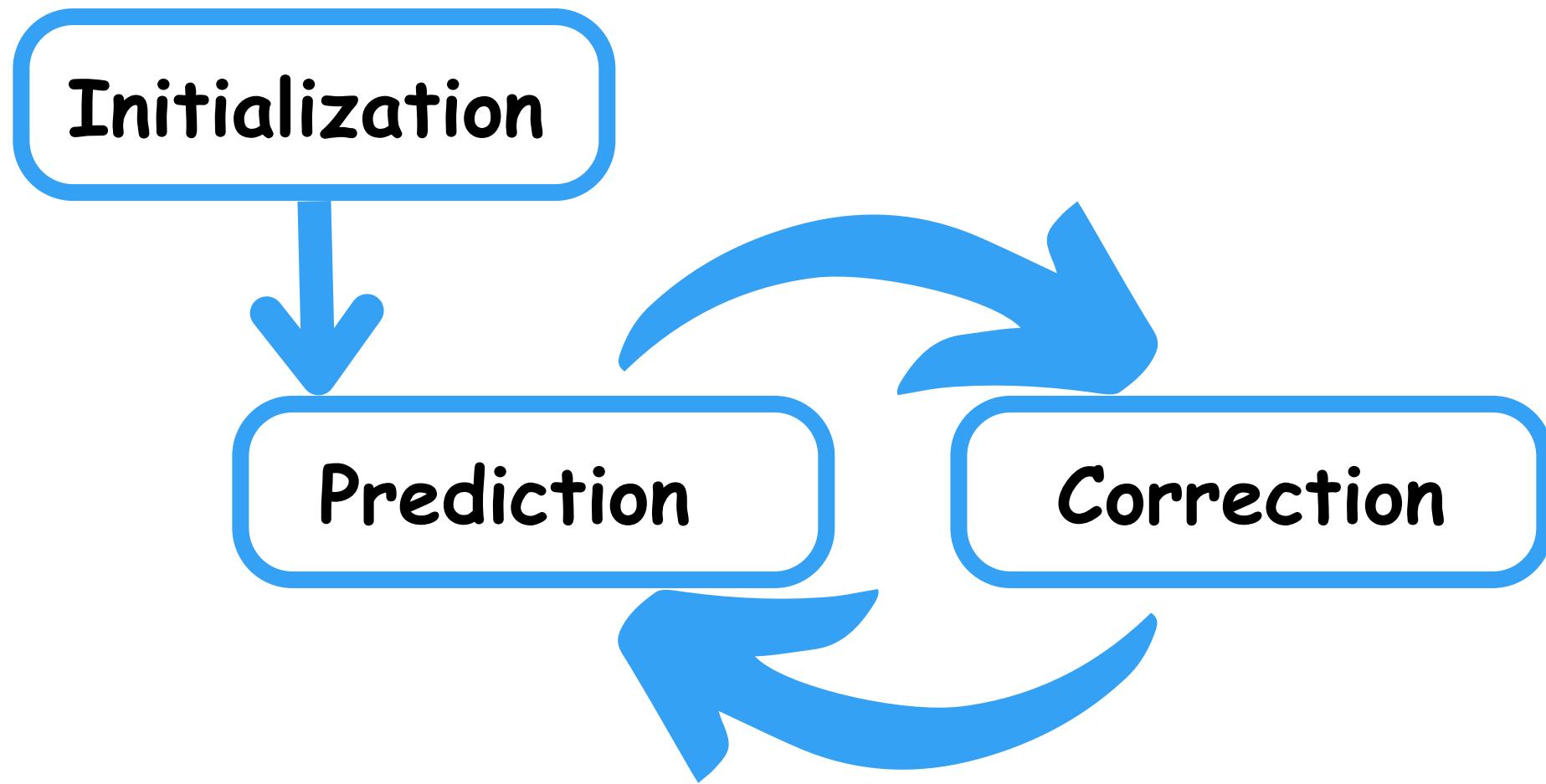
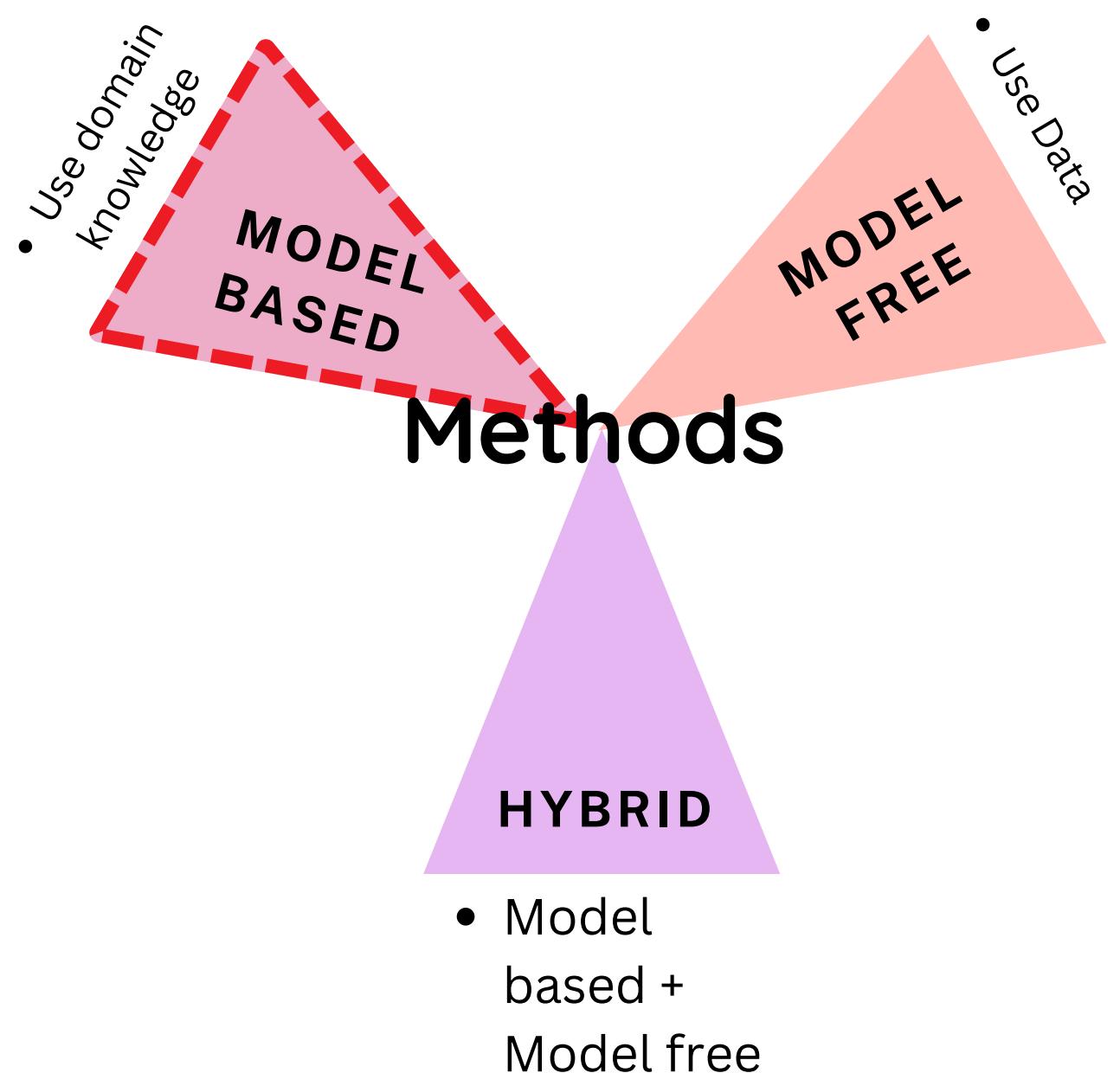
- Calculate "Disparity"
- Disparity > Threshold (Anomaly Alert!)



- Calculate "Disparity"
- Disparity > Threshold (Anomaly Alert!)



- Calculate "Disparity"
- Disparity > Threshold (Anomaly Alert!)



- Calculate "Disparity"
- Disparity > Threshold (Anomaly Alert!)

• Use domain
knowledge

**MODEL
BASED**

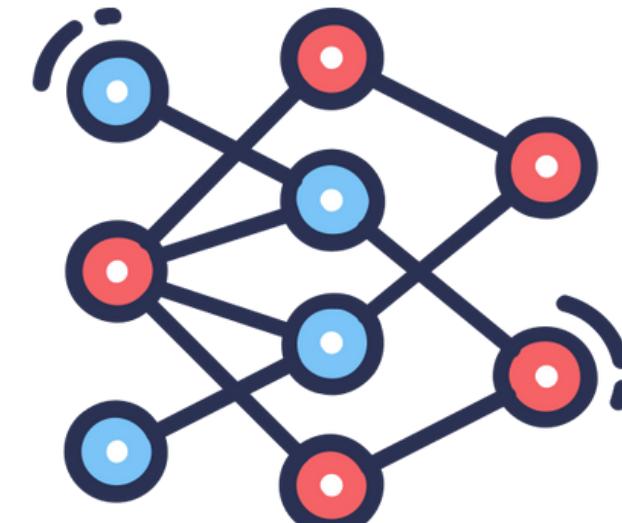
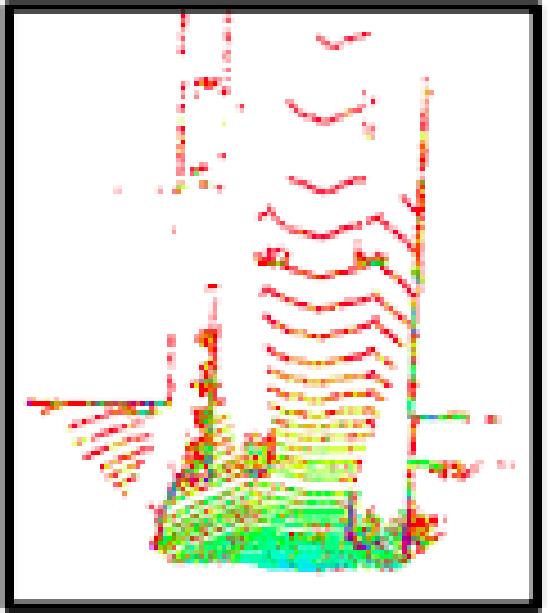
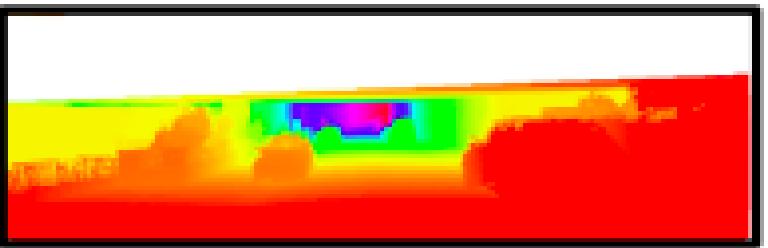
• Use Data

**MODEL
FREE**

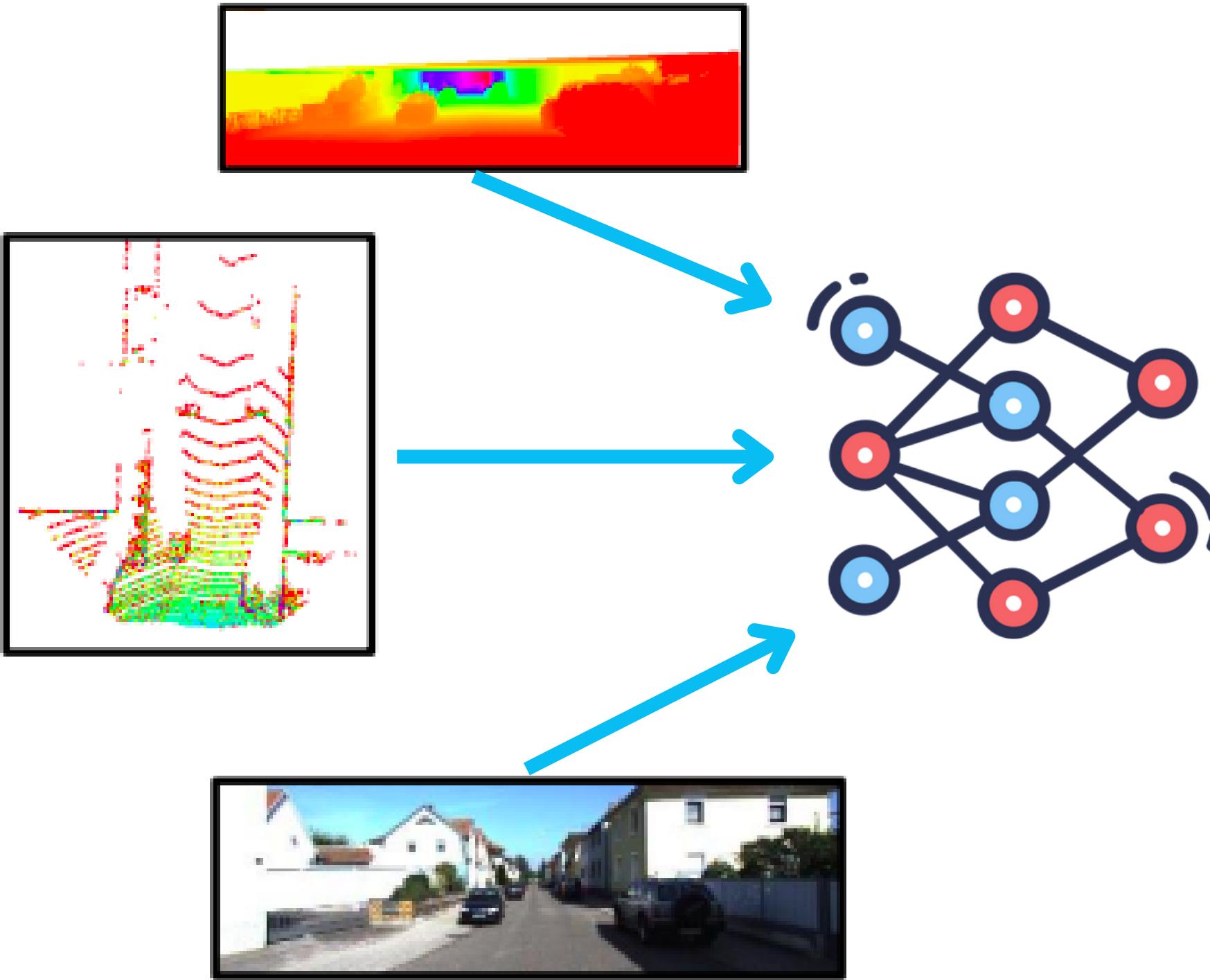
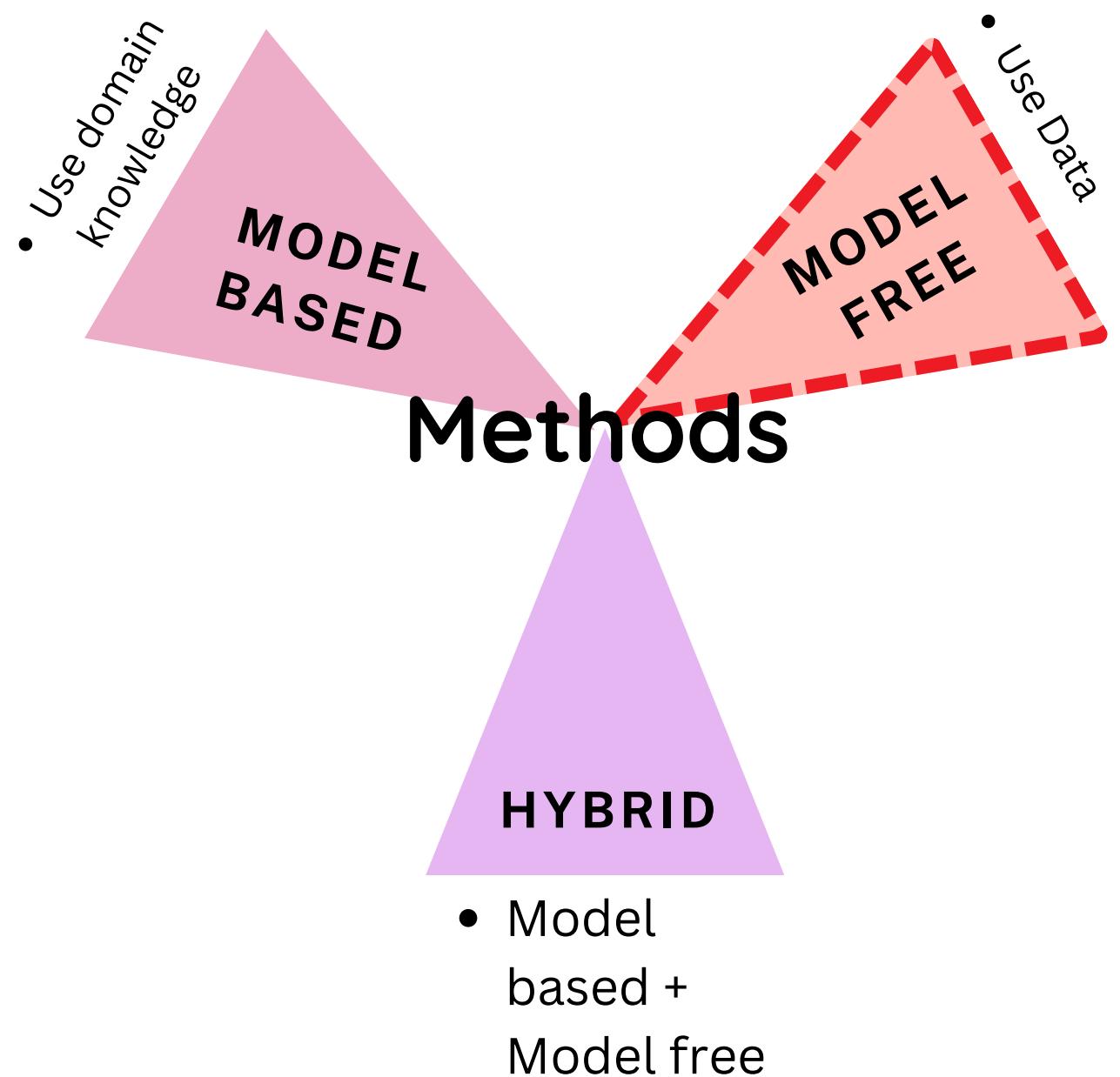
Methods

HYBRID

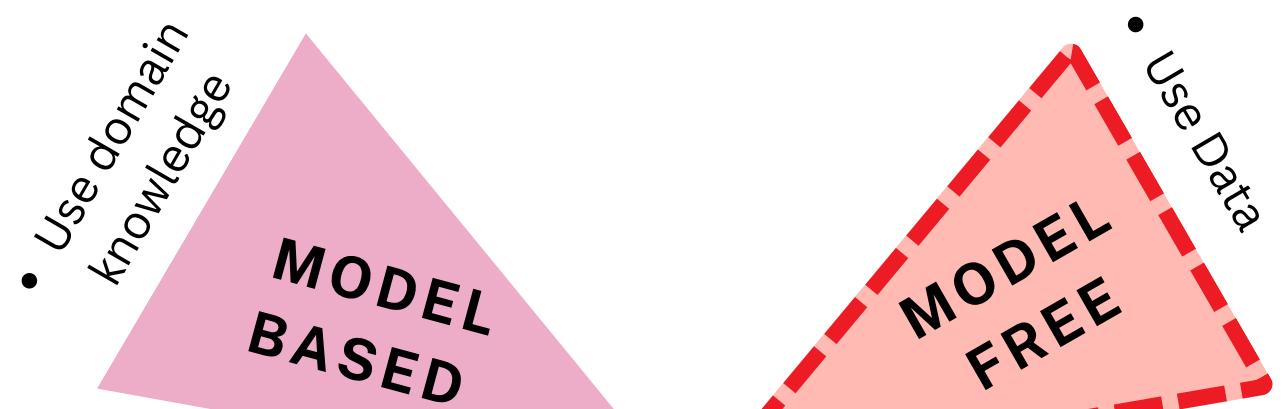
- Model
based +
Model free



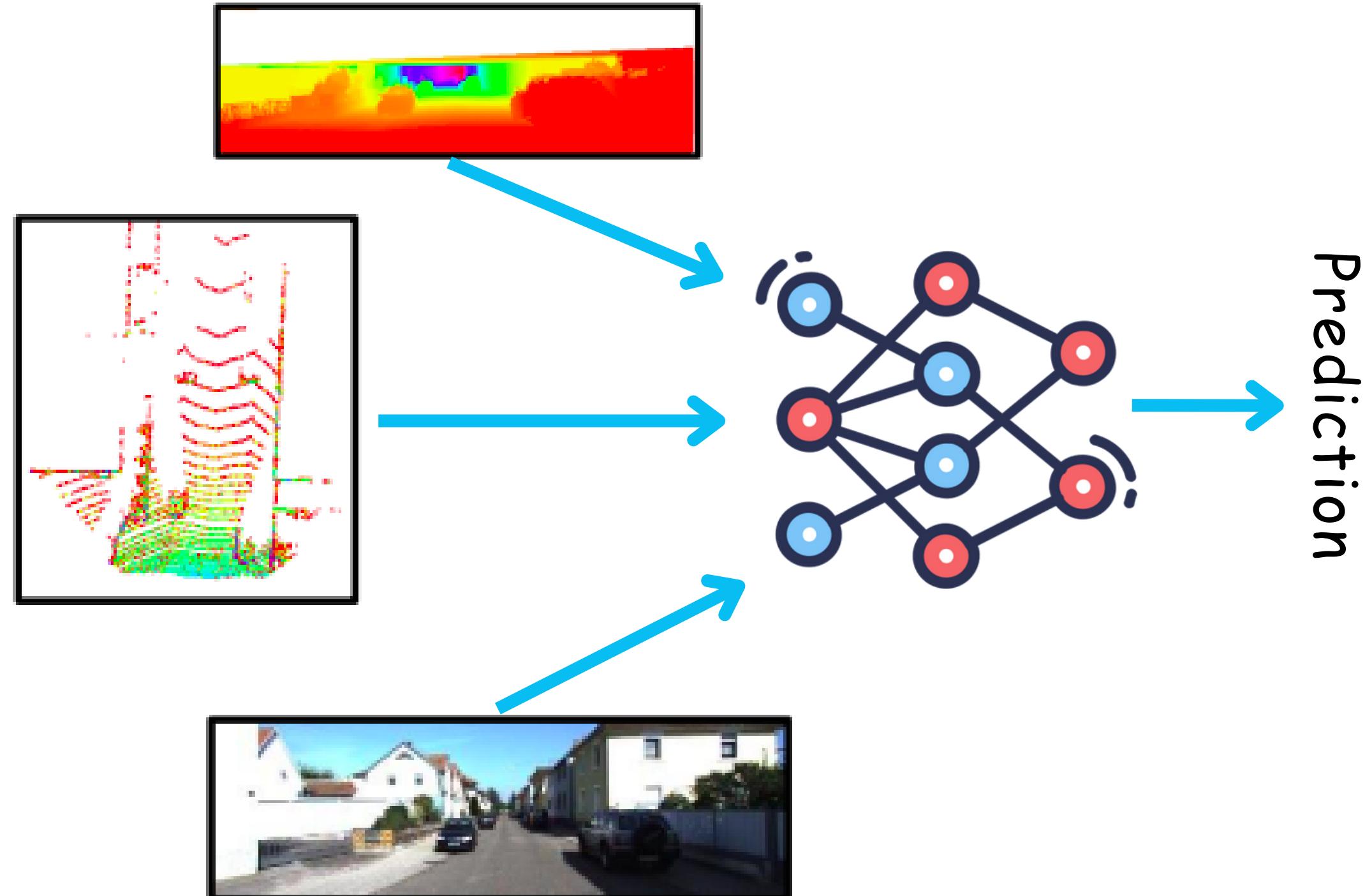
- Prediction based on sensor
input



- Prediction based on sensor input

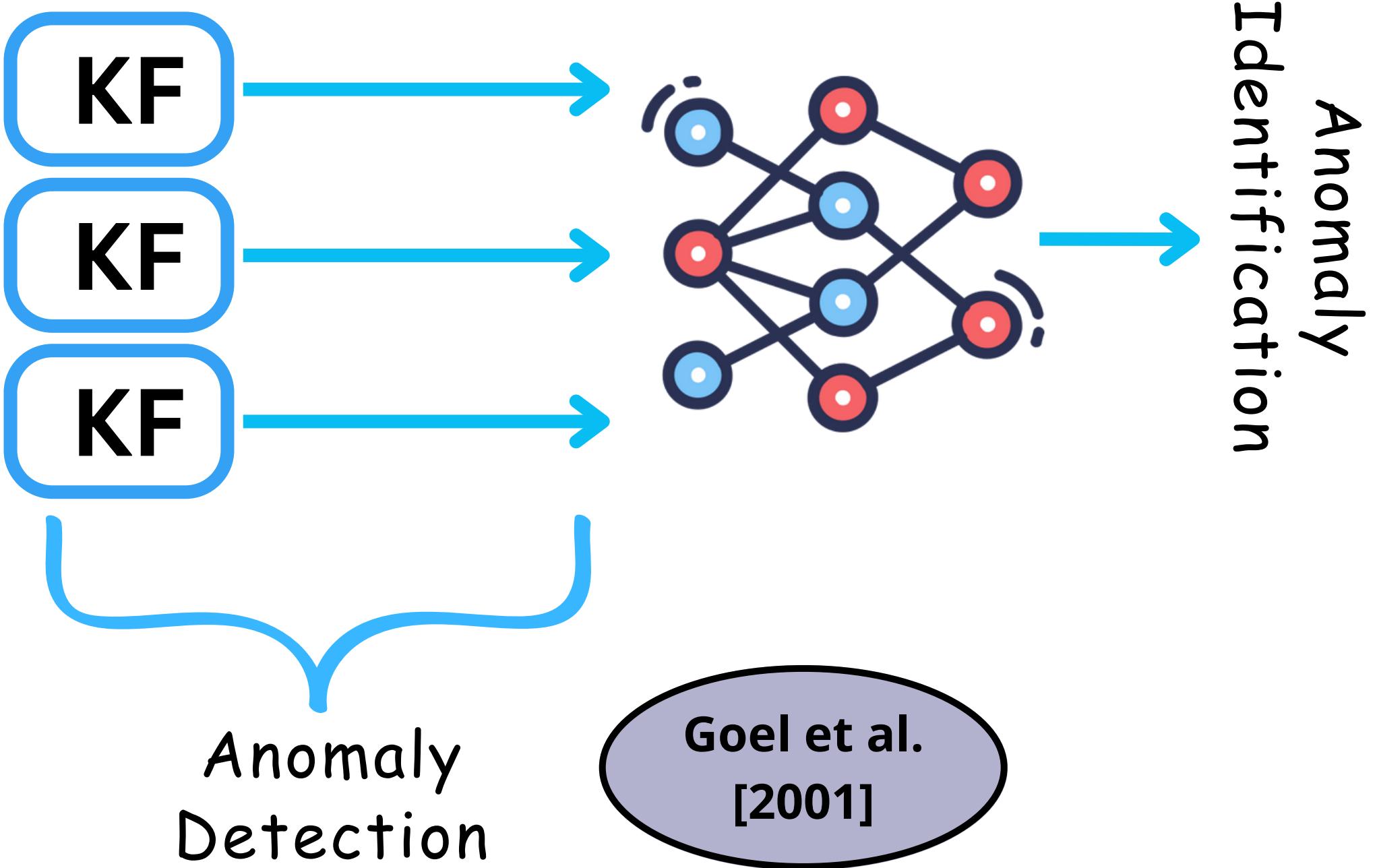
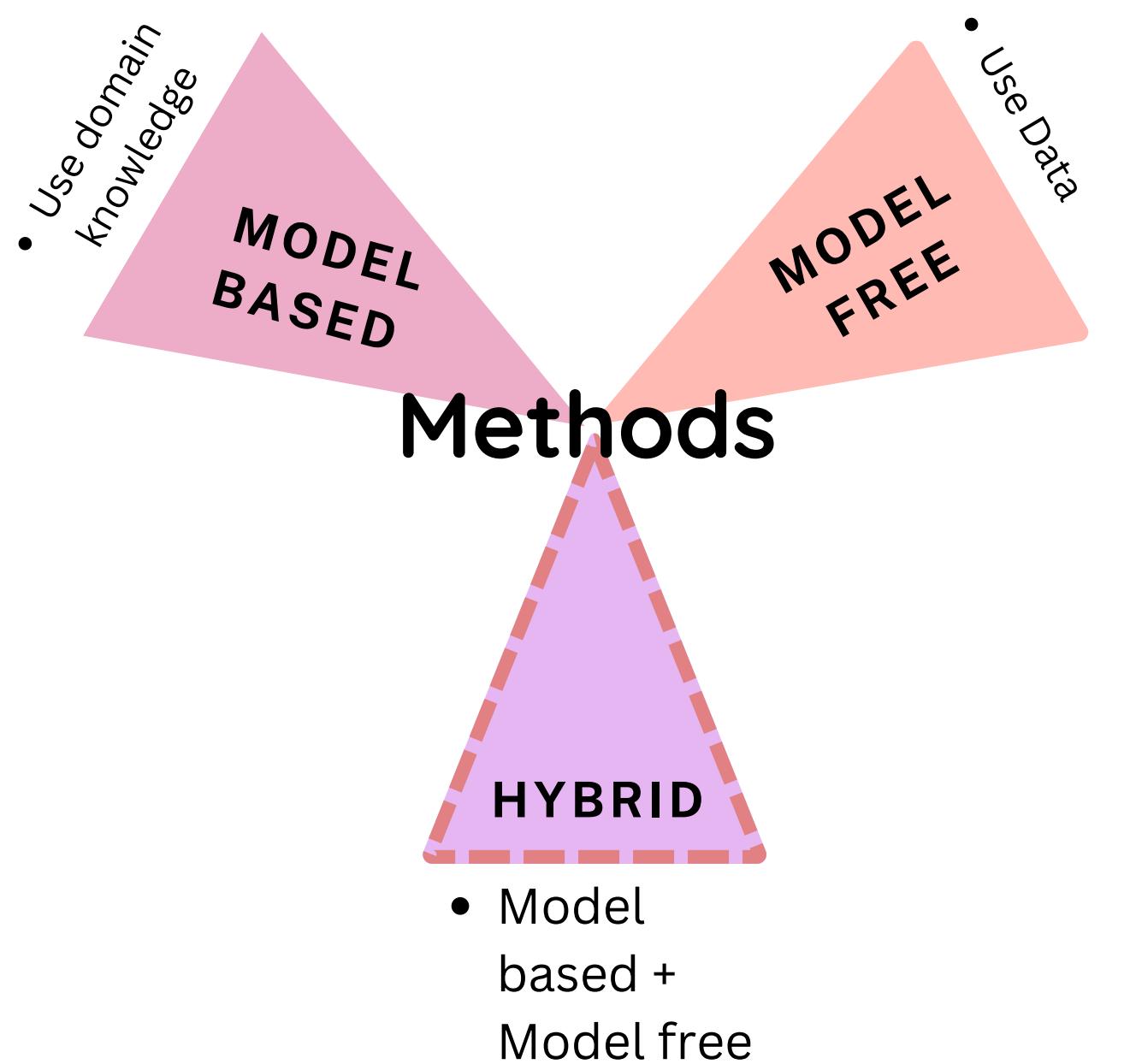


Methods

- 
- A diagram illustrating a prediction process. It shows three inputs: a thermal image of clouds, a 3D point cloud of a building, and a street view image. Arrows point from each input to a central neural network diagram, which then points to a final arrow labeled "Prediction".
- Model based +
Model free

- Prediction based on sensor input

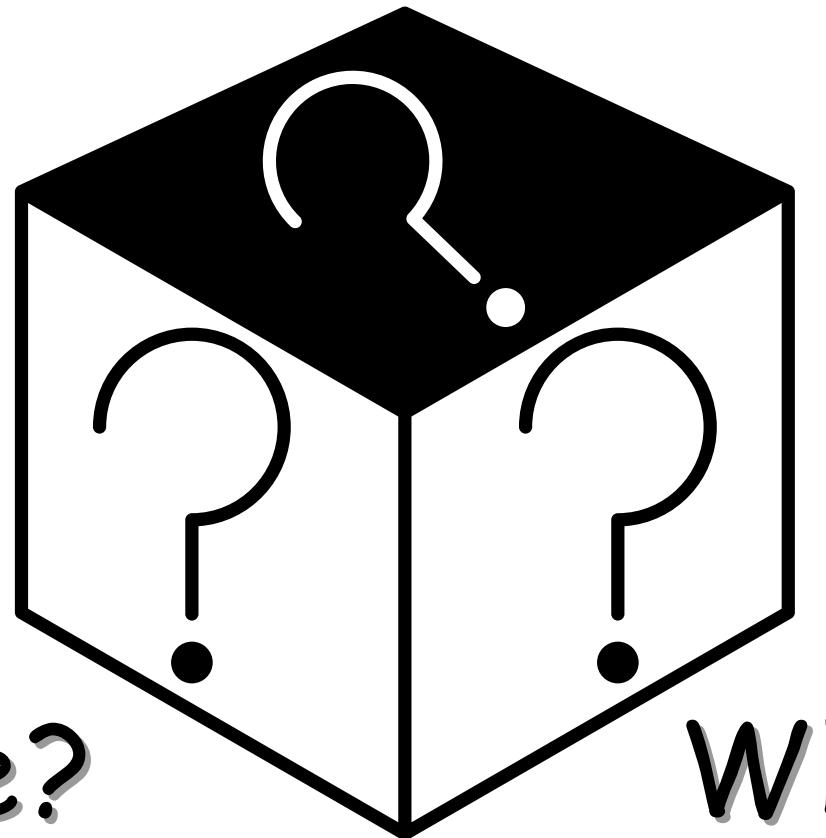
Prediction



Multimodal Sensor Fusion

- Use information from multiple sensors

What to fuse?

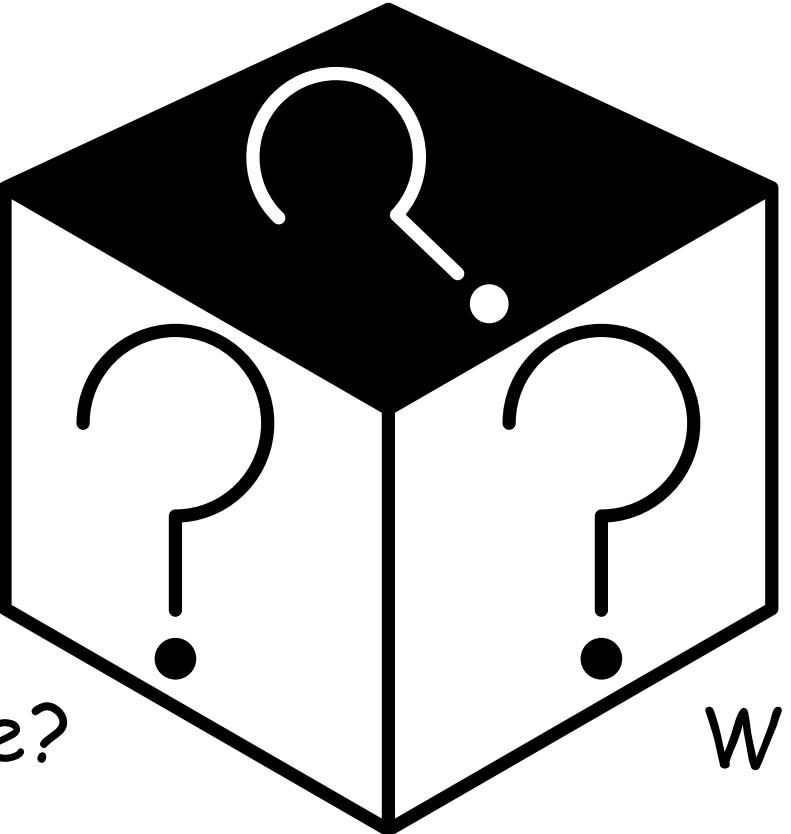


How to fuse?

When to fuse?

Multimodal Sensor Fusion

What to fuse?



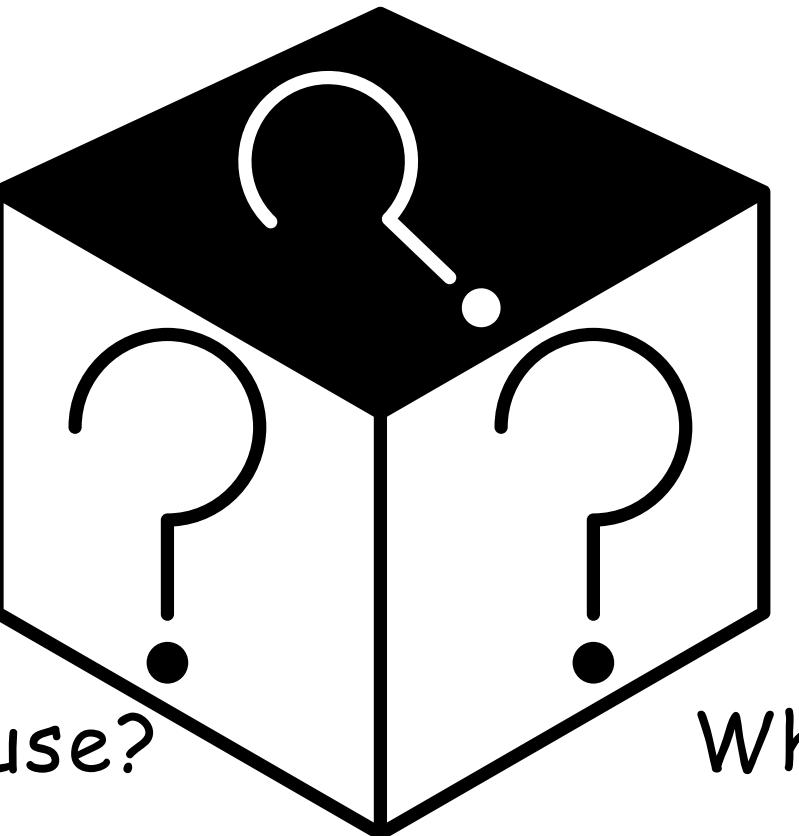
How to fuse?

When to fuse?

- Task dependent
- Example:
 - Vision + Tactile + sound for Pouring

Multimodal Sensor Fusion

How to fuse?



What to fuse?

When to fuse?

CONCATENATION

Often used
in early and
middle

Fusion Operation

**MIXTURE
OF
EXPERTS**

Often used
in middle and
late fusion

**ADDITION
OR
AVERAGE**

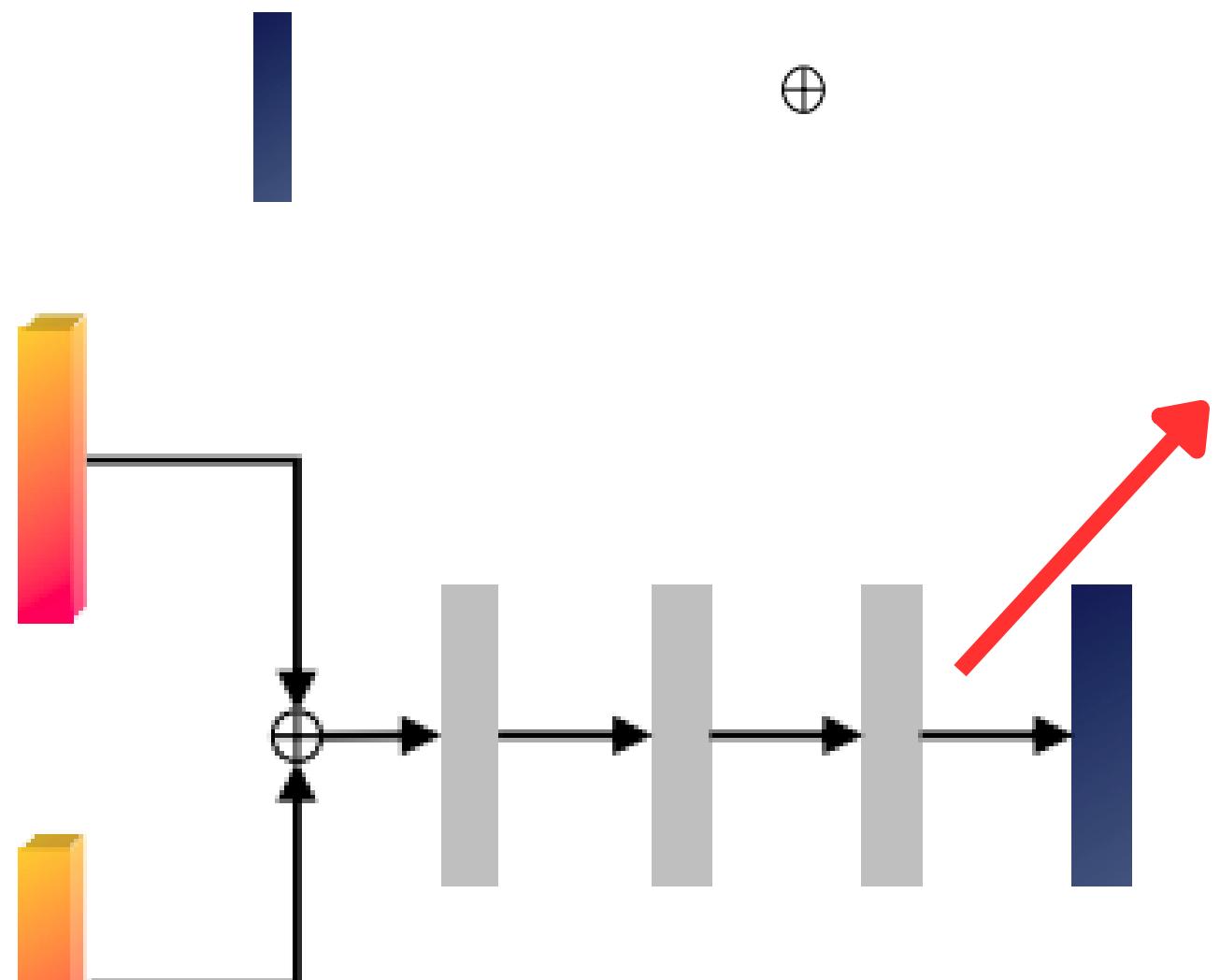
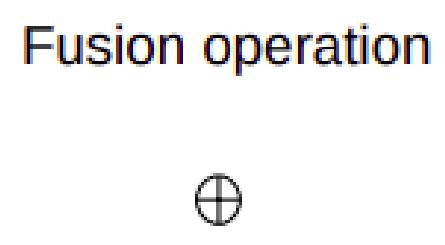
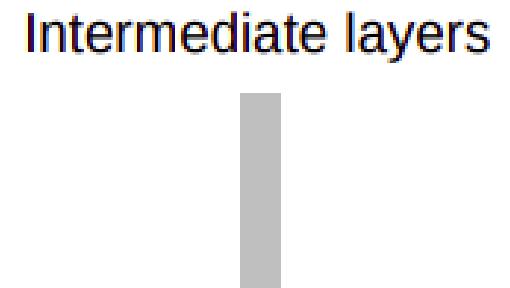
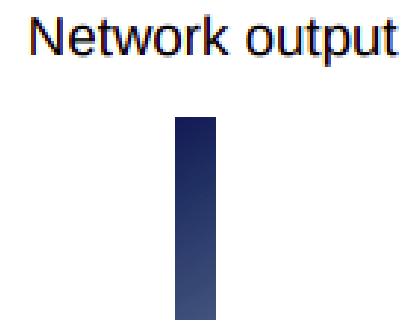
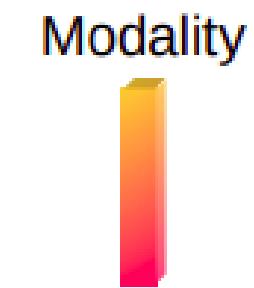
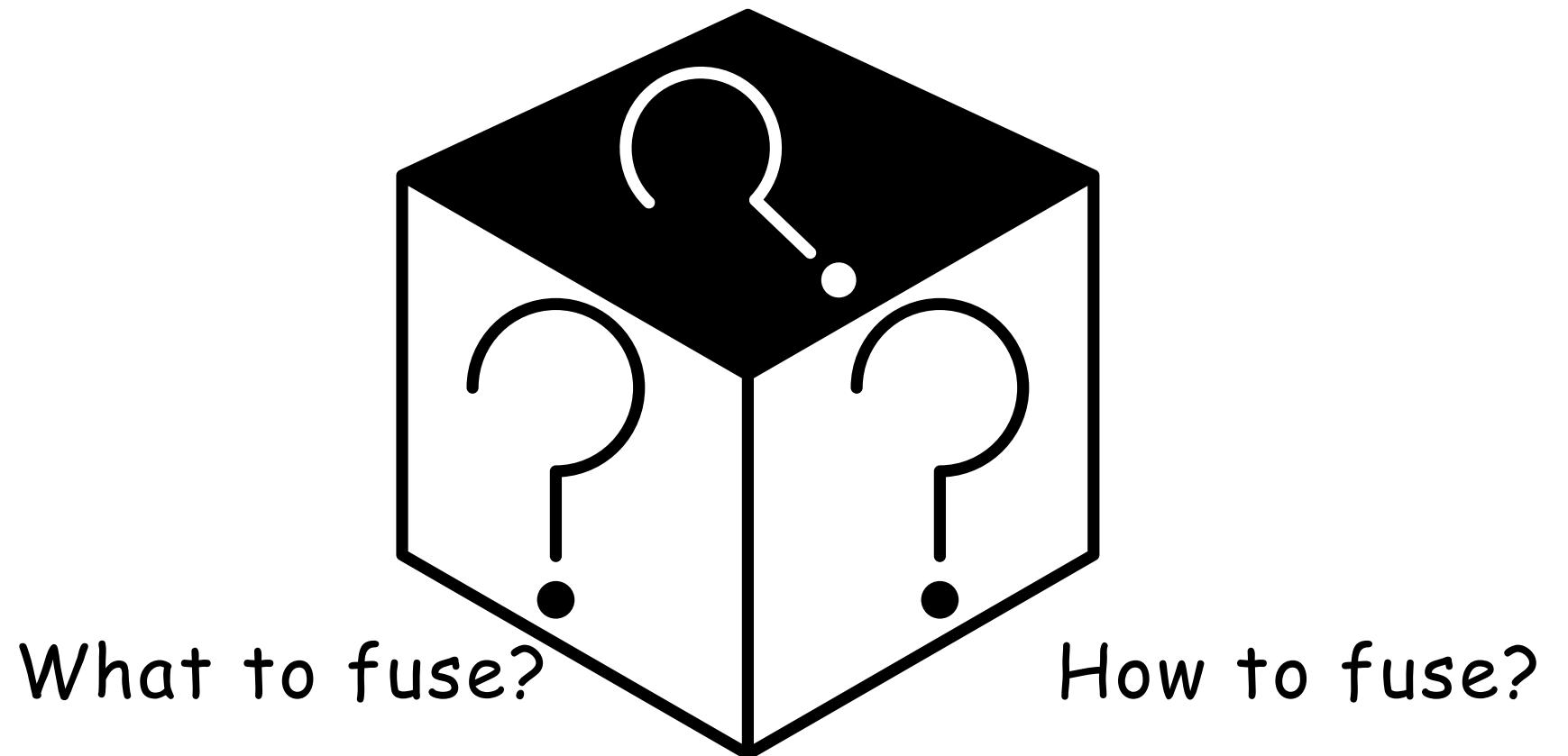
Used in
middle
fusion

ENSEMBLE

Often used
in middle and
late fusion

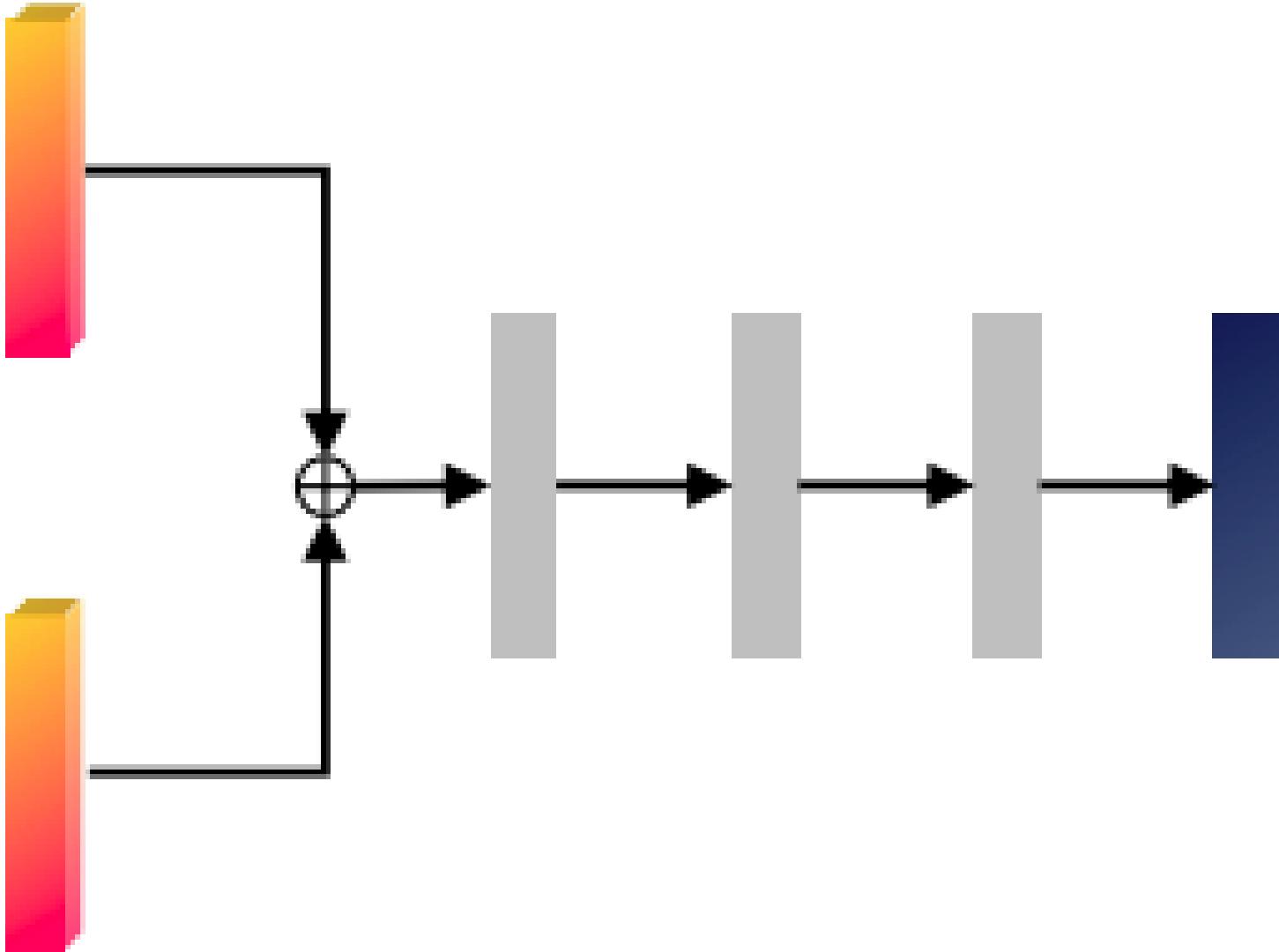
Multimodal Sensor Fusion

When to fuse?



Early Fusion

- Combine raw observations



LOW
COMPUTATION
REQUIREMENTS

ADVANTAGE
1



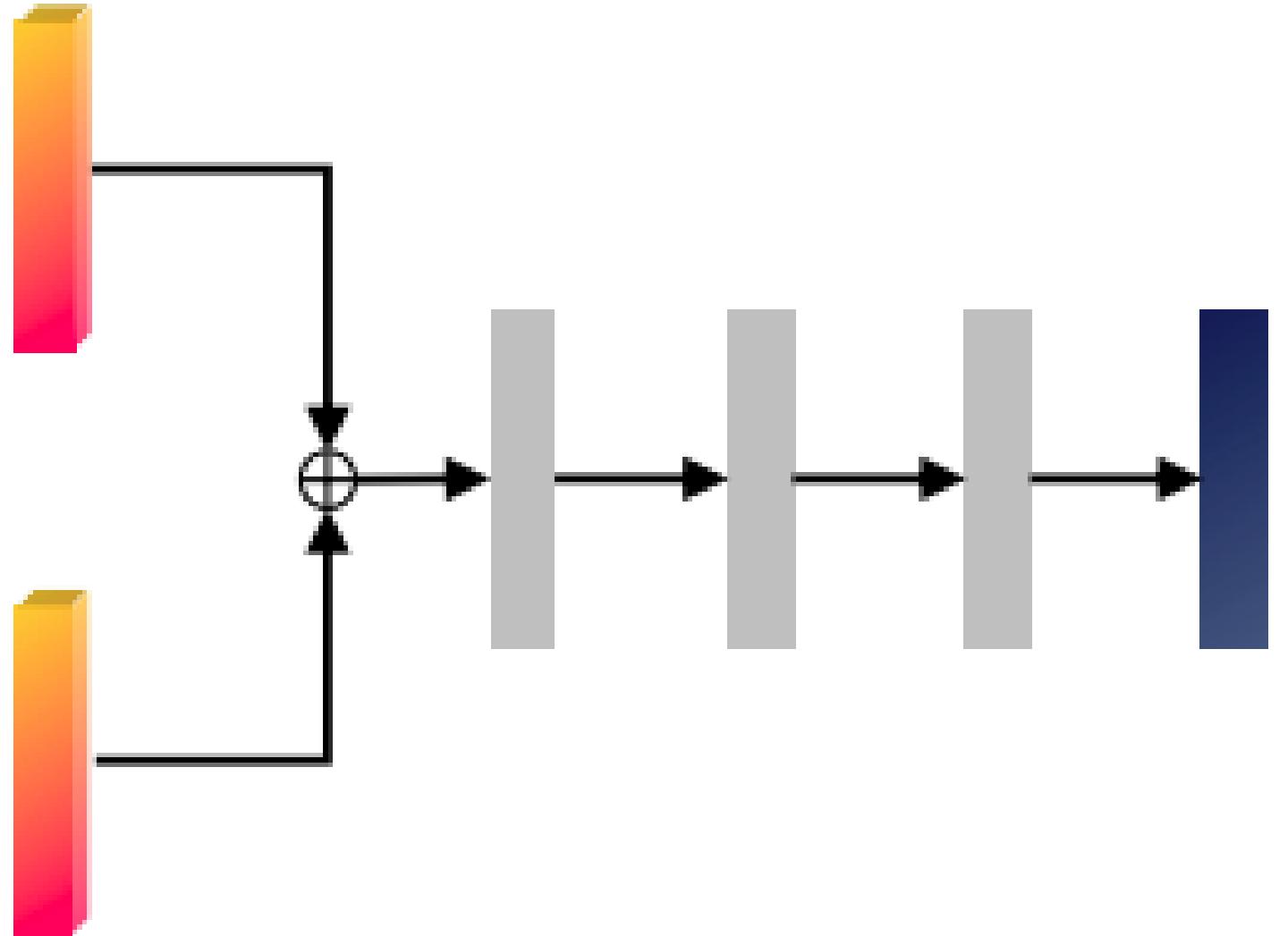
DISADVANTAGE
1



MODEL
INFLEXIBILITY

Early Fusion

- Combine raw observations



LOW MEMORY
BUDGET

ADVANTAGE
2



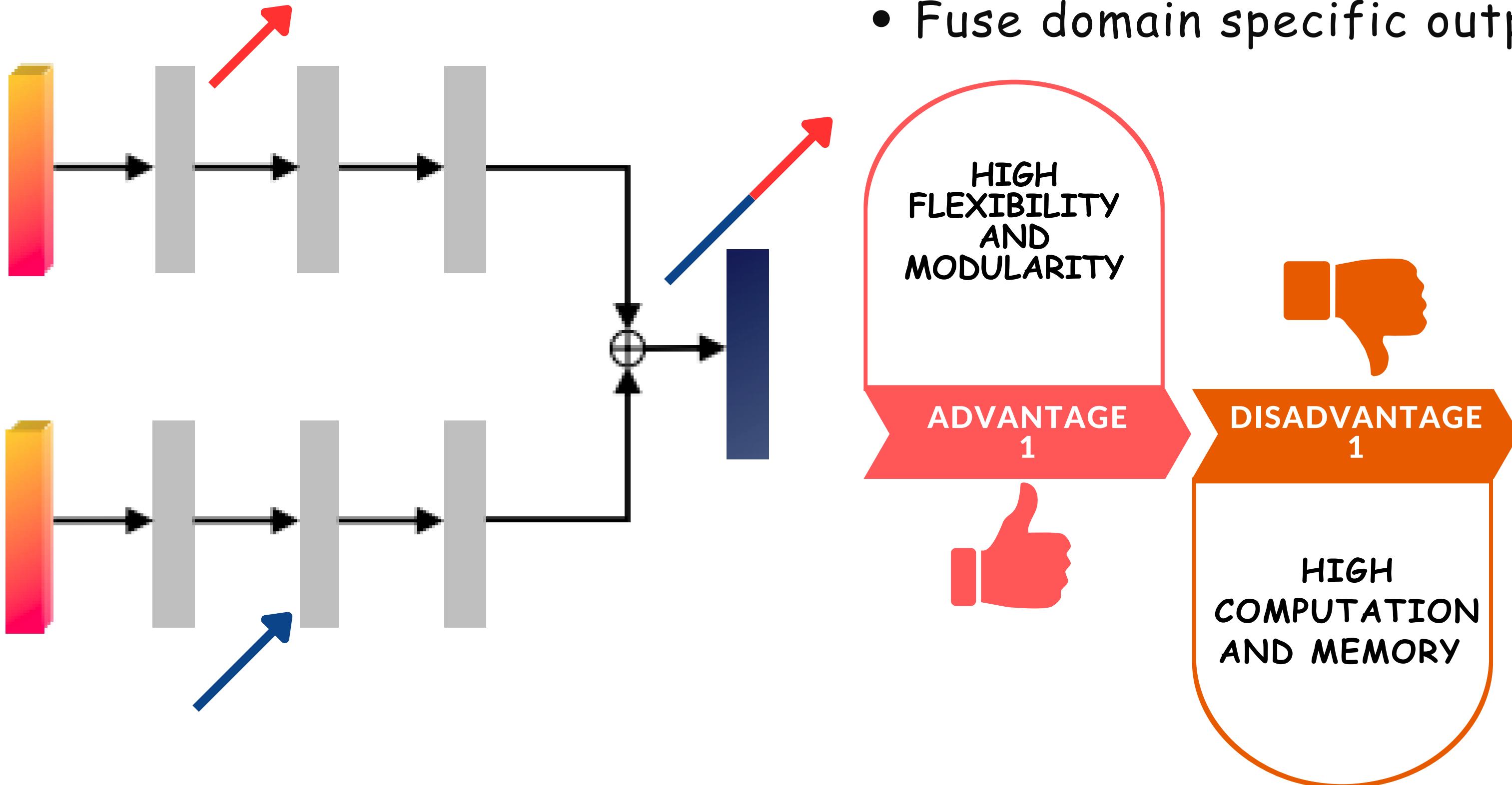
DISADVANTAGE
2

DATA
MISALIGNMENT



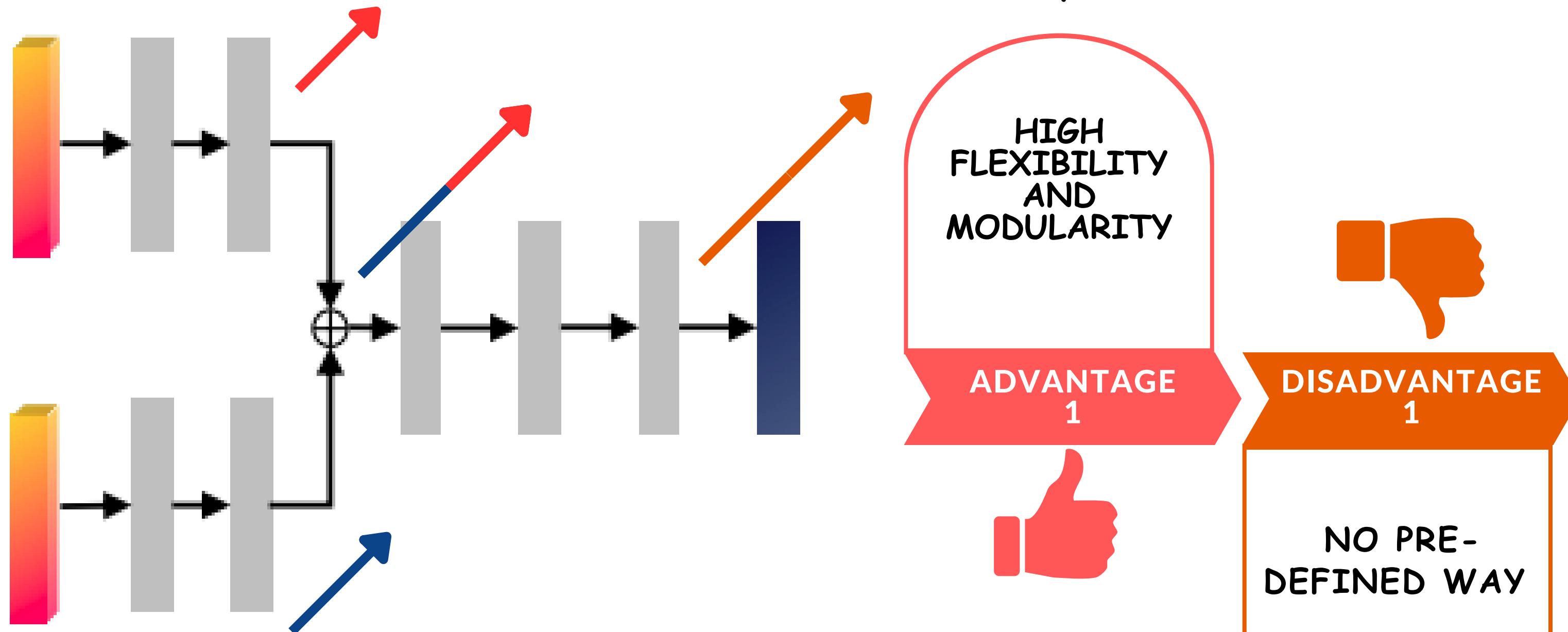
Late Fusion

- Fuse domain specific output



Middle Fusion

- Early fusion + Late fusion



SUPERVISED ANOMALY DETECTION

Dataset with
normal and
anomaly
instances

SEMI- SUPERVISED ANOMALY DETECTION

Dataset with
normal
instances
only

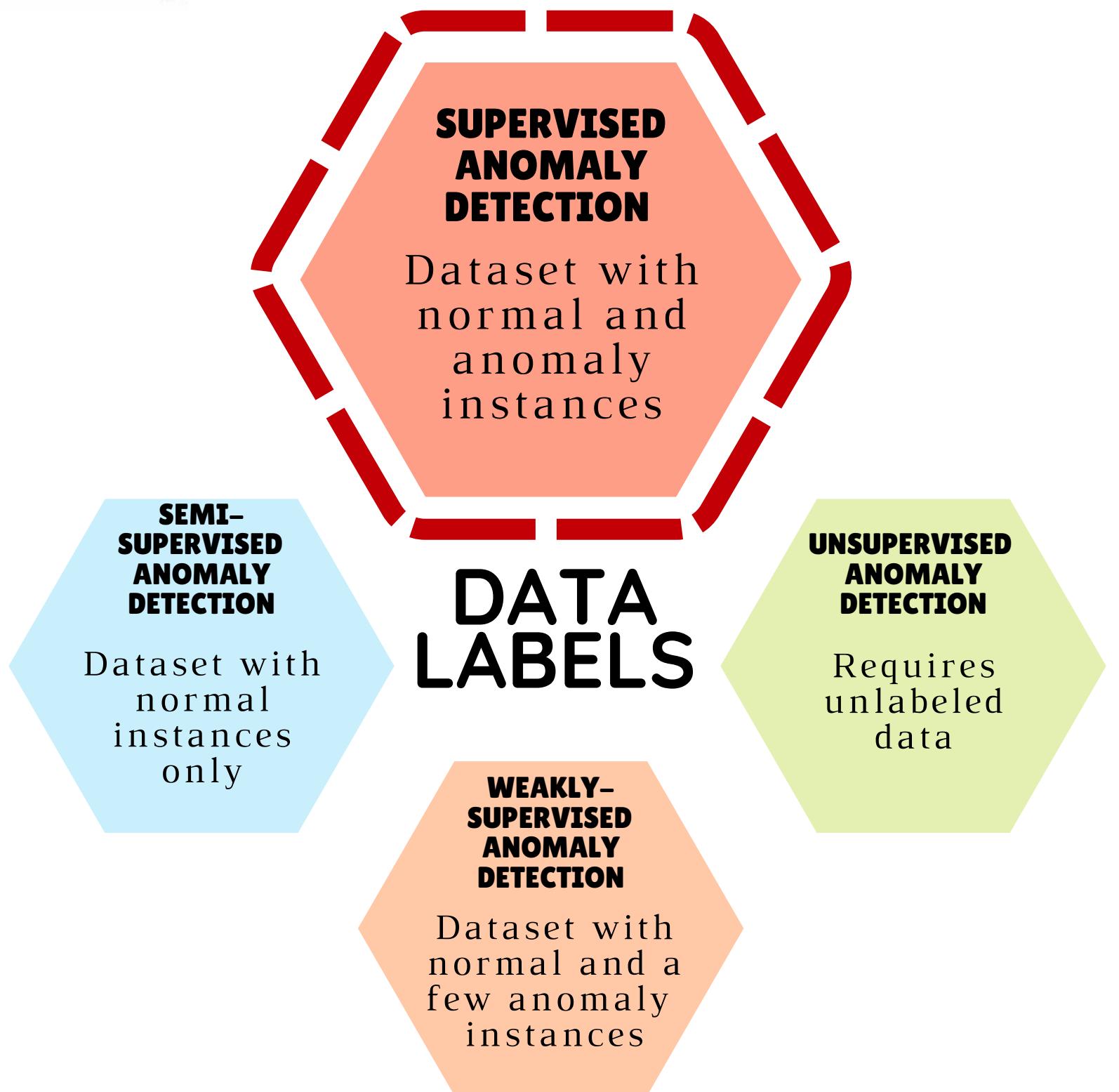
DATA LABELS

WEAKLY- SUPERVISED ANOMALY DETECTION

Dataset with
normal and a
few anomaly
instances

UNSUPERVISED ANOMALY DETECTION

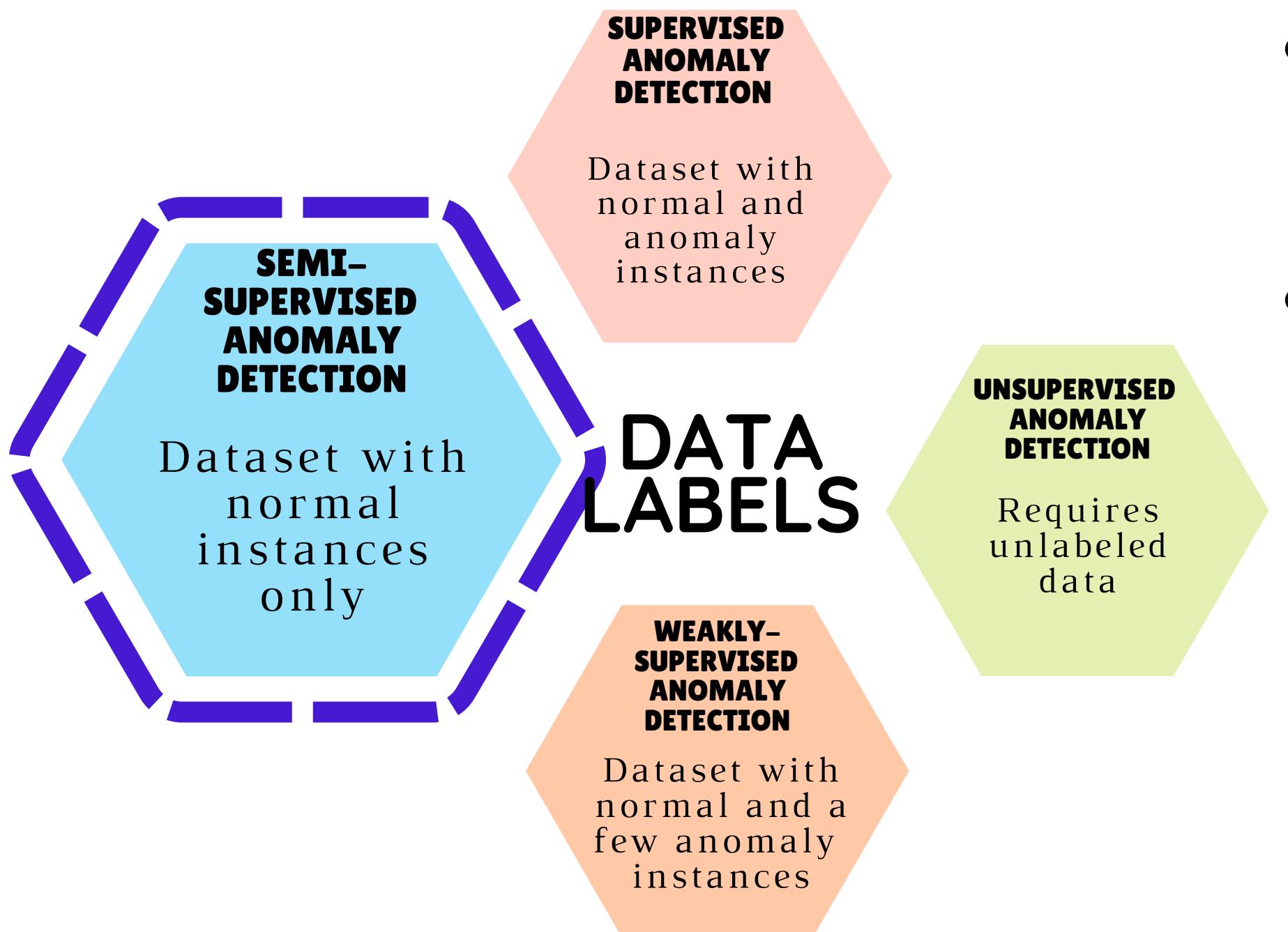
Requires
unlabeled
data



Supervised Anomaly Detection

- Requires balanced dataset
- Requires large dataset

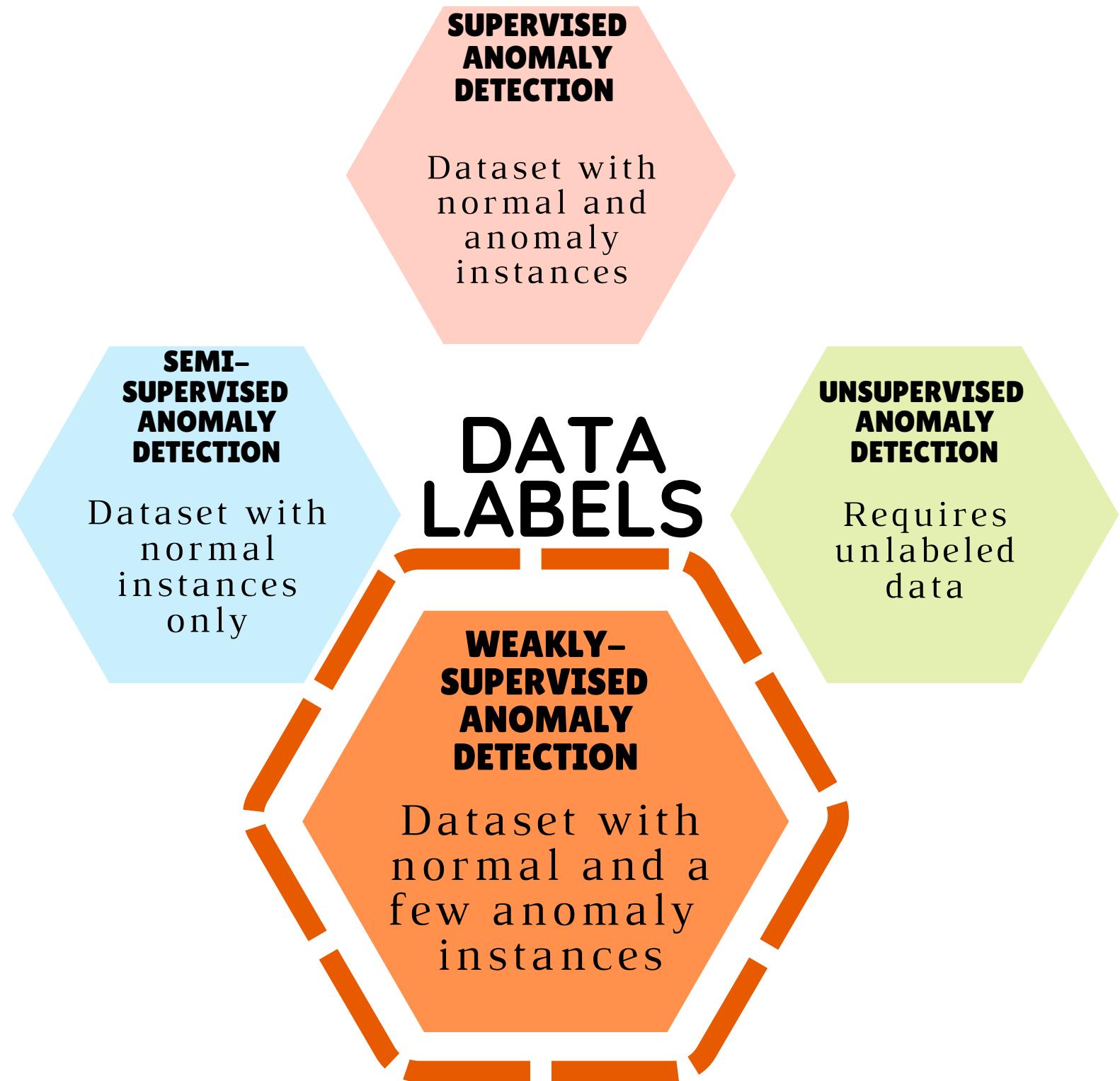
Semi-Supervised Anomaly Detection



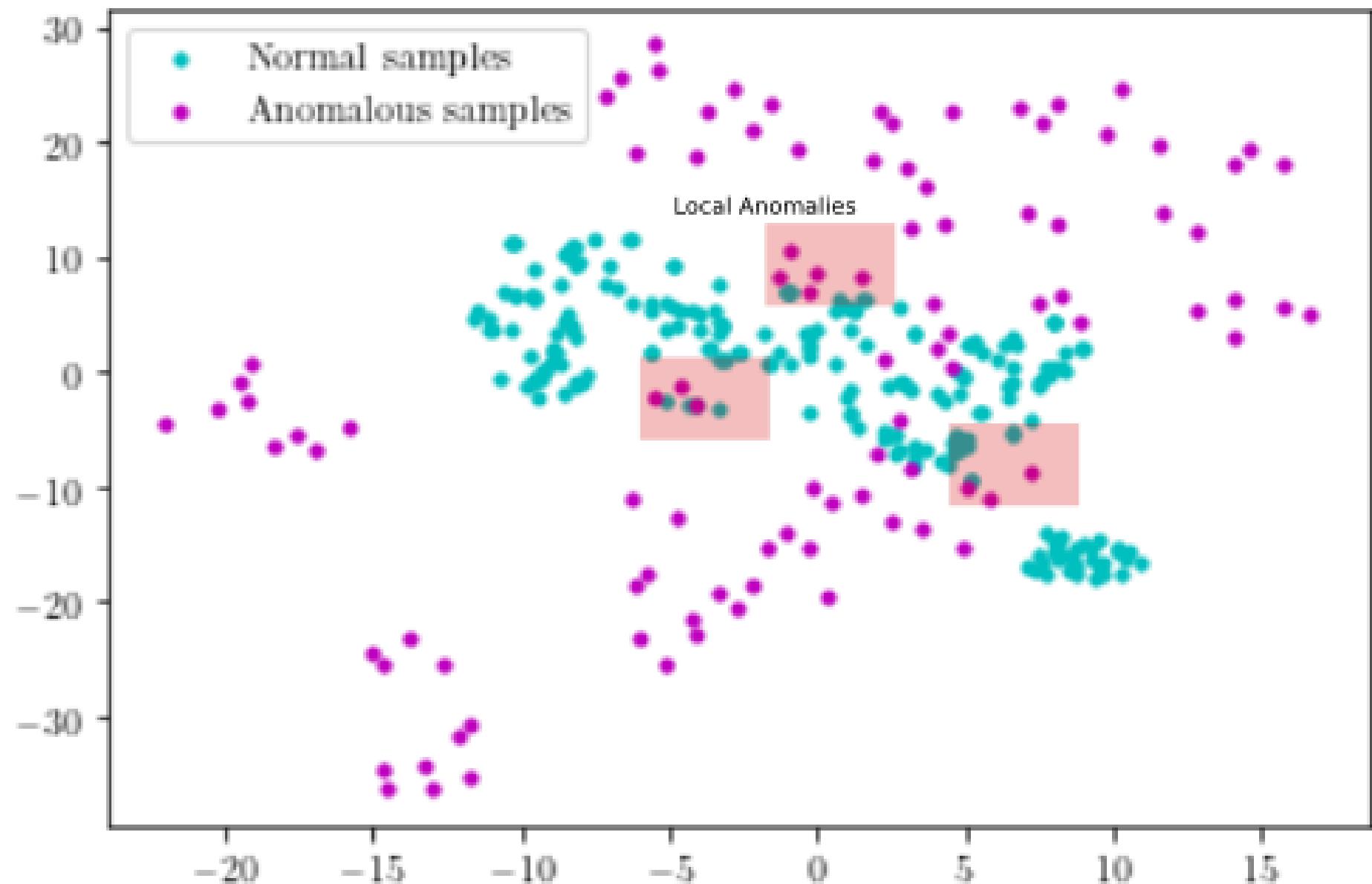
- Train data contains normal instances only
- Preferred training method



Weakly-Supervised Anomaly Detection

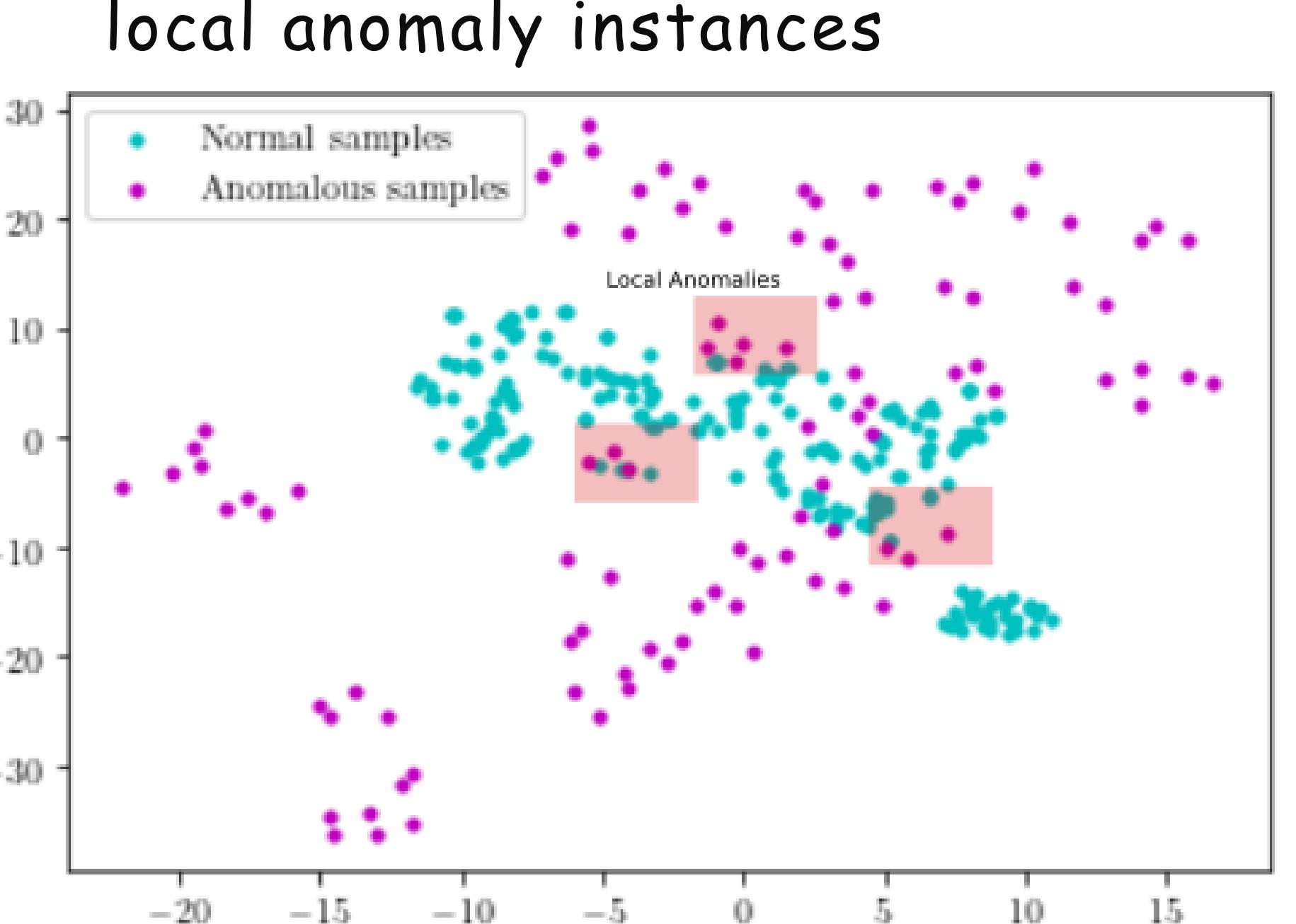
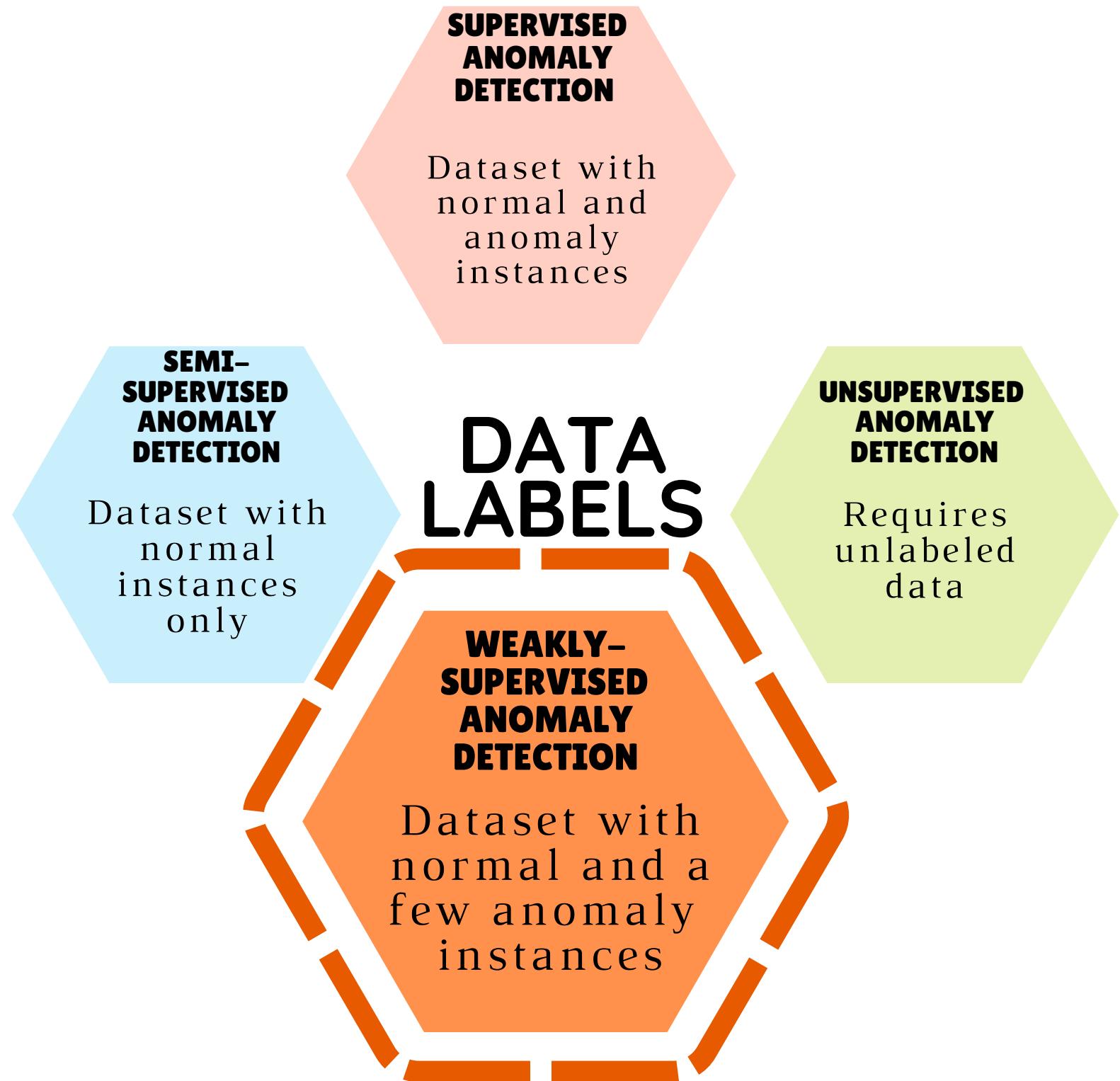


- Train data has normal instances and local anomaly instances

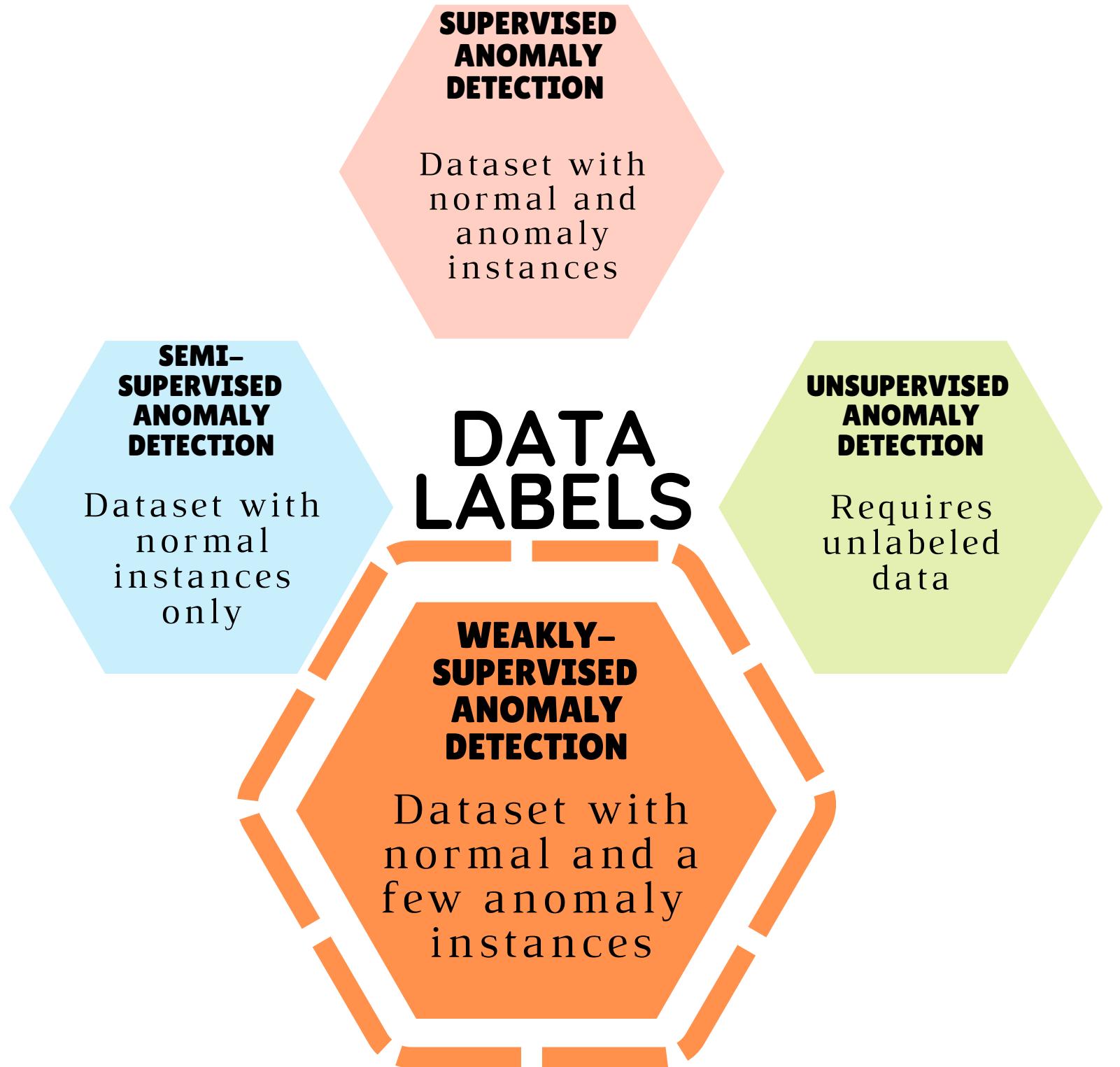




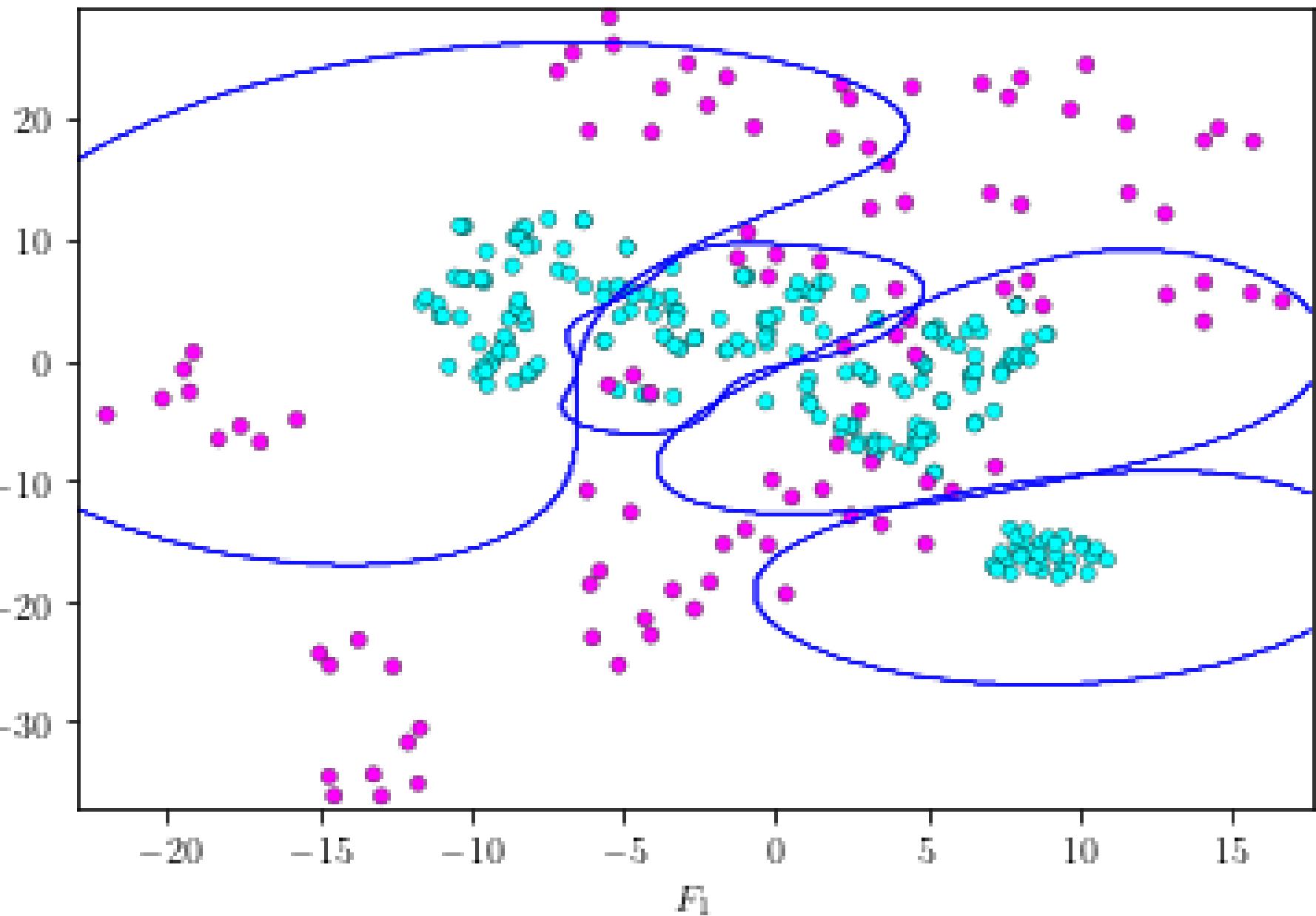
Weakly-Supervised Anomaly Detection



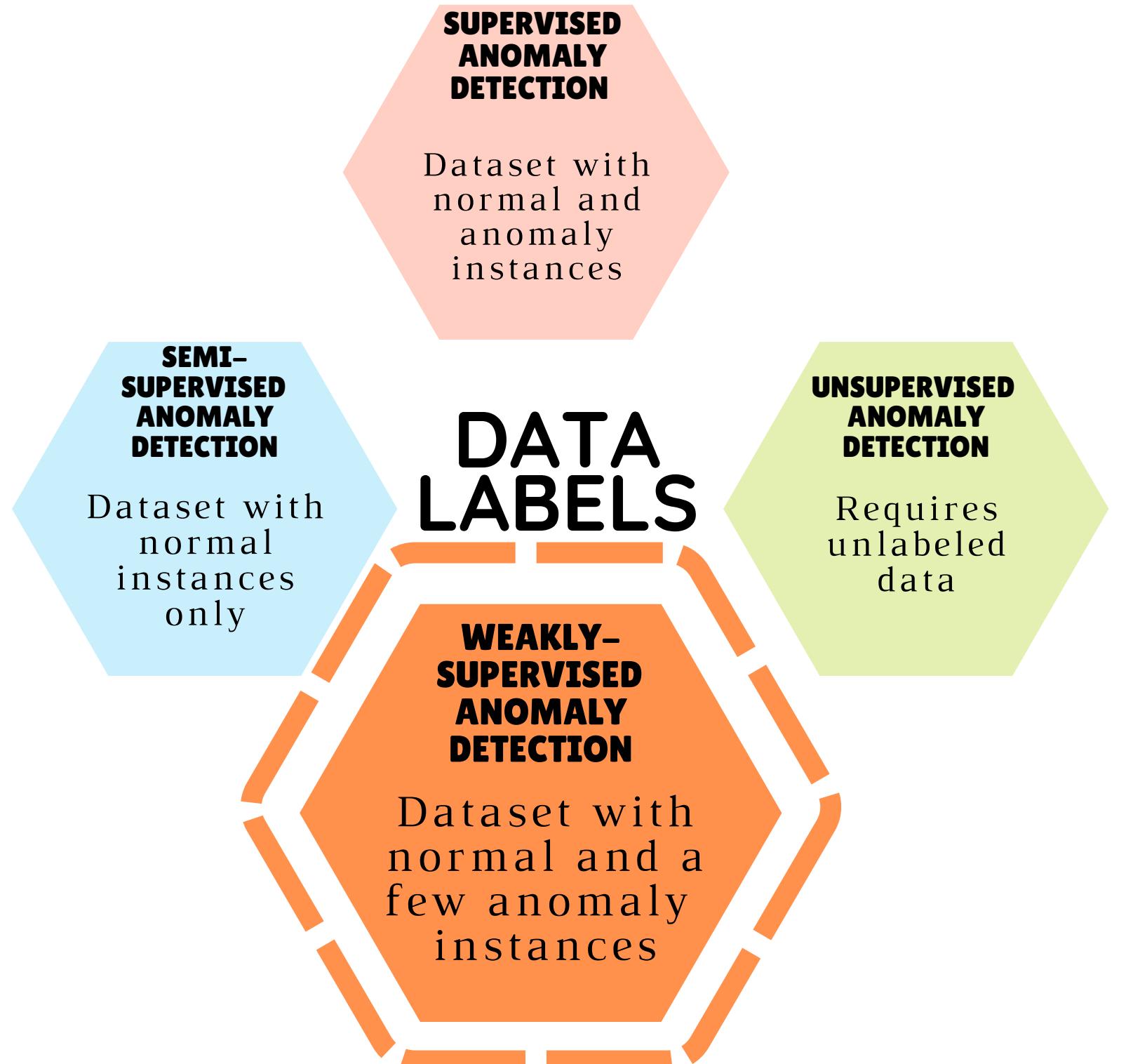
Weakly-Supervised Anomaly Detection



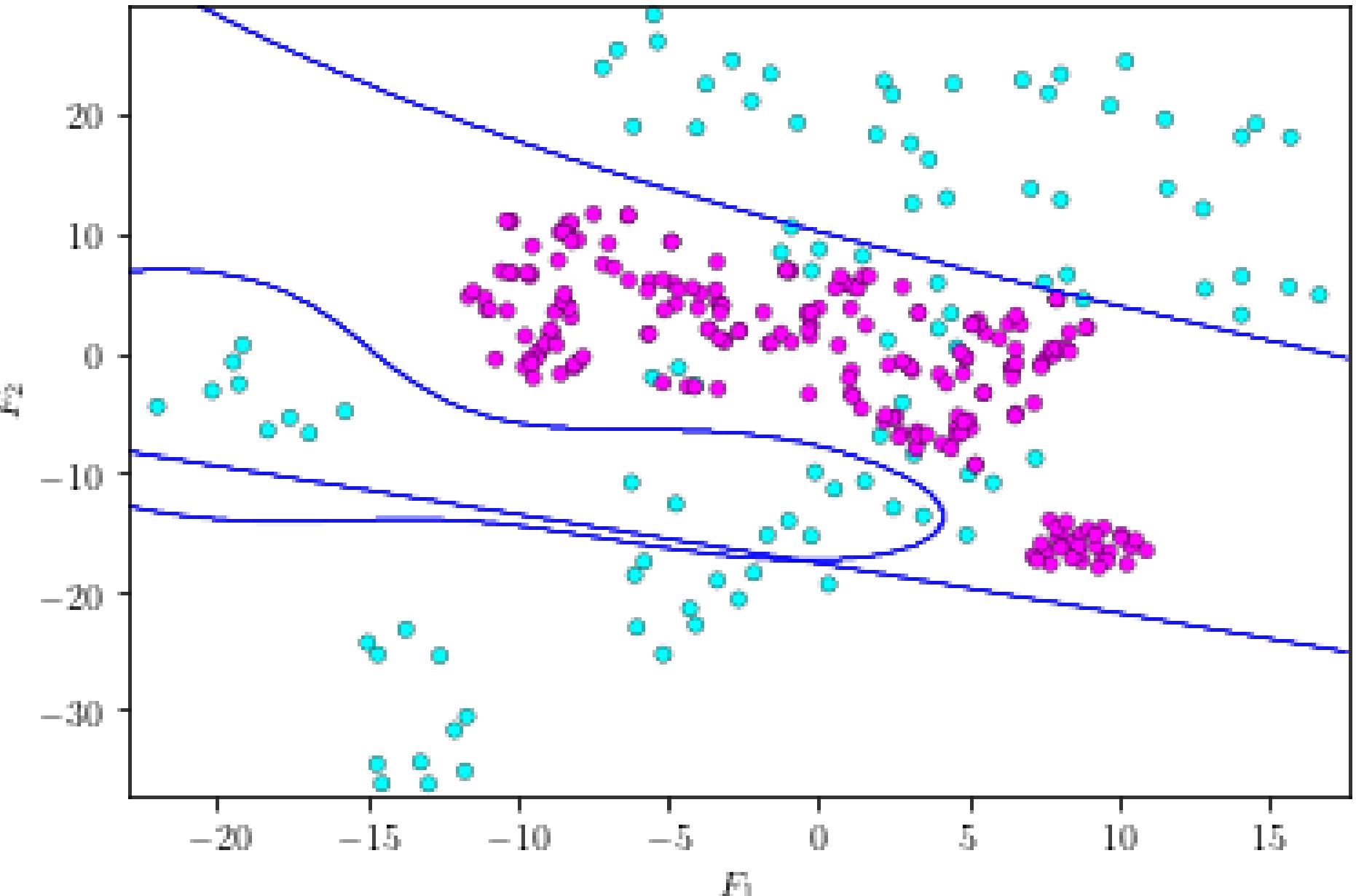
- Train data with normal instances only



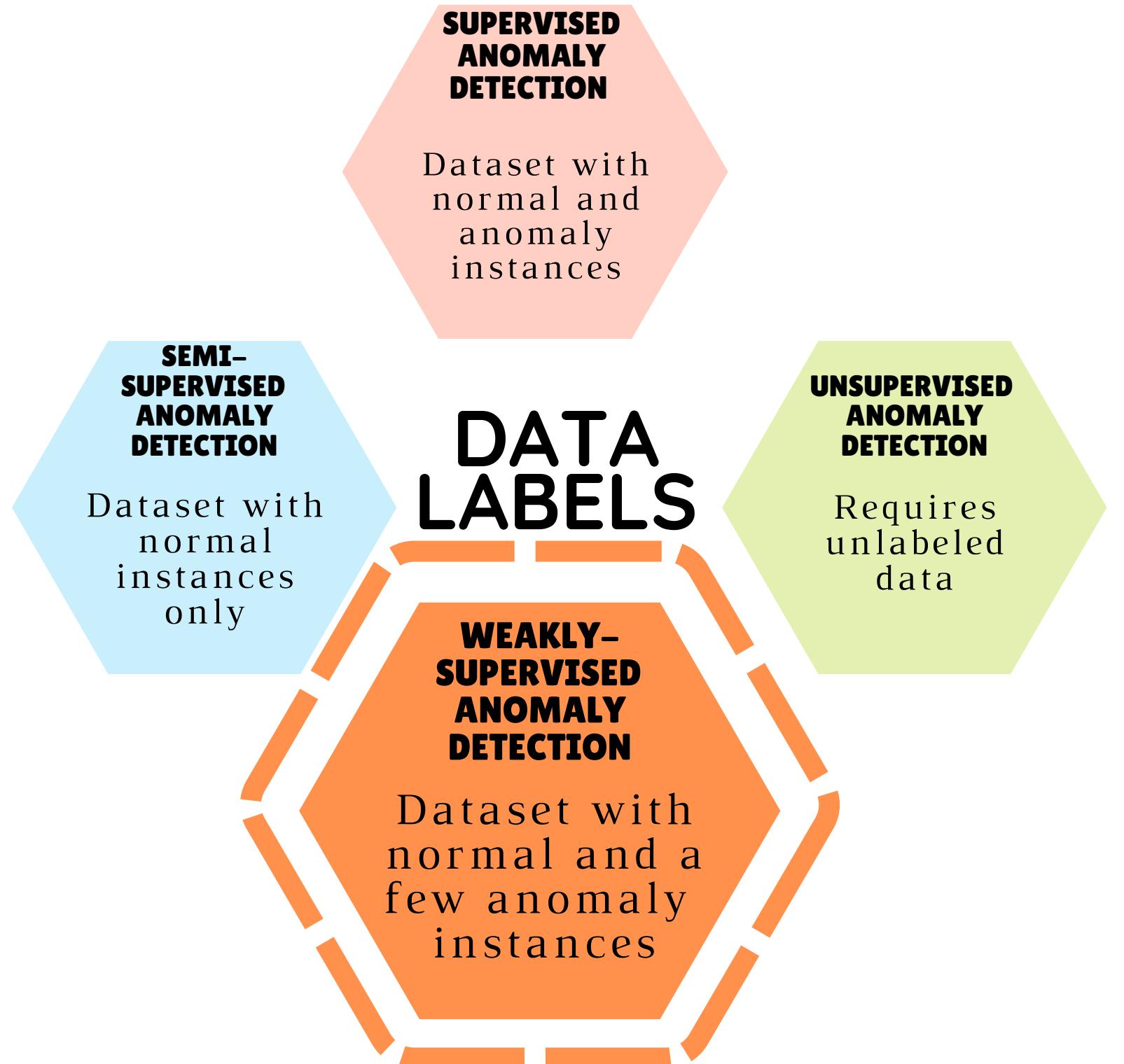
Weakly-Supervised Anomaly Detection



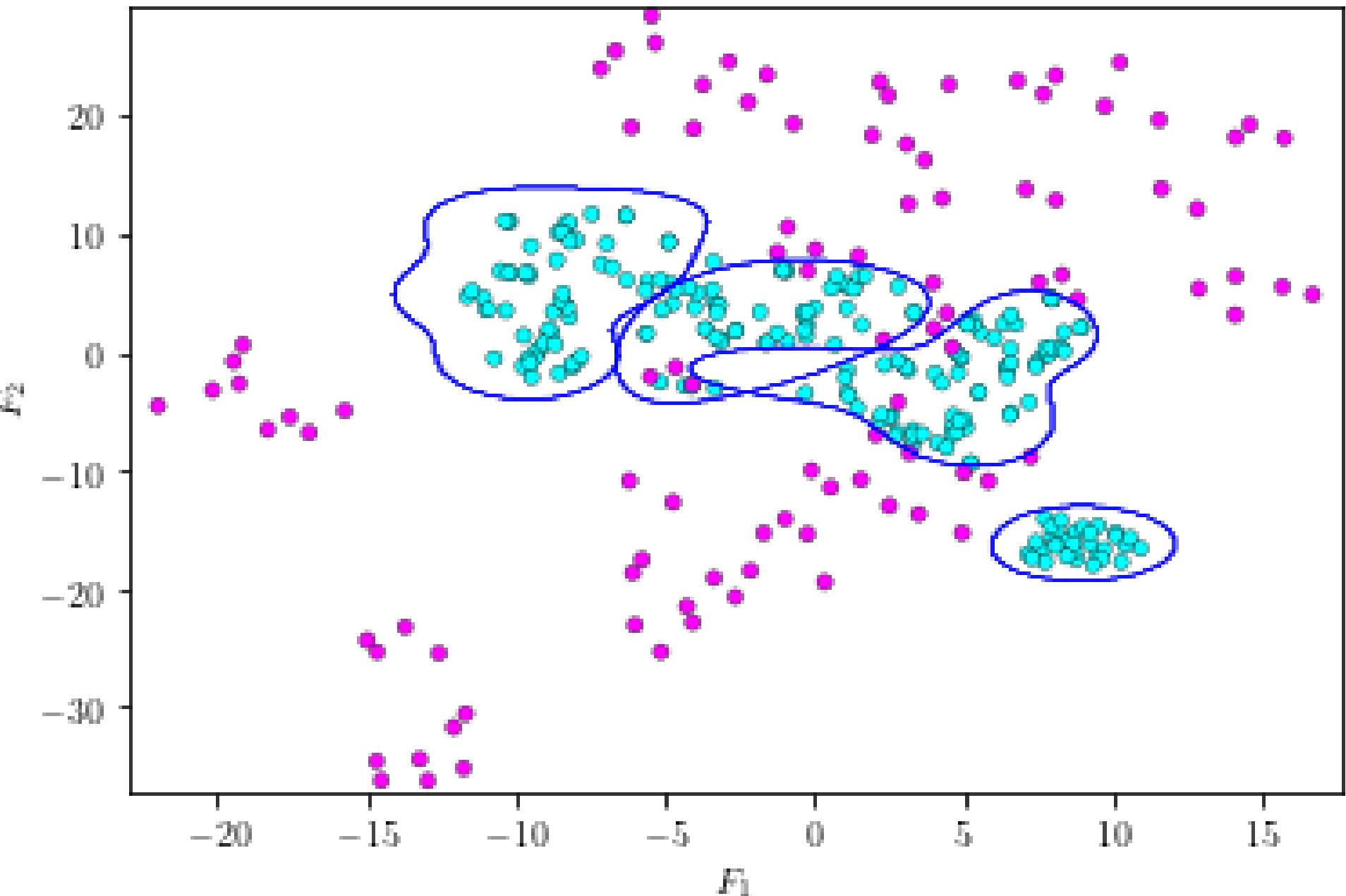
- Train data with anomaly instances only



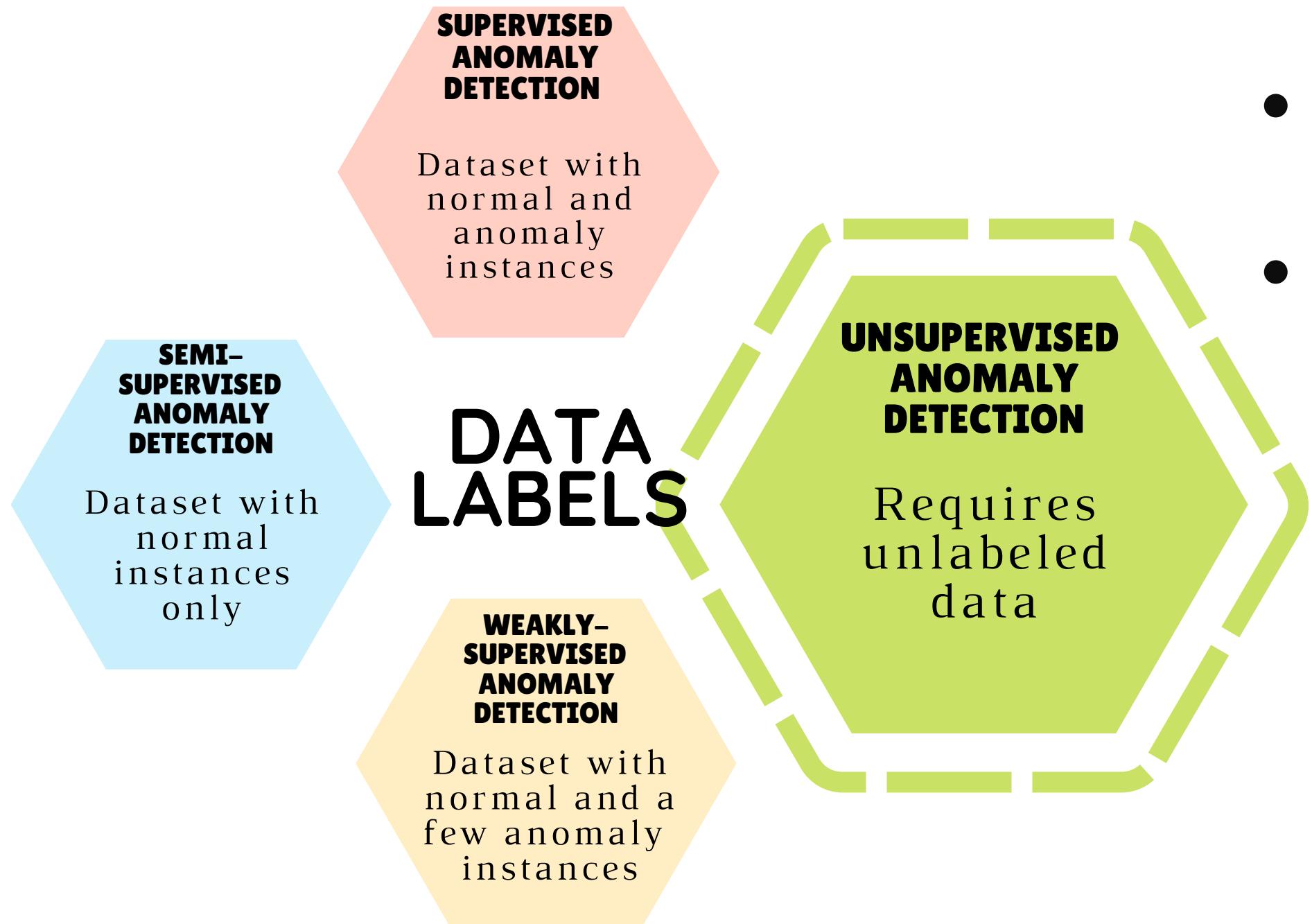
Weakly-Supervised Anomaly Detection



- train data has normal instances and local anomaly instances

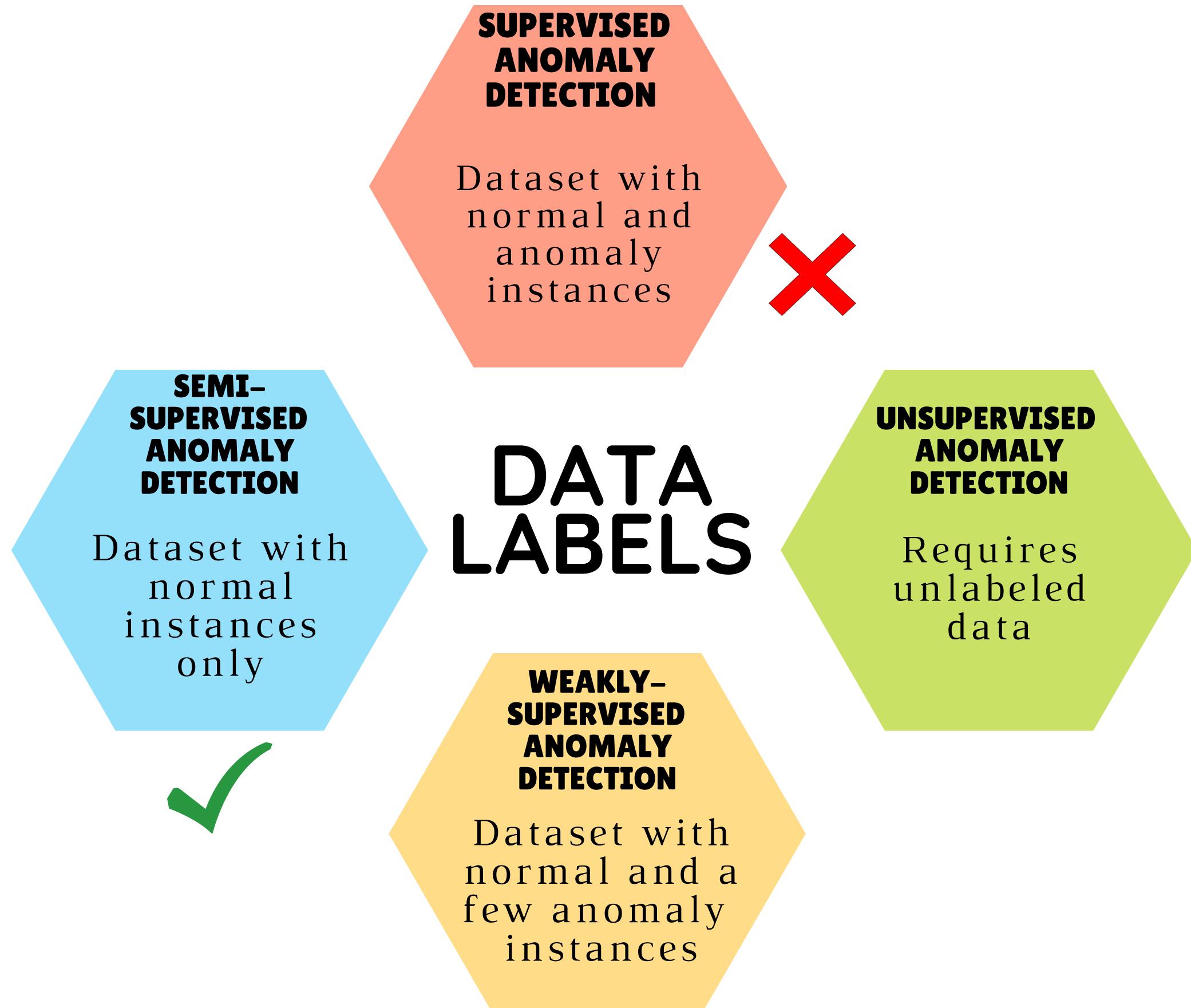


Unsupervised Anomaly Detection



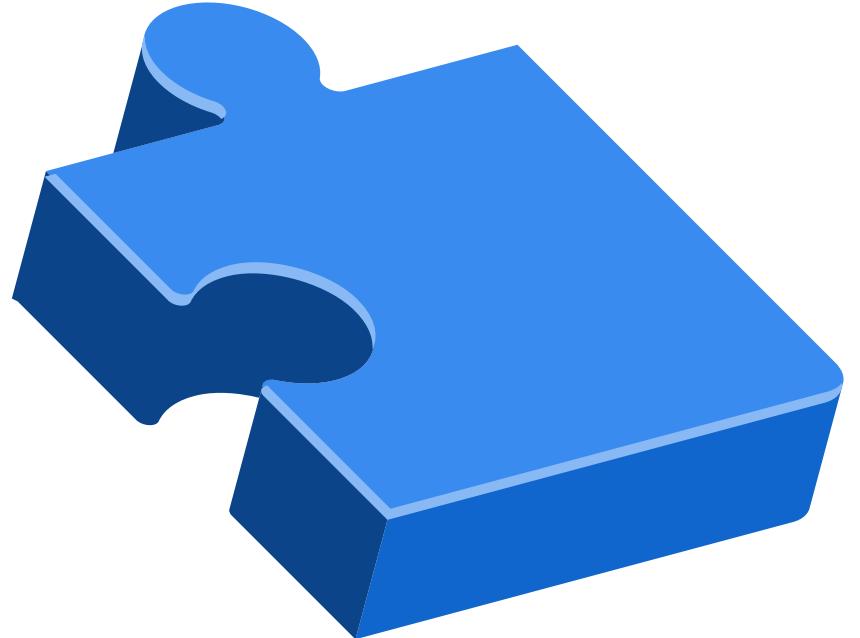
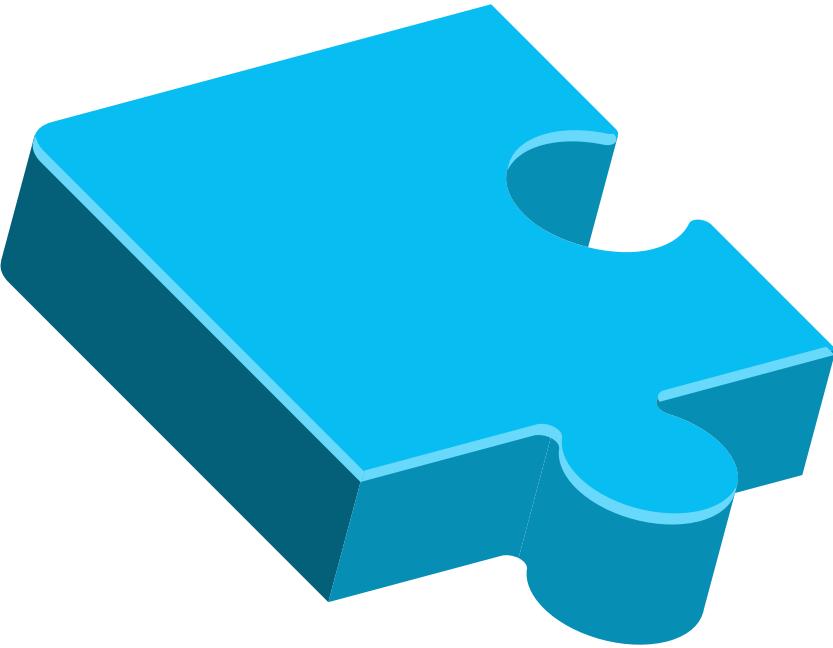
- No dependency of labeled data
- Requires prior knowledge of data distribution
 - $n(\text{Normal}) \gg n(\text{Anomalies})$
 - Anomalies are outliers

Training methods

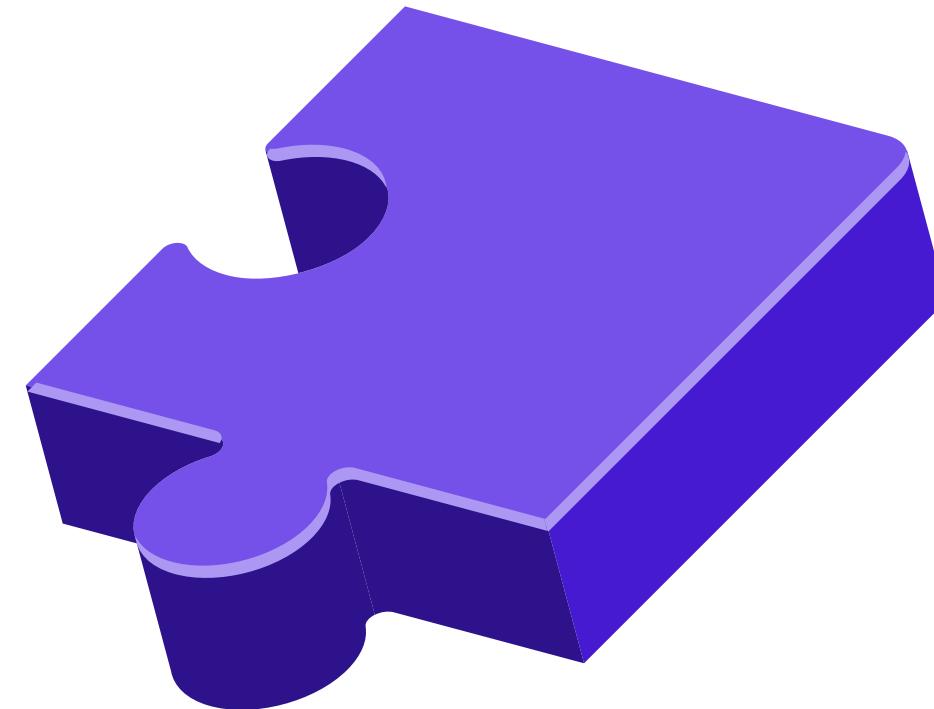


Anomaly Detection: Problem Complexities

Anonymous



Rare

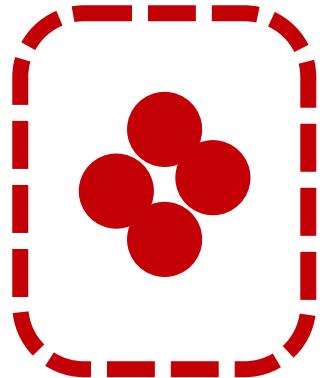


Heterogeneous

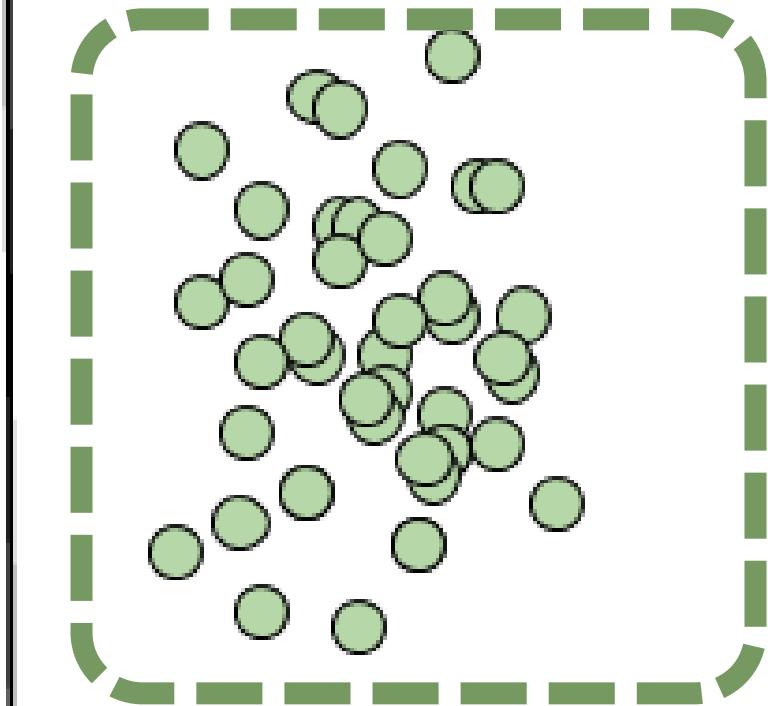
Crowd outbreaks



ANOMALY



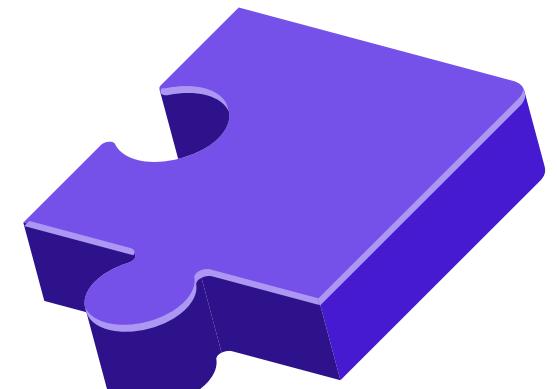
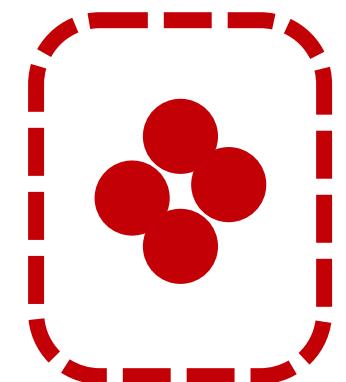
NORMAL



Gun attacks



ANOMALY



Heterogeneous



Rareness

Anonymous

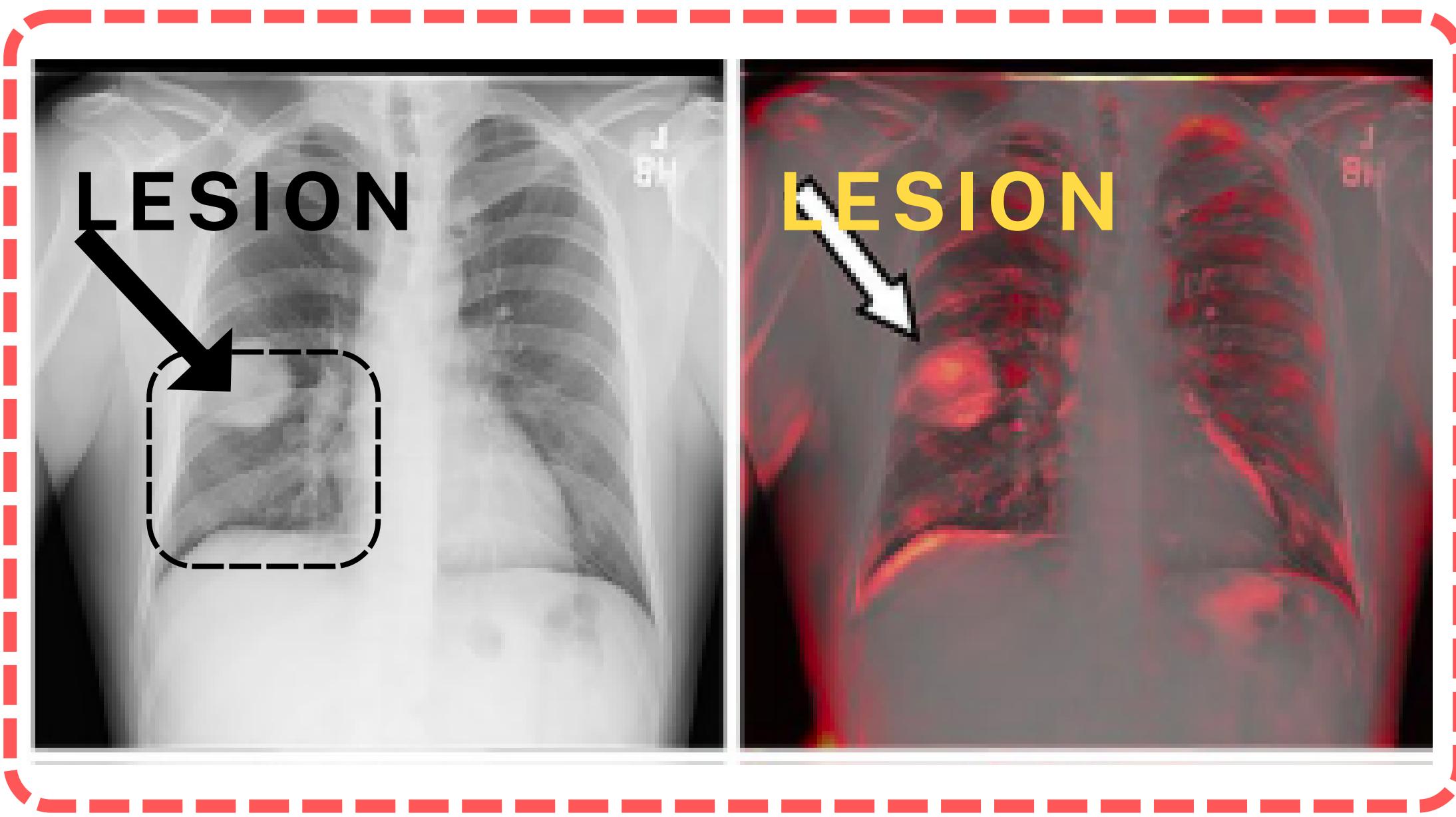
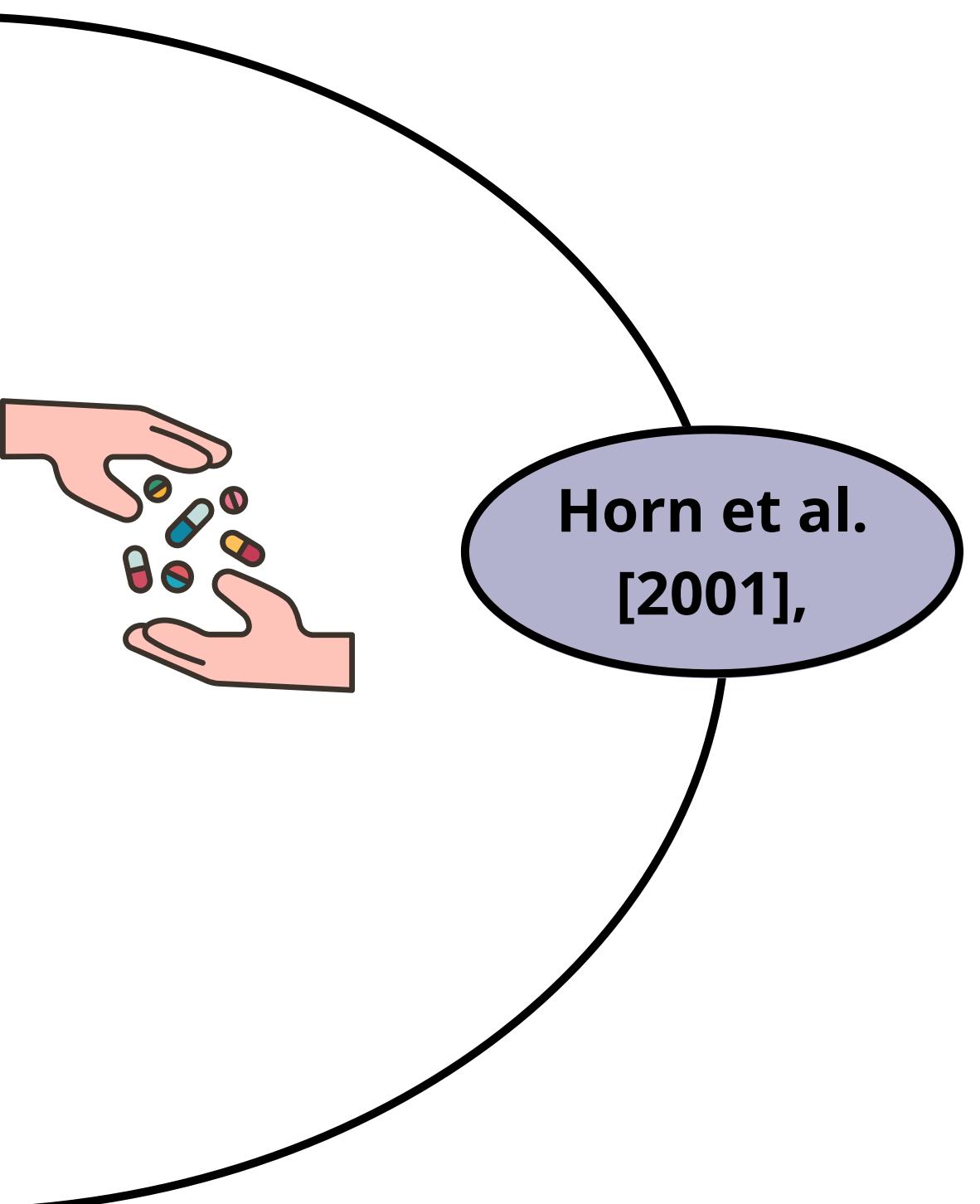


"HAPPY FAMILIES ARE ALL ALIKE;
EVERY UNHAPPY FAMILY IS UNHAPPY
IN ITS OWN WAY"
—ANNA KARENINA

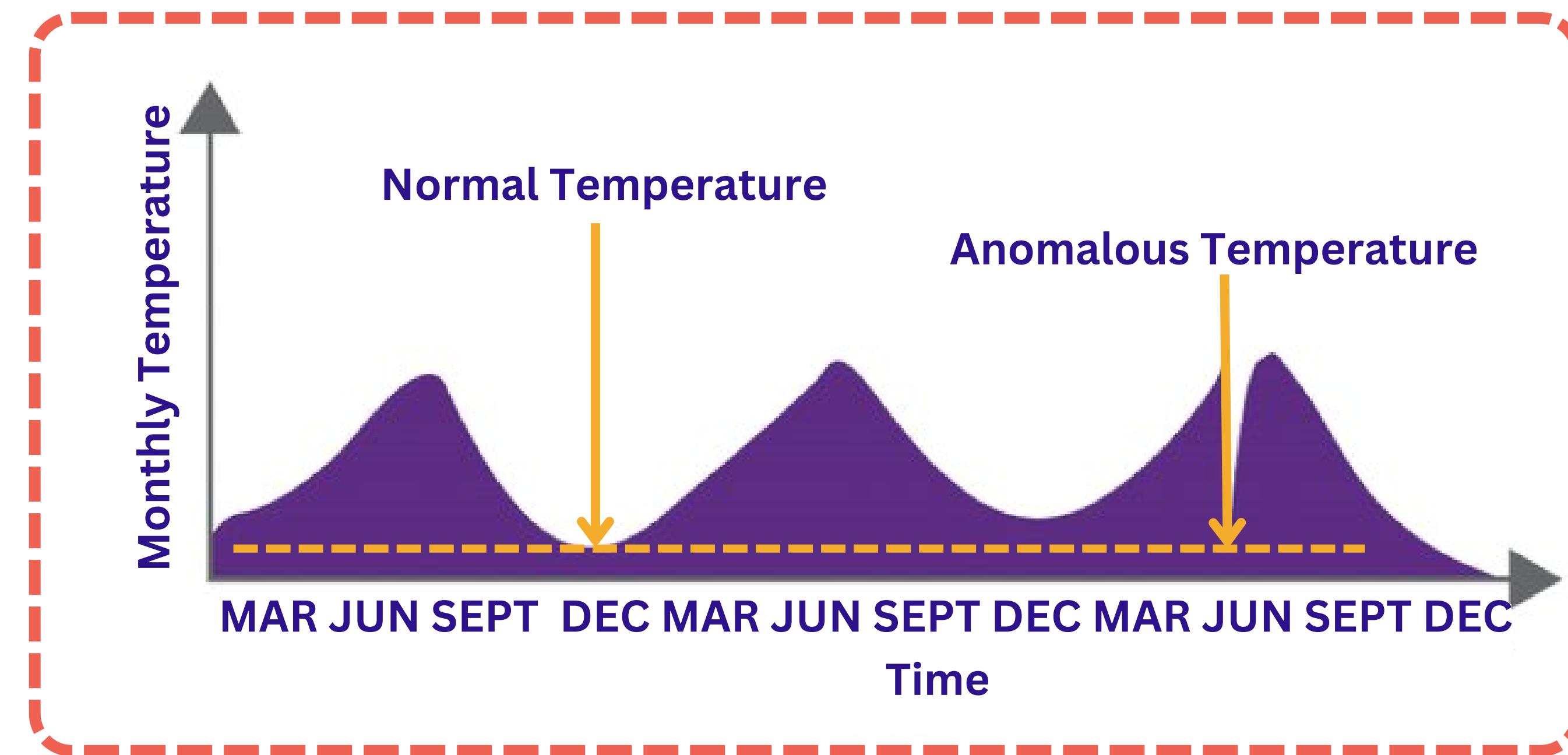


Copyright: Canva

Anomaly Types: Point Anomalies

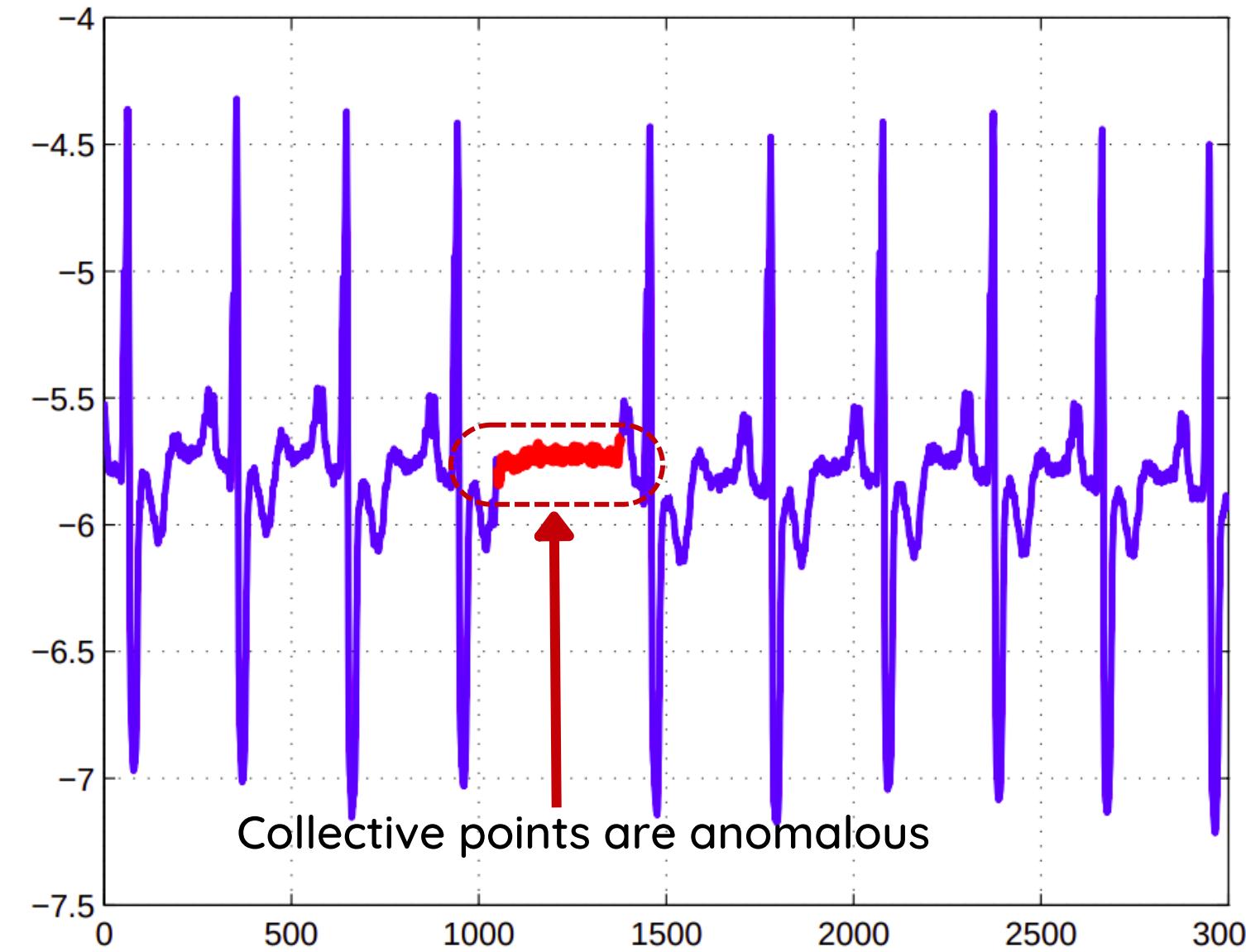
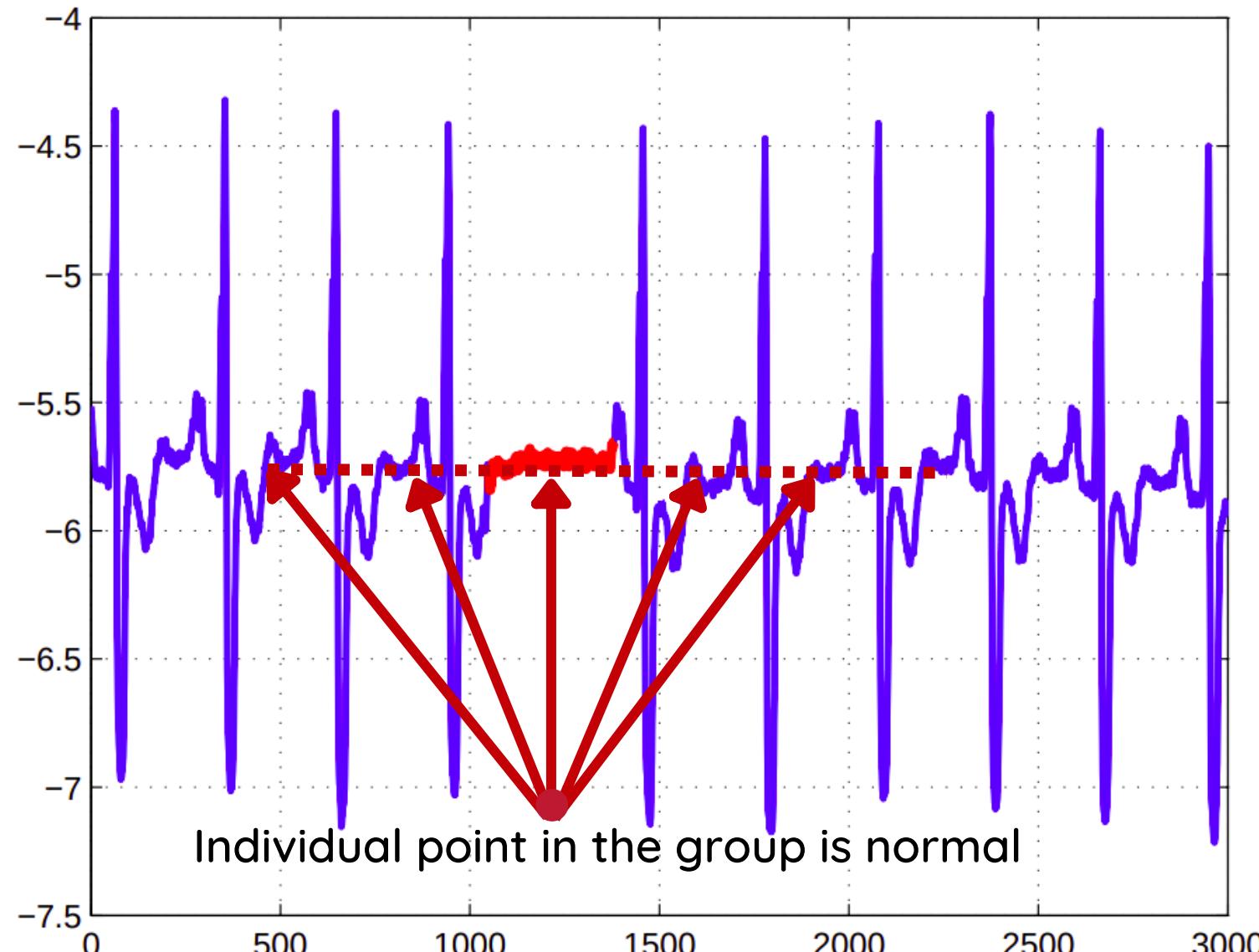


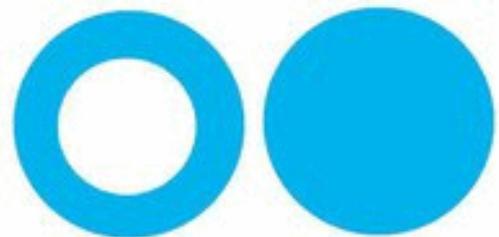
Anomaly Type: Conditional Anomalies



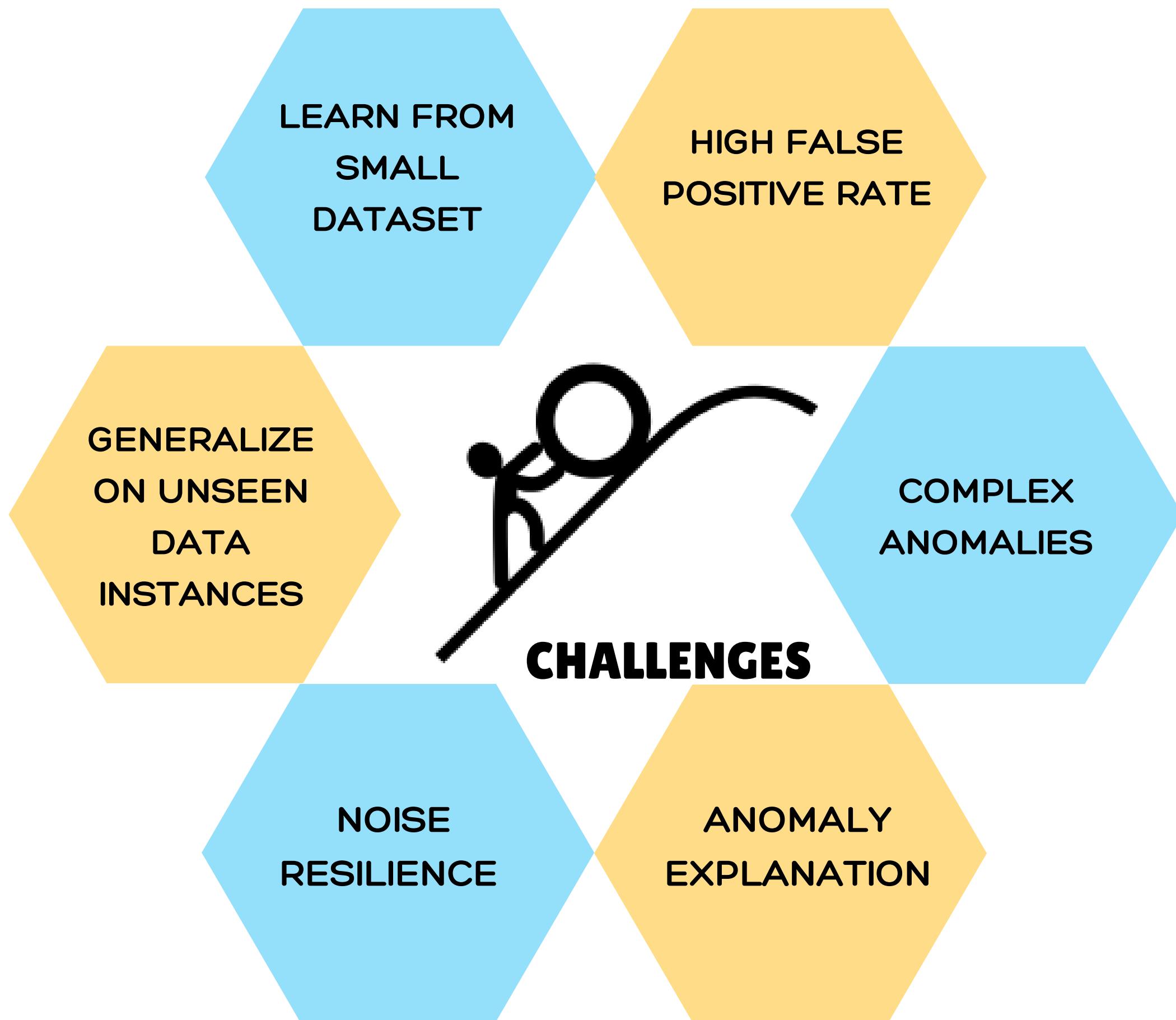
Anomaly Type: Collective Anomalies

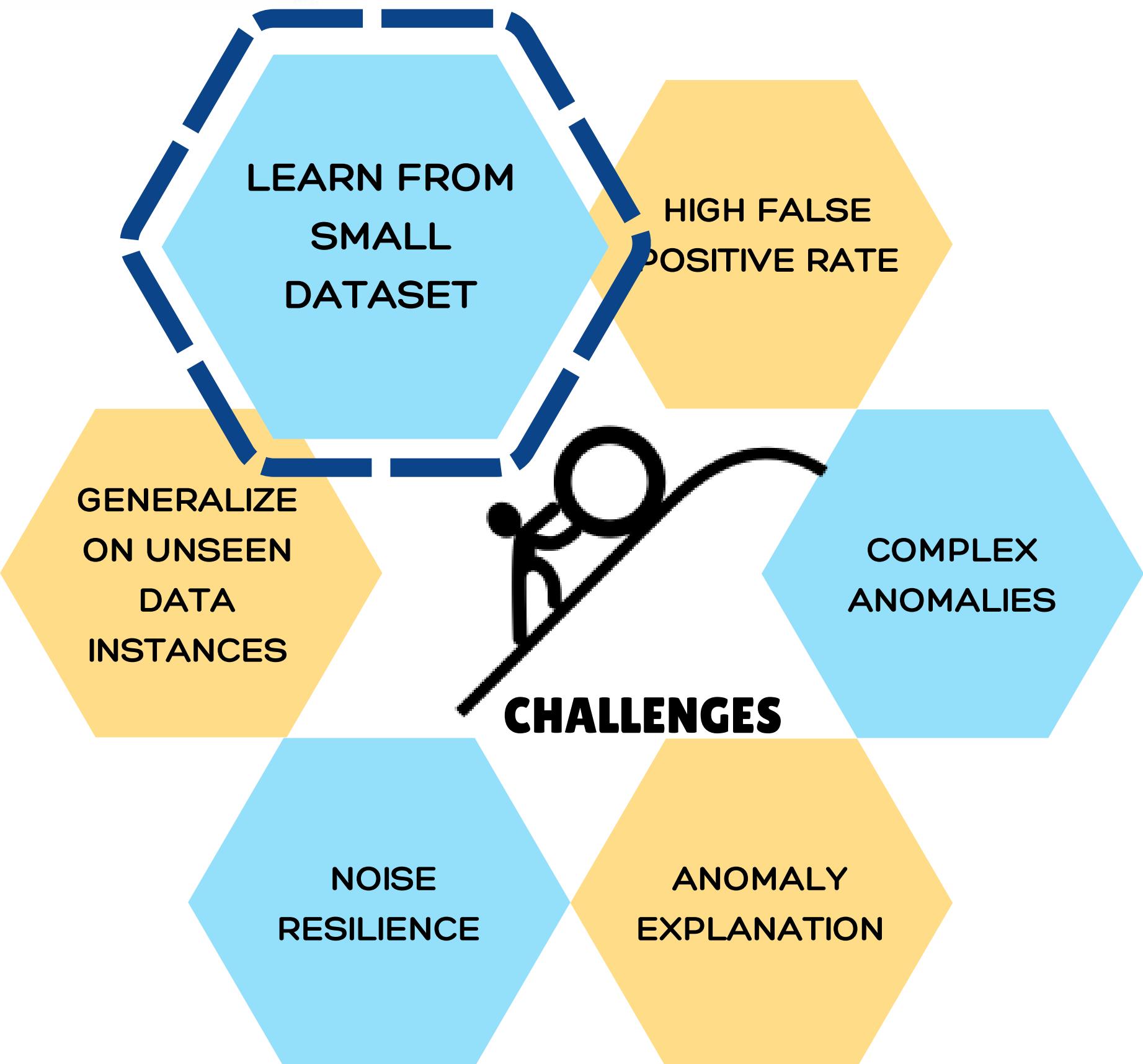
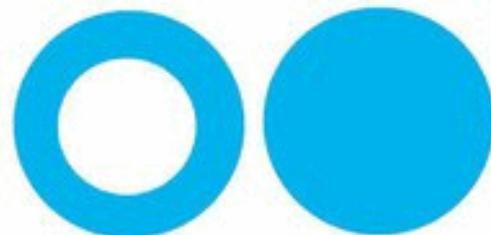
HUMAN ELECTROCARDIOGRAM OUTPUT





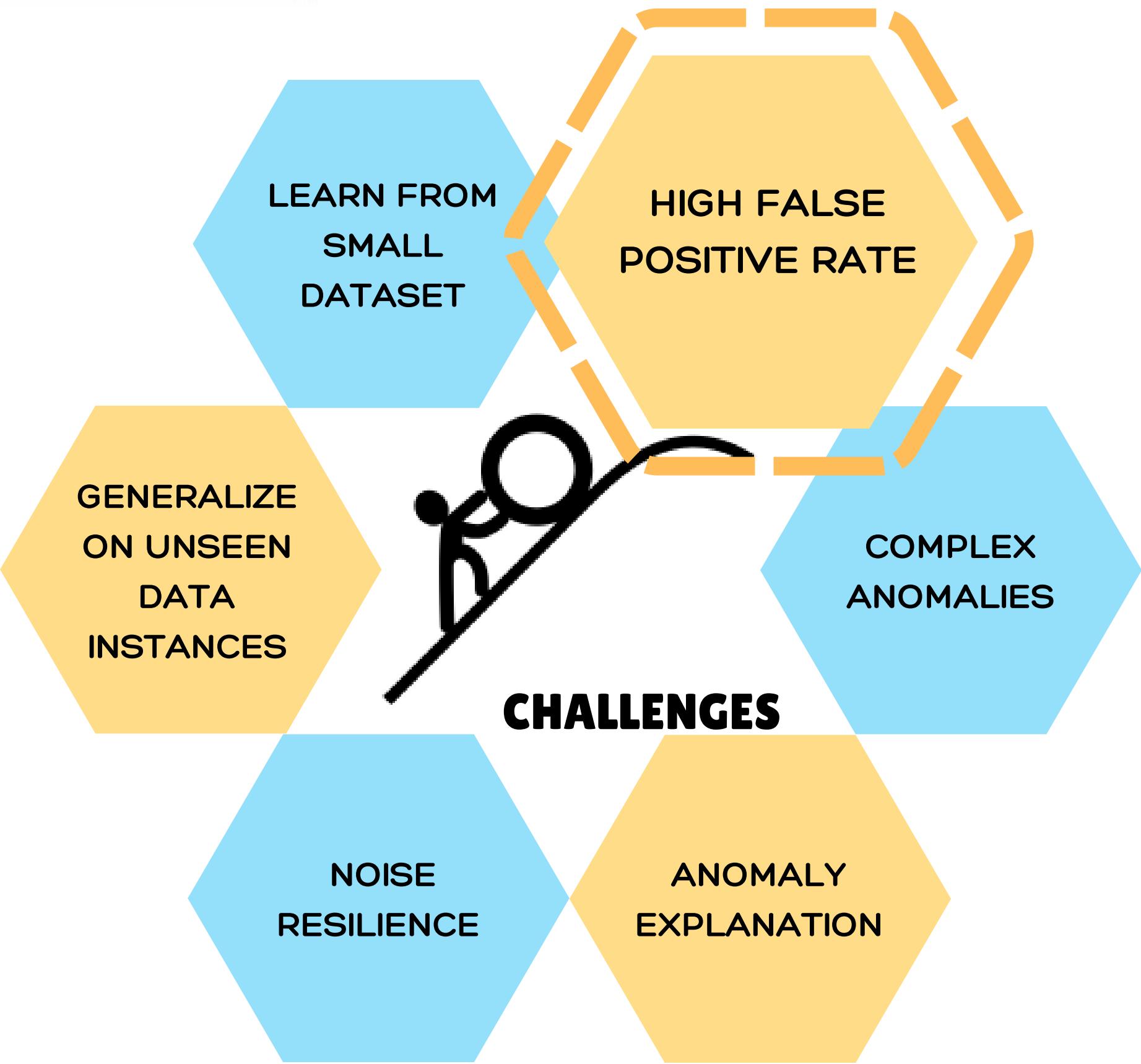
Challenges Addressed by Deep Learning





CH1: Sample-Efficient Learning

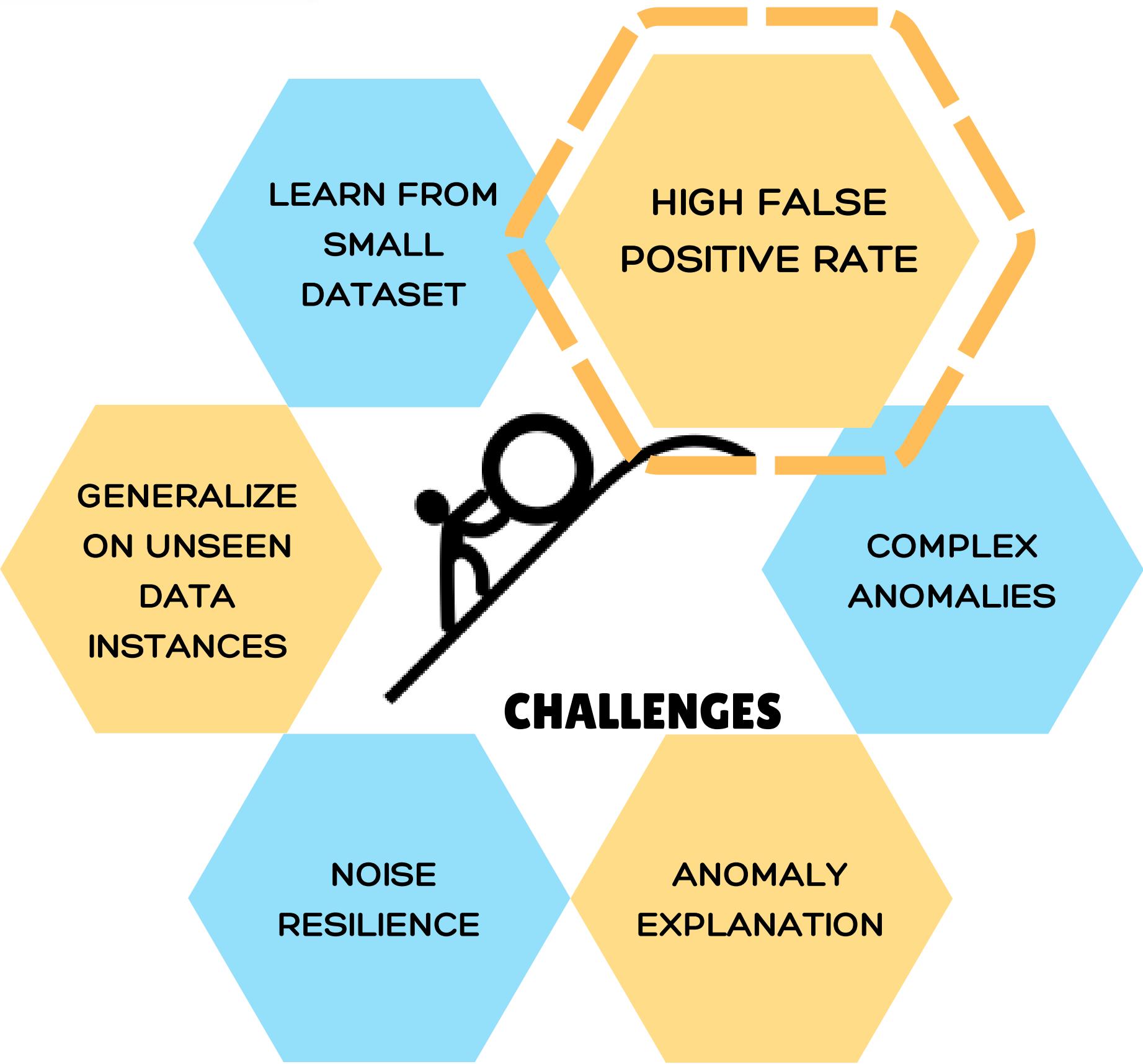
- No dependency on large dataset
- Labeling is expensive



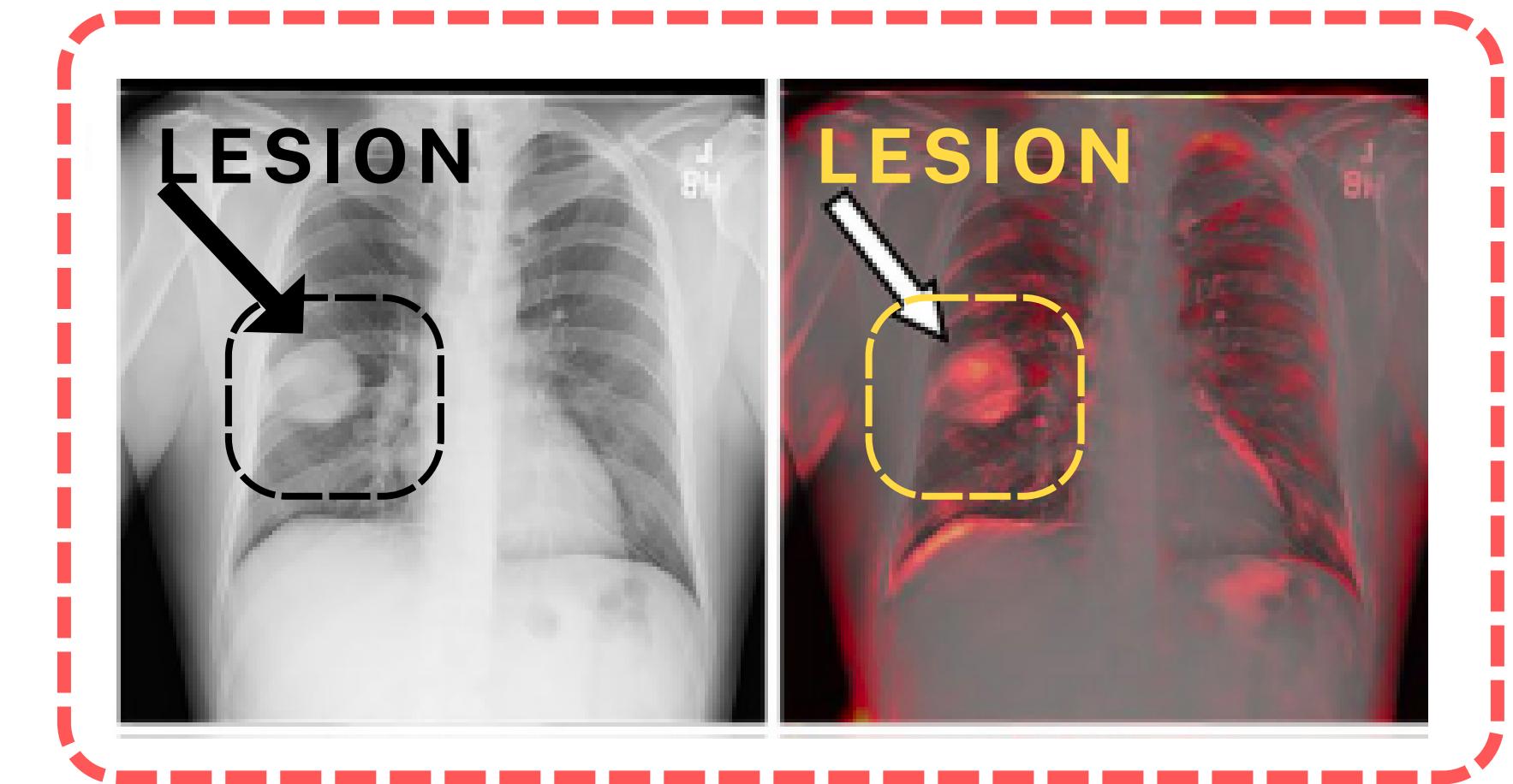
CH2: High False Positive Rate

- Anomalies are
 - Rare
 - Heterogeneous
- Anomalies are not correctly reported

CH1: High False Positive Rate

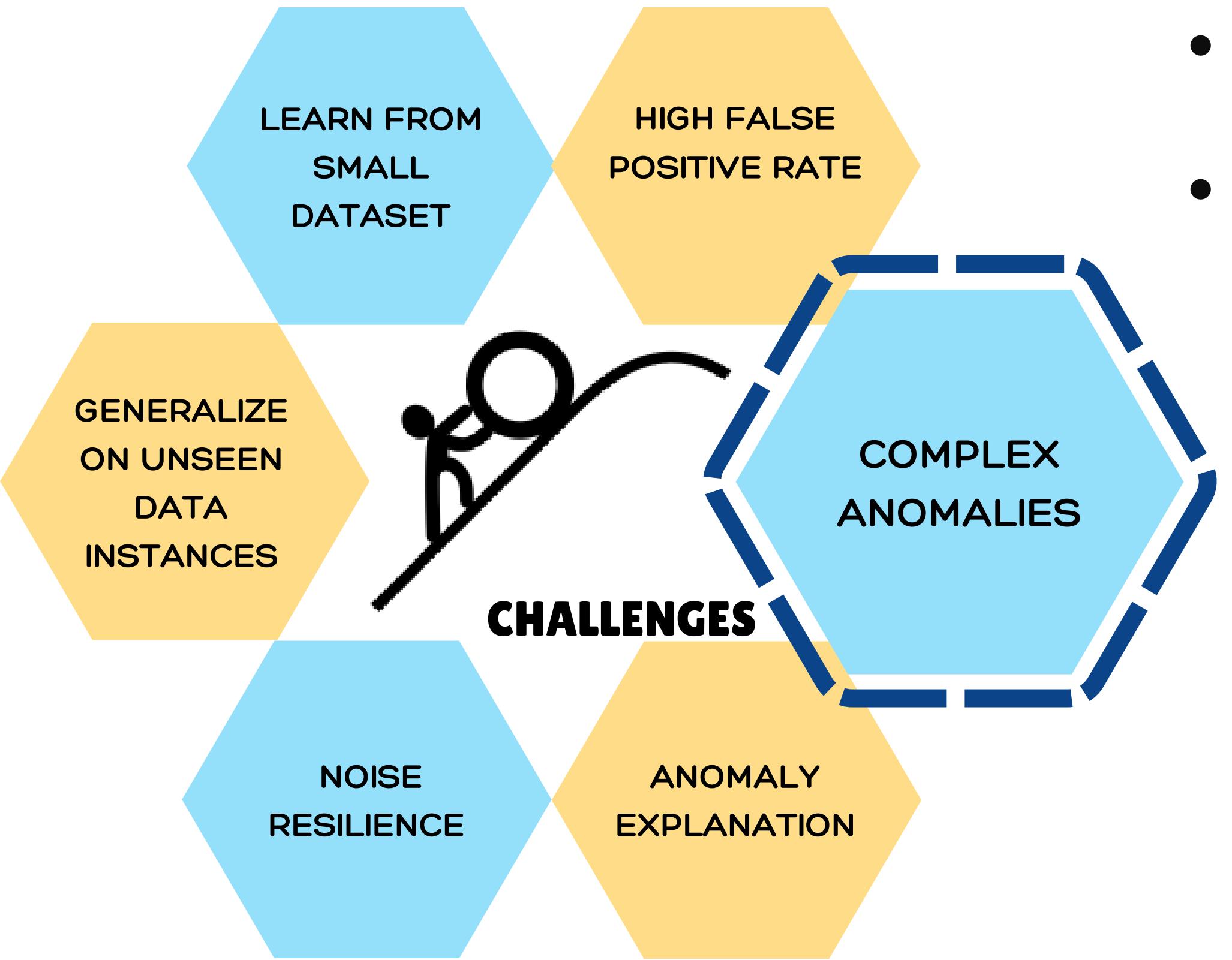


- Anomalies are not correctly reported

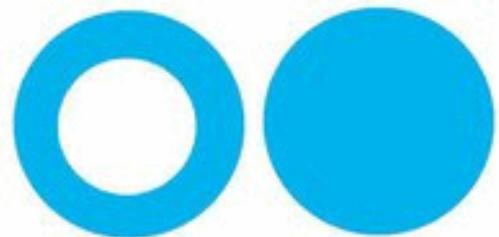


Copyright: Horn et al. [2001]

CH1: Complex Anomalies



- Requires multiple clues
- Varied anomaly types
 - Point
 - Contextual
 - Group



Methods for Comparative Study

Thoduka et al.
[2021]

Proposed: Action Execution Monitoring

- Hybrid of Model-based and Model free method
- Detect motion anomalies related to book placement

Architecture: Probabilistic U-Net

- Conditional variational autoencoder + U-Net
- Relatively easier to train with less training parameters



Key-points:

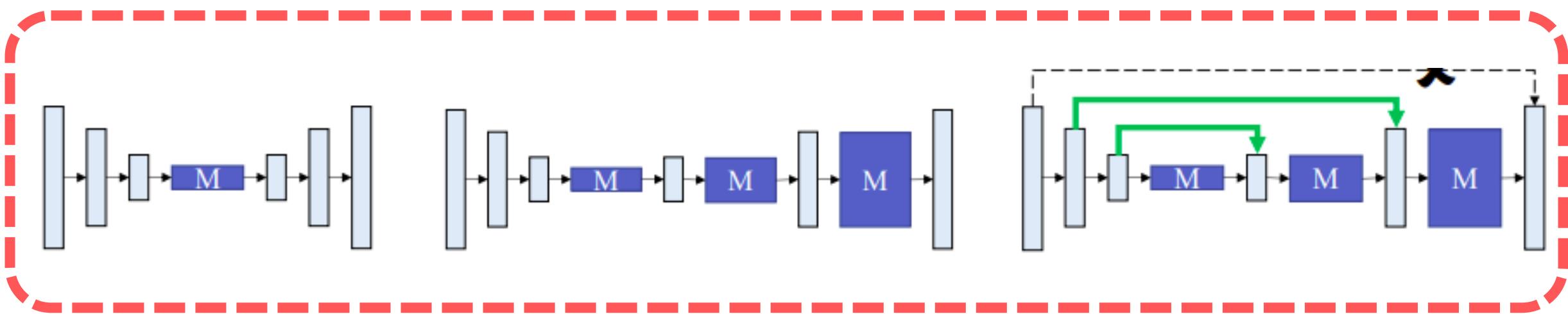
- No dependency on large dataset
- Used domain knowledge

Key-points:

- Exhibited superior against two other methods
- Contributed a new robotic dataset

Methods for Comparative Study

Liu et al.
[2021]



Proposed: Video Anomaly Detection

- Reconstruction method + future frame prediction method
- Proposed multi level memory augmentation

Architecture:

- Autoencoder for reconstruction
- Conditional variational autoencoder for future frame prediction

Key-points:

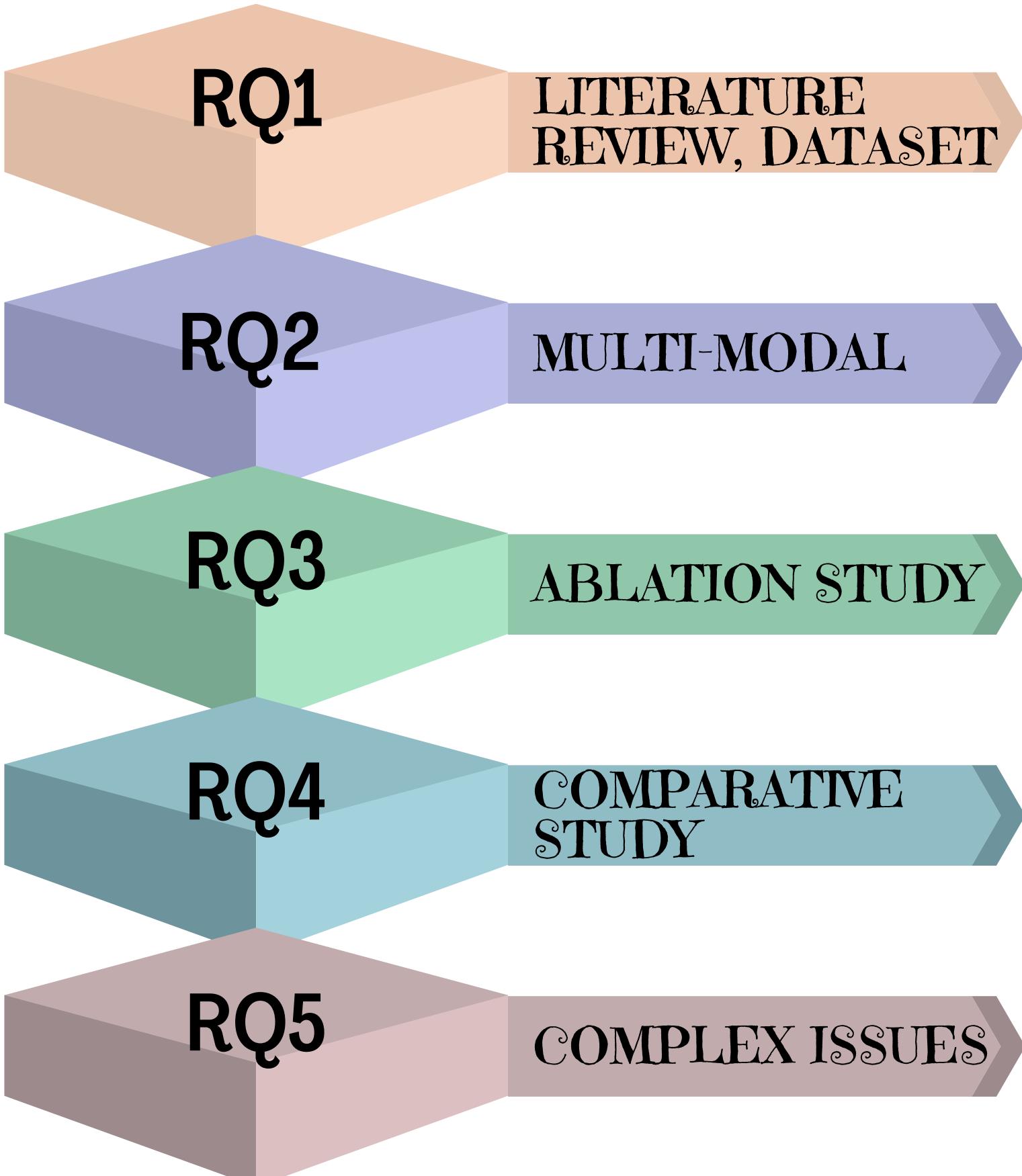
- No generalization for Anomaly instances

Key-points:

- Exhibited superior against 8 other methods
- Superior to reconstruction only and prediction only

Problem Formulation

Research Questions (RQ1-RQ5)



- What are deficits?
- Why do we need a new dataset?
- What to fuse?
- When to fuse?
- How to fuse?
- How do individual sensors help detect anomalies?
- How do the chosen methods perform against the proposed dataset?
- What are the complex issues addressed by methods?



Hochschule
Bonn-Rhein-Sieg

Thank you