

**iihe**  
BRUXELLES BRUSSEL



**ULB**



# Jet-based anomaly detection at the LHC

**Santiago Paredes Saenz**

On behalf of the ATLAS and CMS collaborations

[santiago.paredes@cern.ch](mailto:santiago.paredes@cern.ch)

**ISMD 2022**

August 2022



# Overview

---

- Anomaly detection at the LHC
- Some recent methods using jets
- Ongoing efforts and recent results
- Summary

# Anomaly detection @ LHC

what? why?

---



# Anomaly detection @ LHC

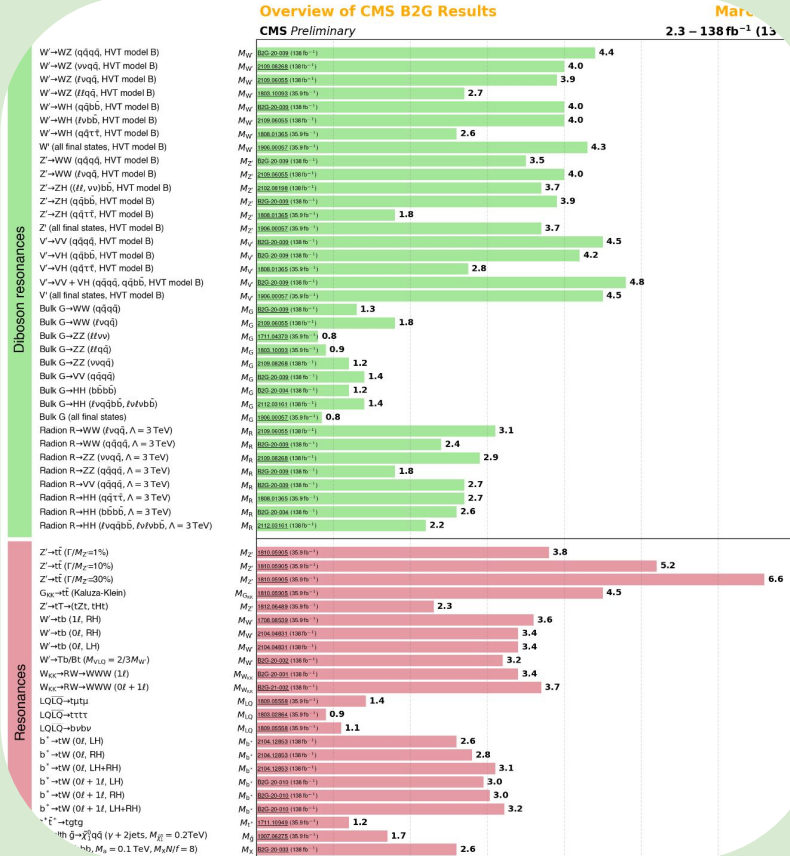
- What we don't understand:
  - ⇒ Dark matter    ⇒ Matter-Antimatter
  - ⇒ Neutrino mass    ⇒ [...]

# Anomaly detection @ LHC

- What we don't understand:

- Dark matter
- Matter-Antimatter
- Neutrino mass
- [...]

- Where we are looking:

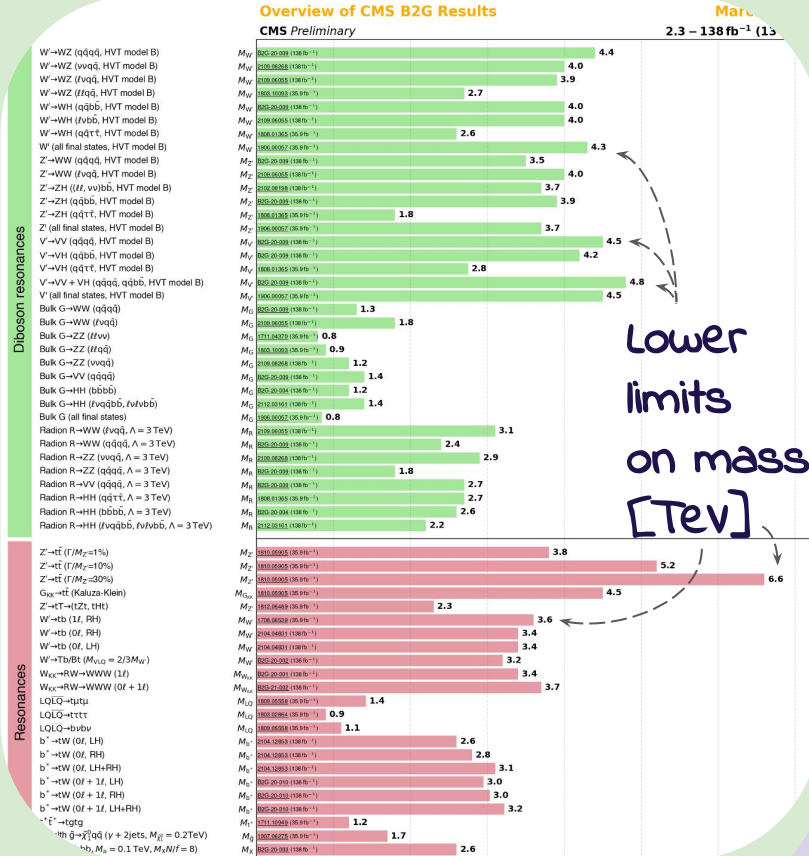


# Anomaly detection @ LHC

## • What we don't understand:

- Dark matter ➤➤ Matter-Antimatter
- Neutrino mass ➤➤ [...]

## • Where we are looking:



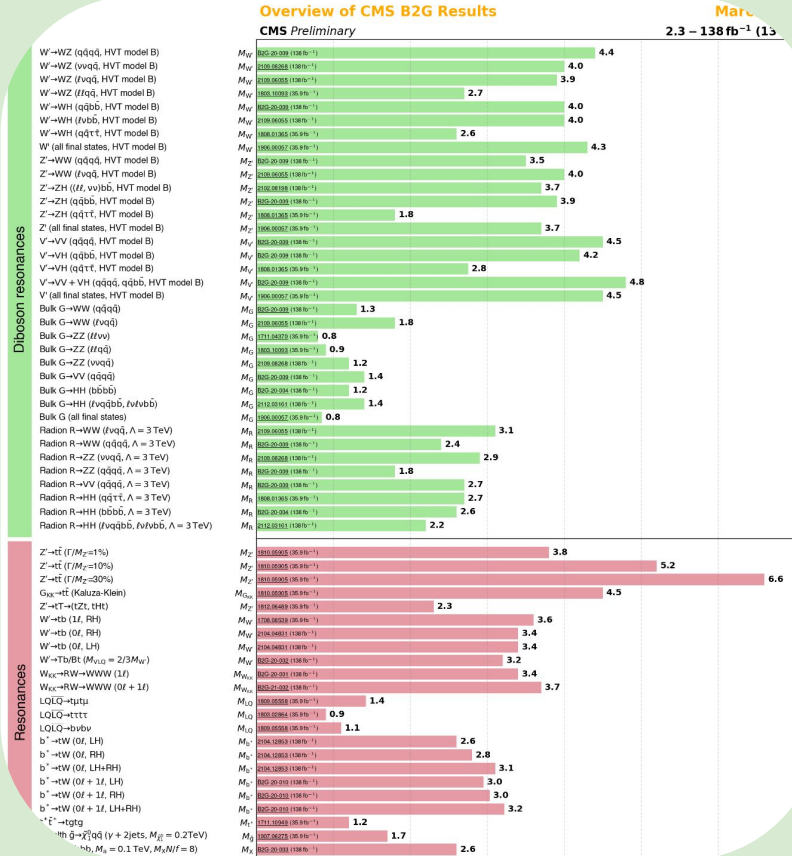
# Anomaly detection @ LHC

- What we don't understand:

- Dark matter
- Matter-Antimatter
- Neutrino mass
- [...]

○ Where the answer could be

- Where we are looking:



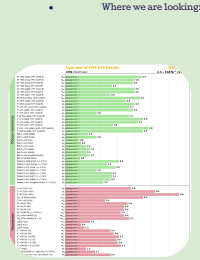




# Anomaly detection @ LHC

- What we don't understand:
  - ⇒ Dark matter    ⇒ Matter-Antimatter
  - ⇒ Neutrino mass    ⇒ [...]

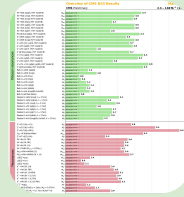
○ Where the answer could be



# Anomaly detection @ LHC

- What we don't understand:
  - Dark matter    ➤➤ Matter-Antimatter
  - Neutrino mass   ➤➤ [...]

○ Where the answer could be



[Model-independent searches]

# Anomaly detection @ LHC

---

- **what?** → A type of **model-independent** search
- **why?** → **Explore phase-space** missed by direct searches
- **How?** → Use the large dataset (+ some\* assumption about signal) to find rare events

# Anomaly detection @ LHC

- **what?** → A type of model-independent search
- **why?** → Explore phase-space missed by direct searches
- **How?** → Use the **large dataset** (+ some\* assumption about signal) to find rare events



# Anomaly detection @ LHC

- **what?** → A type of model-independent search
- **why?** → Explore phase-space missed by direct searches
- **How?** → Use the large dataset (+ some\* assumption about signal) to find **rare events**



# Anomaly detection @ LHC

- **what?** → A type of model-independent search
- **why?** → Explore phase-space missed by direct searches
- **How?** → Use the large dataset (+ some\* assumption about signal) to find rare events ⇒ great task for Machine Learning!

# Anomaly detection @ LHC

- **what?** → A type of model-independent search
- **why?** → Explore phase-space missed by direct searches
- **How?** → Use the large dataset (+ some\* assumption about signal) to find rare events ⇒ great task for Machine Learning!

Supervised

→ labelled data



Weakly-supervised

→ No 'exact' labels



Unsupervised

→ No labels

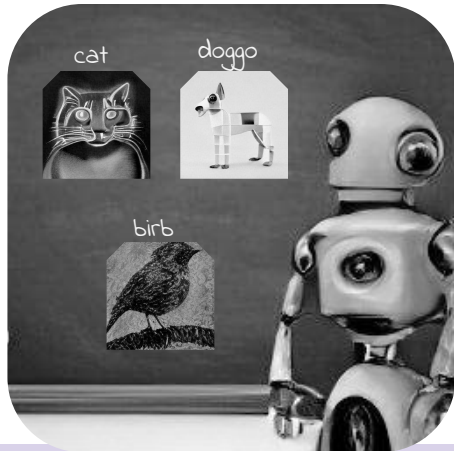


# Anomaly detection @ LHC

- **what?** → A type of model-independent search
- **why?** → Explore phase-space missed by direct searches
- **How?** → Use the large dataset (+ some\* assumption about signal) to find rare events ⇒ great task for Machine Learning!

Supervised

→ labelled data



Weakly-supervised

→ No 'exact' labels



Unsupervised

→ No labels



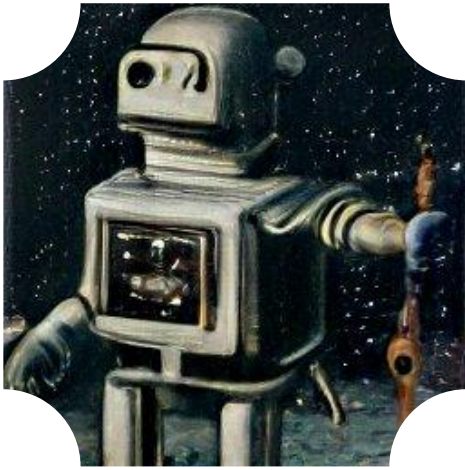


# Current methods using jets

---

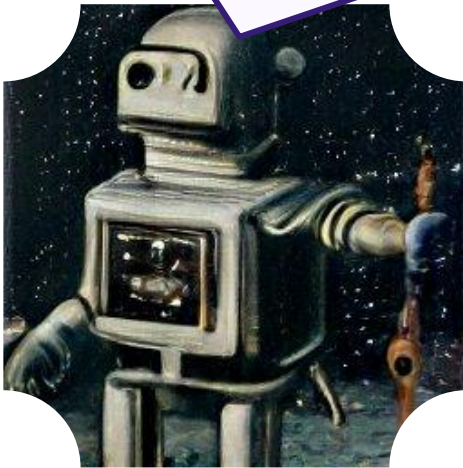


# Anomaly detection methods - weak supervision



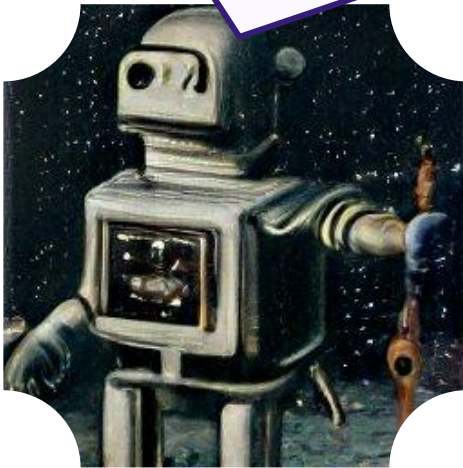
# Anomaly detection methods - weak supervision

Tell me the data  
labels without telling  
me the data labels



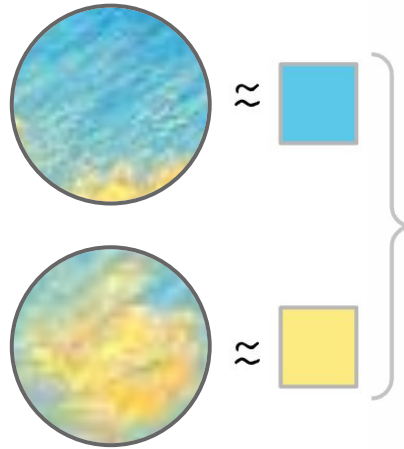
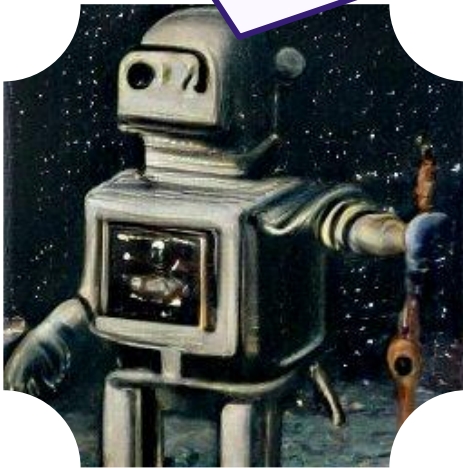
# Anomaly detection methods - weak supervision

Tell me the data labels without telling me the data labels



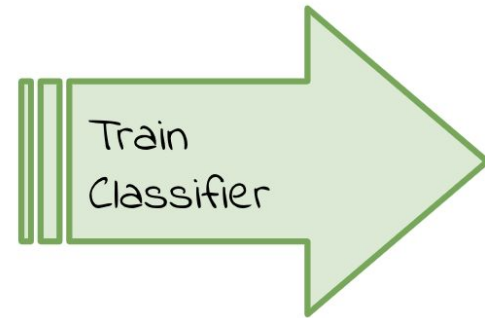
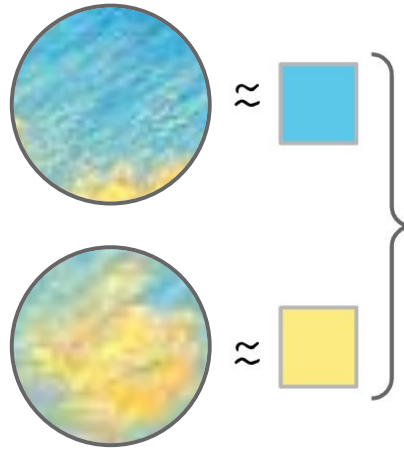
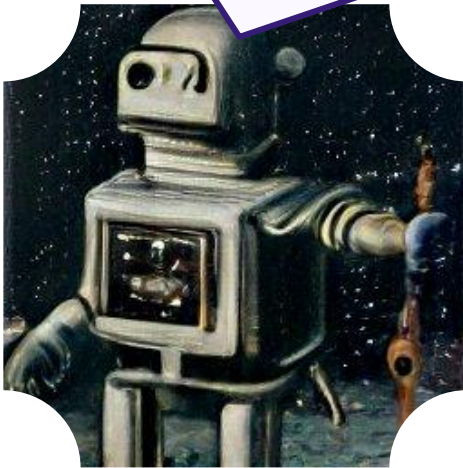
# Anomaly detection methods - weak supervision

Tell me the data labels without telling me the data labels



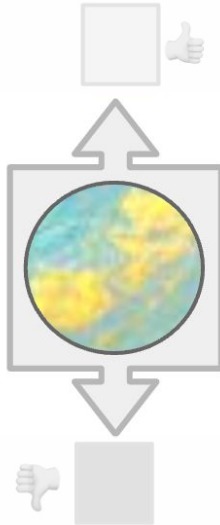
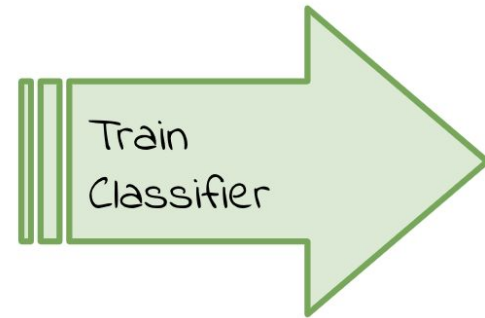
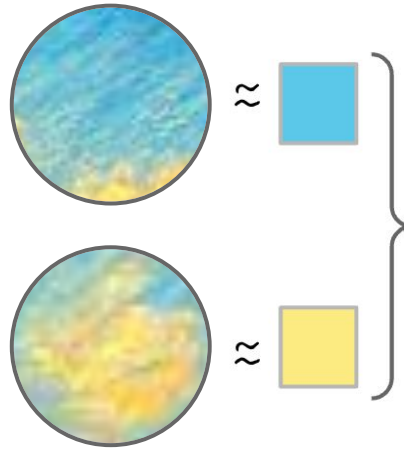
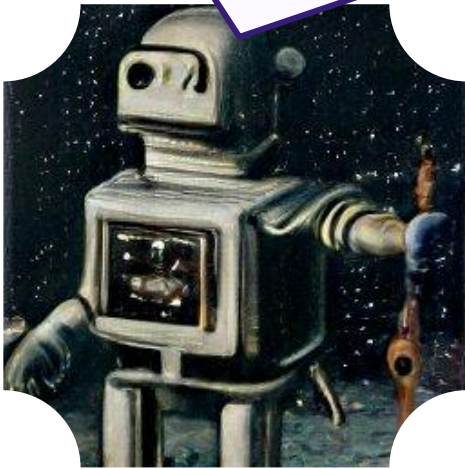
# Anomaly detection methods - weak supervision

Tell me the data labels without telling me the data labels



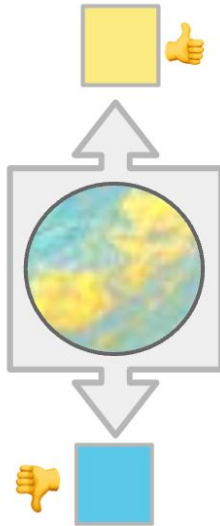
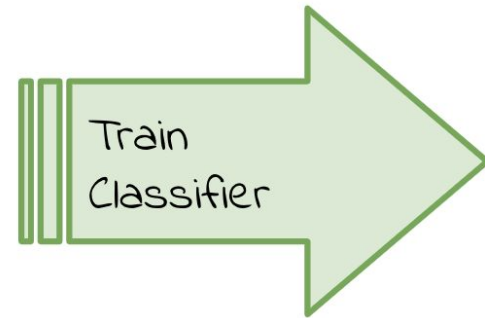
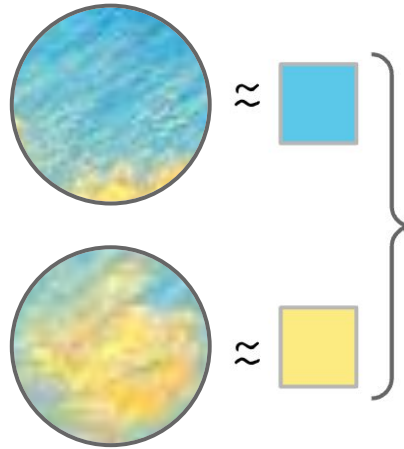
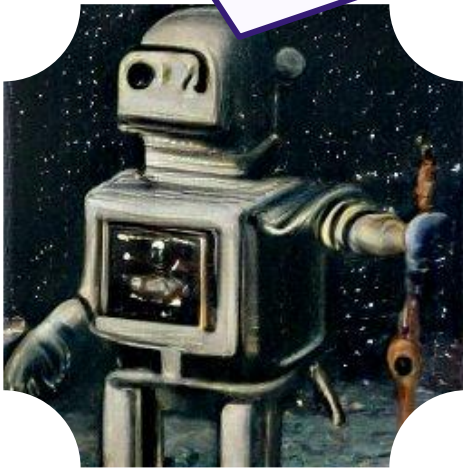
# Anomaly detection methods - weak supervision

Tell me the data labels without telling me the data labels



# Anomaly detection methods - weak supervision

Tell me the data labels without telling me the data labels





# Anomaly detection methods - weak supervision

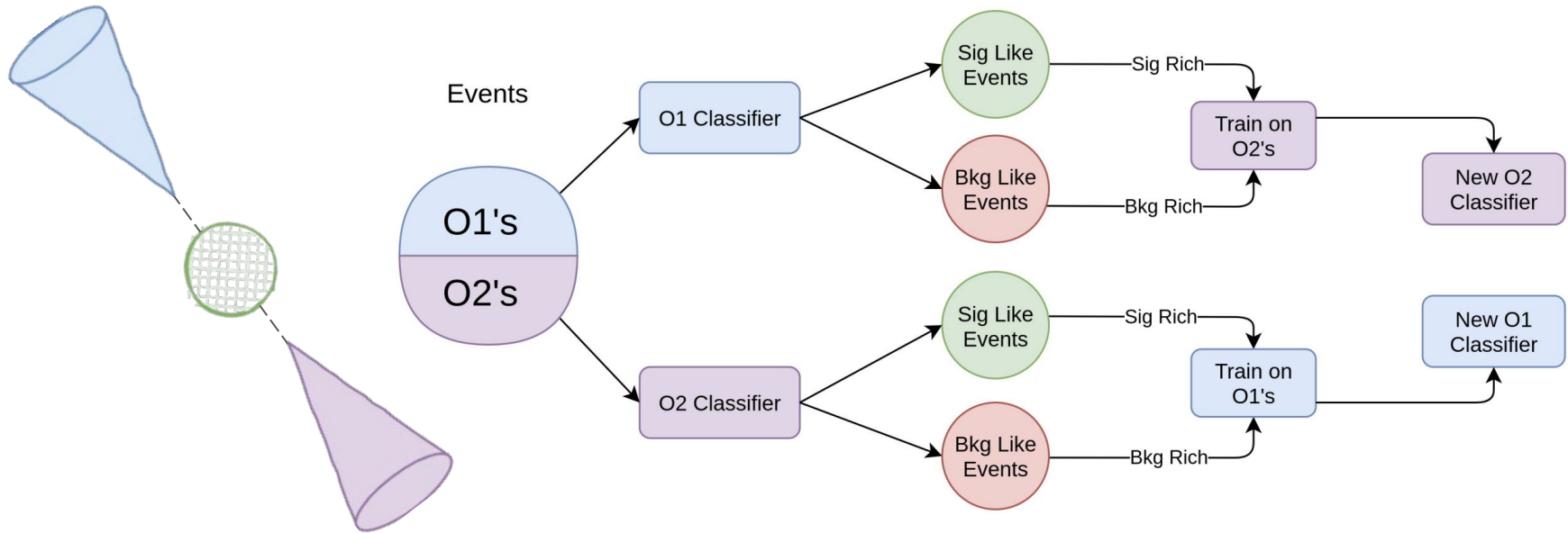
---

Several proposals for weak/un supervised ML @ LHC!

# Anomaly detection methods - weak supervision

Several proposals for weak/un supervised ML @ LHC!

## »» Tag N' Train



# Anomaly detection methods - weak supervision

---

Several proposals for weak/un supervised ML @ LHC!

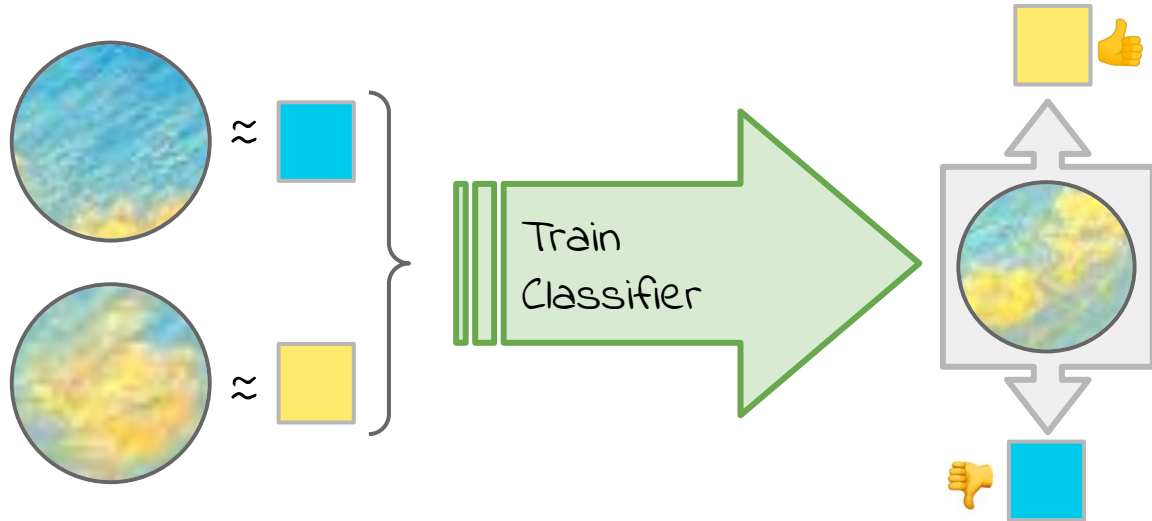
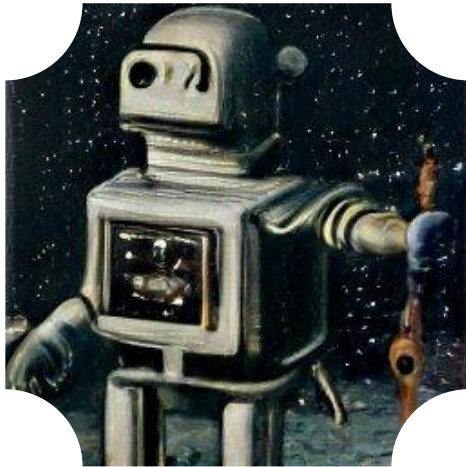
»»→ Tag N' Train

»»→ Classification Without Labels

# Anomaly detection methods - weak supervision

Several proposals for weak/un supervised ML @ LHC!

- »» Tag N' Train
- »» Classification Without Labels

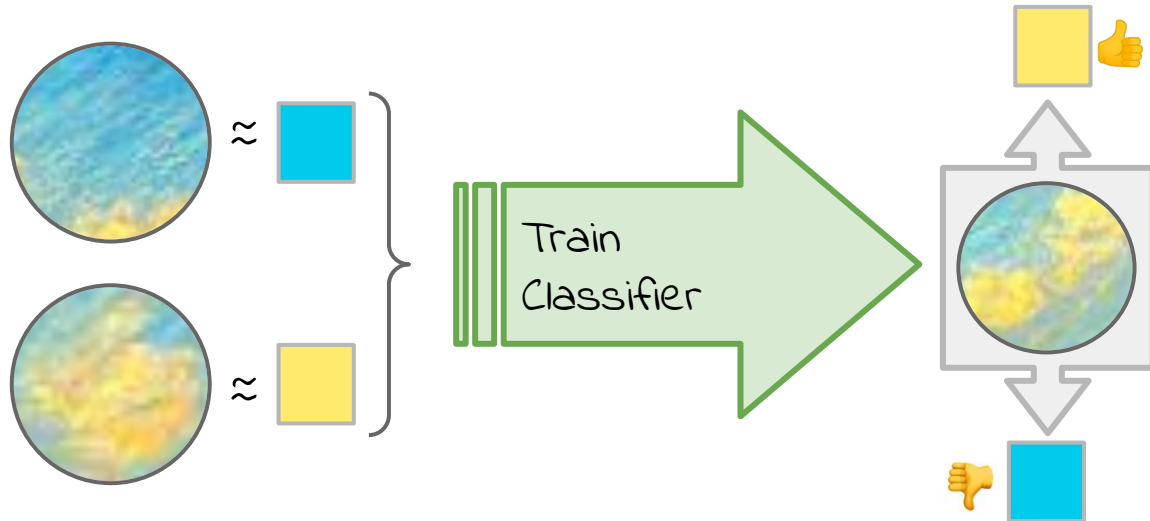
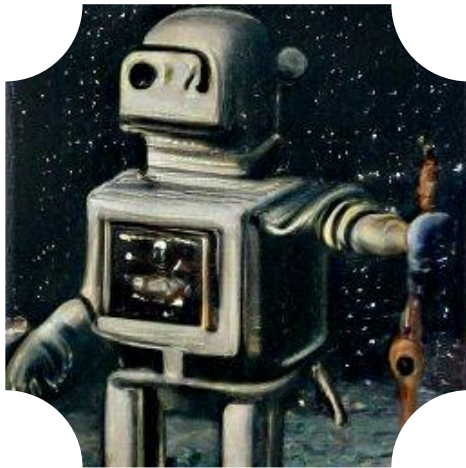


# Anomaly detection methods - weak supervision

Several proposals for weak/un supervised ML @ LHC!

»» Tag N' Train

»» CWoLa

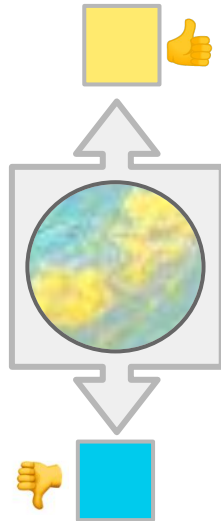
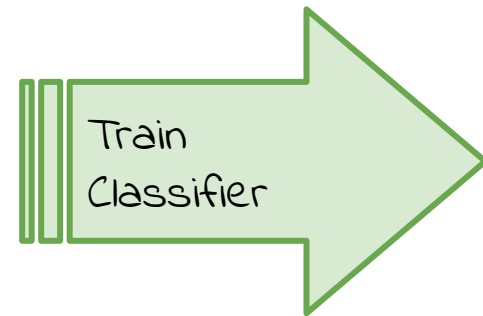
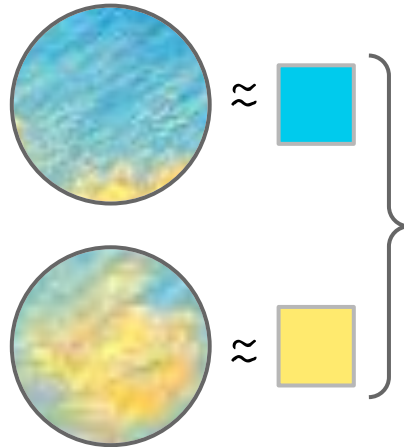
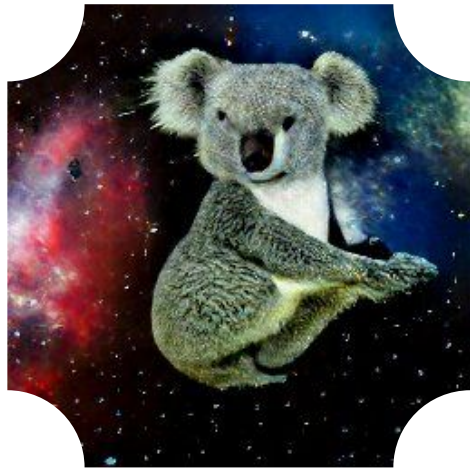


# Anomaly detection methods - weak supervision

Several proposals for weak/un supervised ML @ LHC!

»» Tag N' Train

»» CWoLa

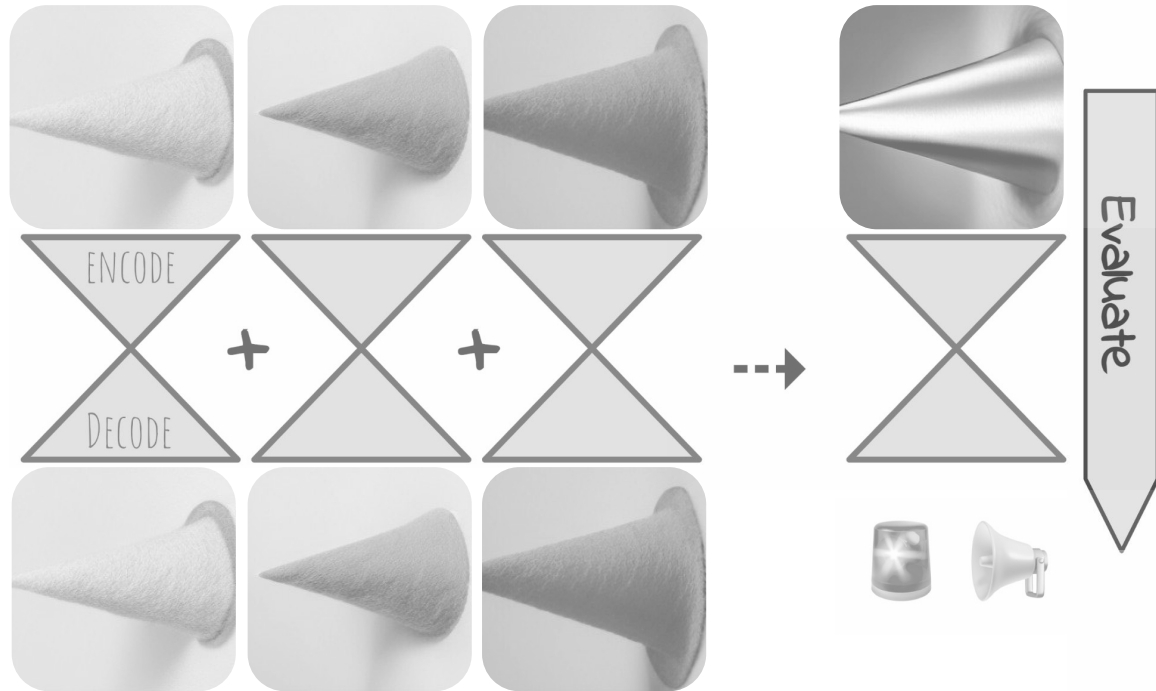


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!

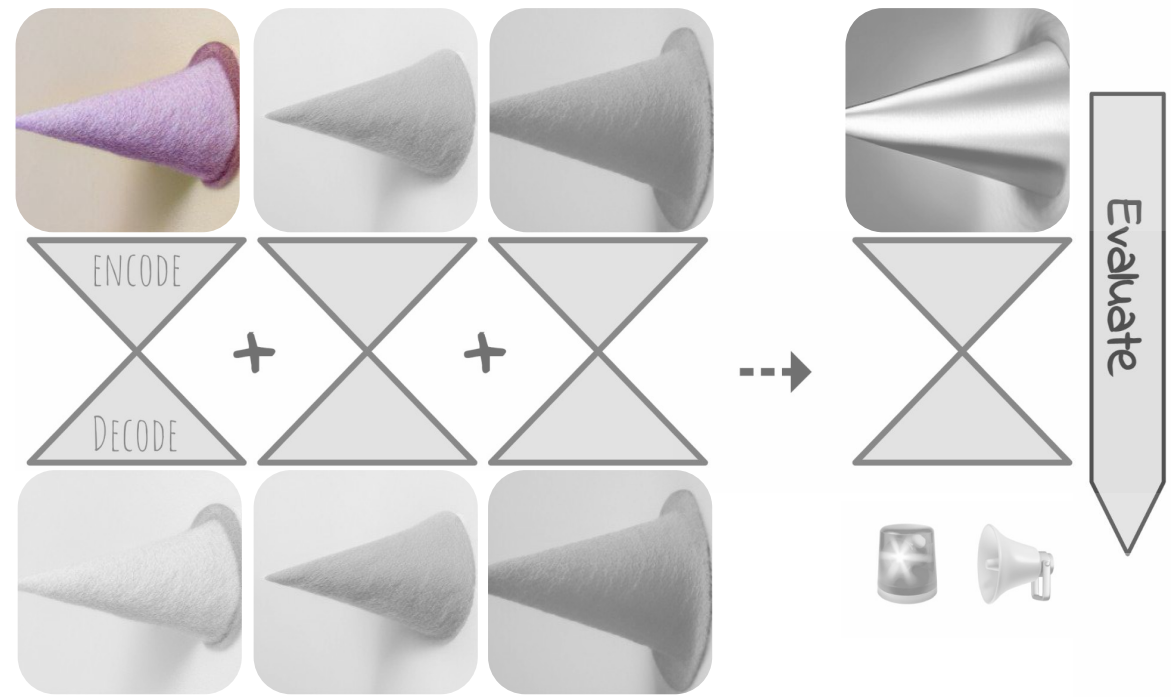


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!



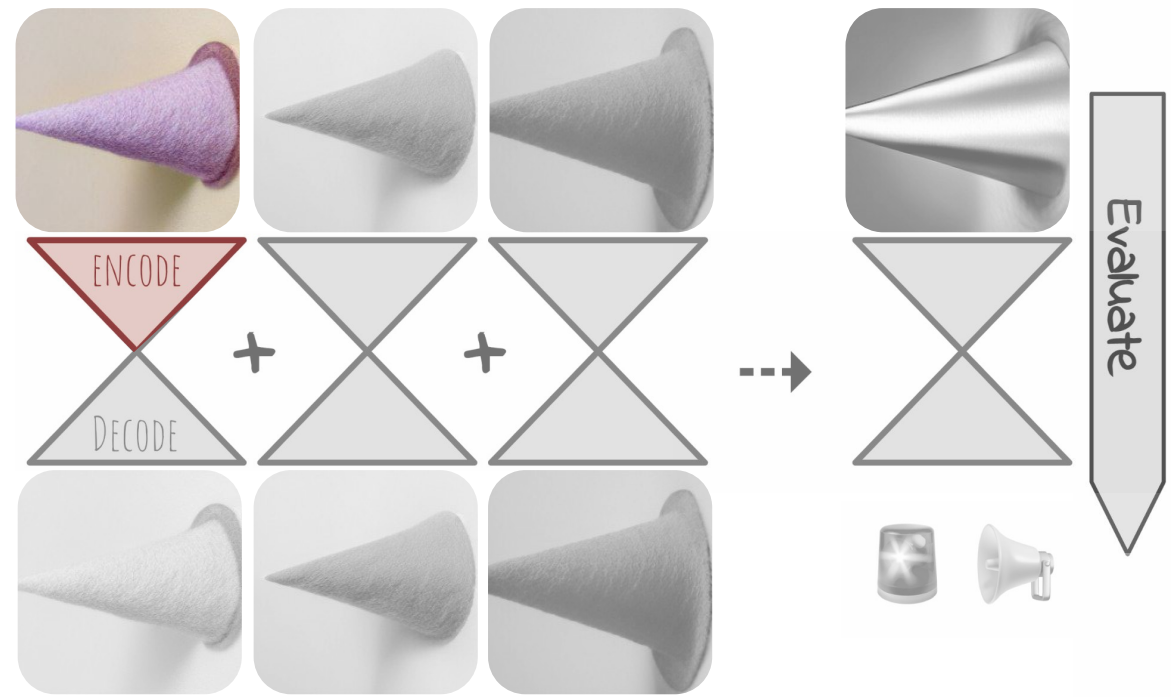


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!

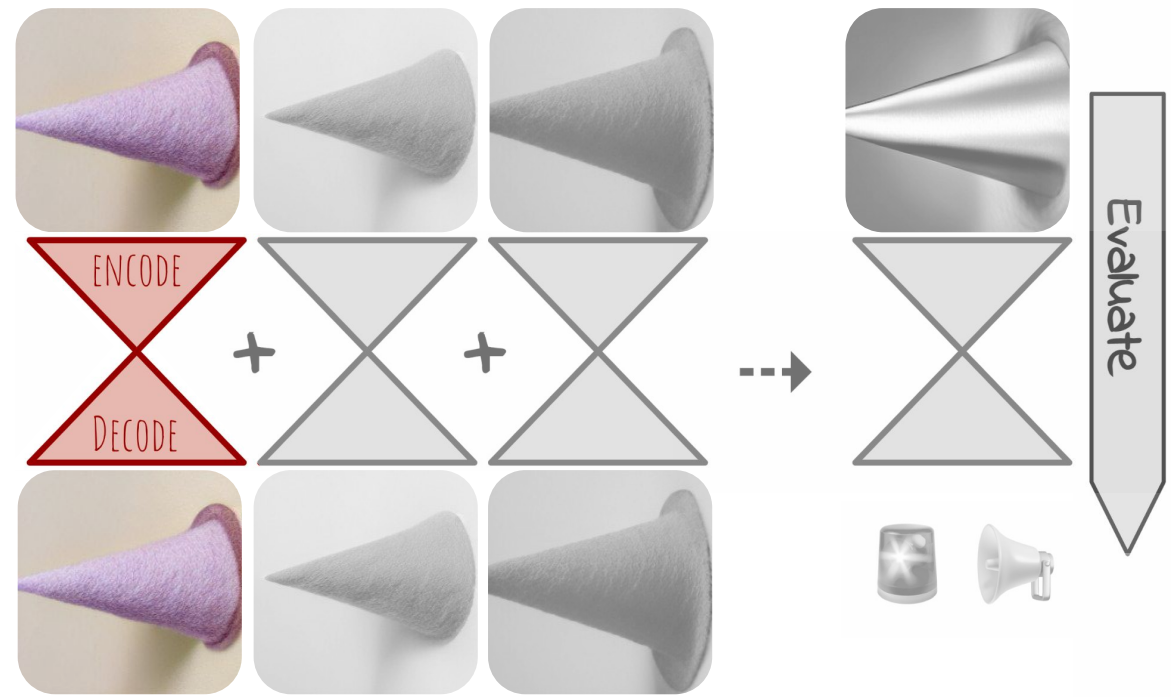


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!

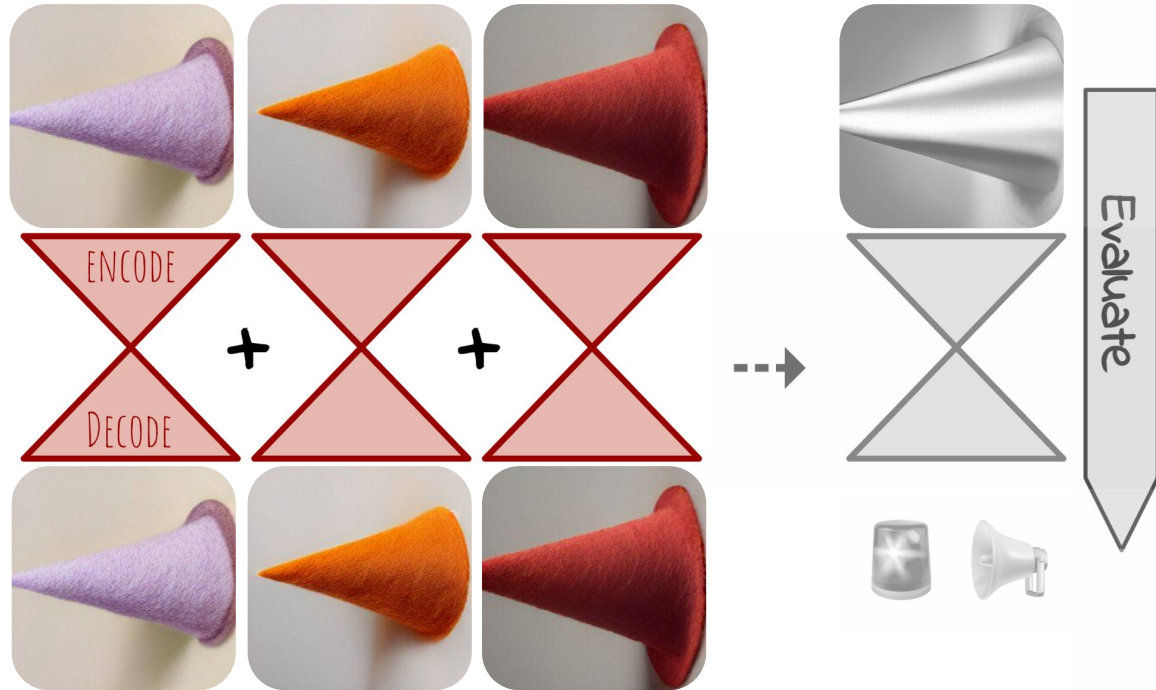


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!



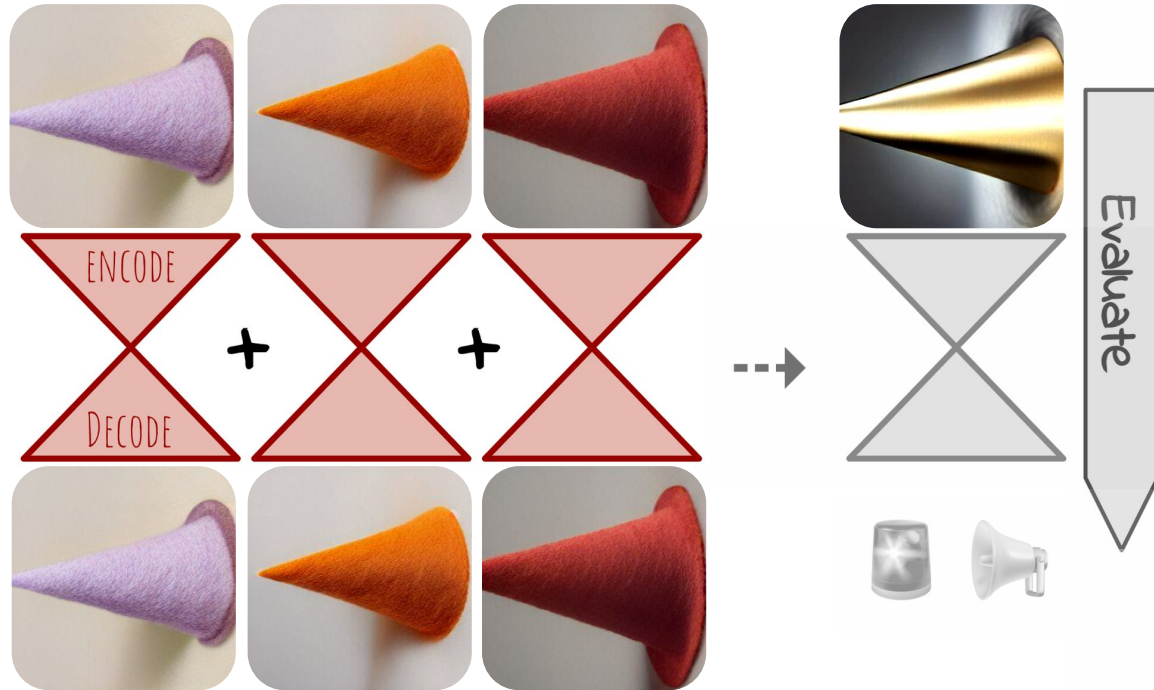
Train

# Anomaly detection methods - Autoencoders



Learn what background jets look like

» Tell us if something is not that!

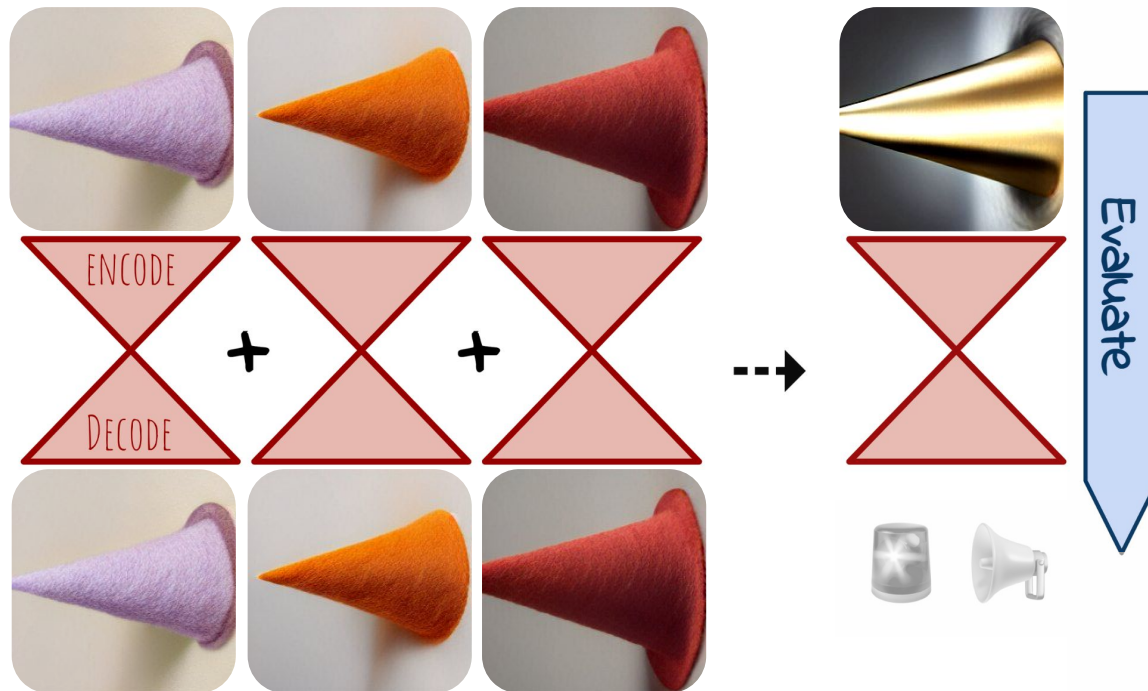


# Anomaly detection methods - Autoencoders



Learn what background jets look like

»» Tell us if something is not that!

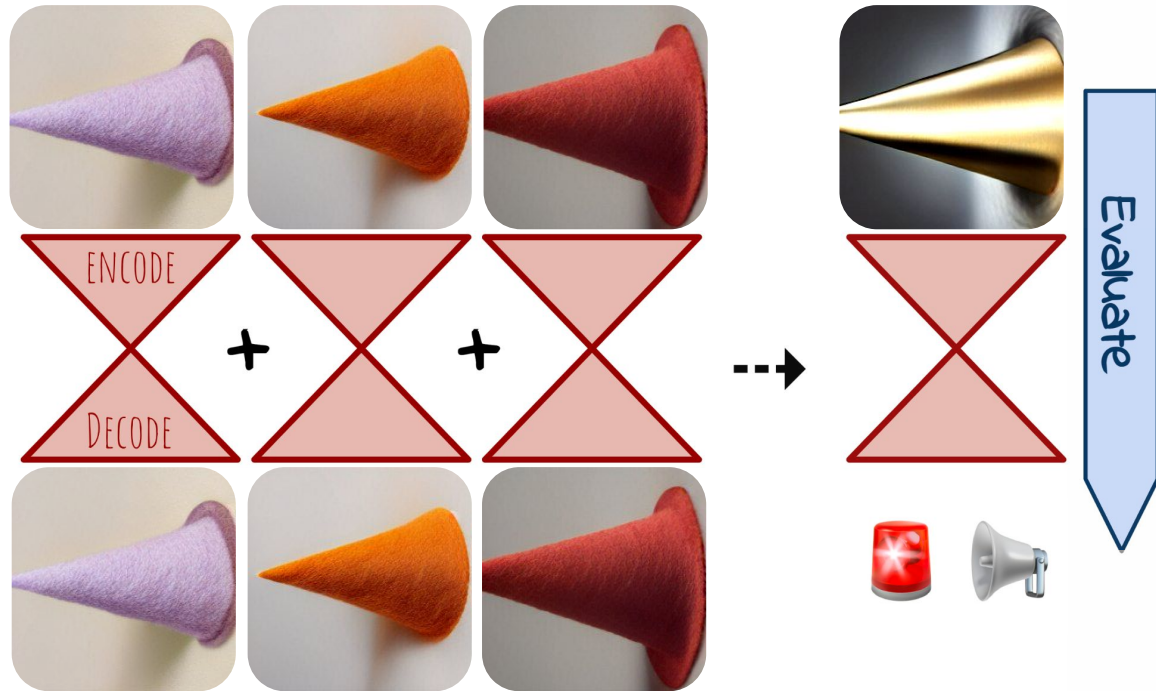


# Anomaly detection methods - Autoencoders



Learn what background jets look like

» Tell us if something is not that!



Train

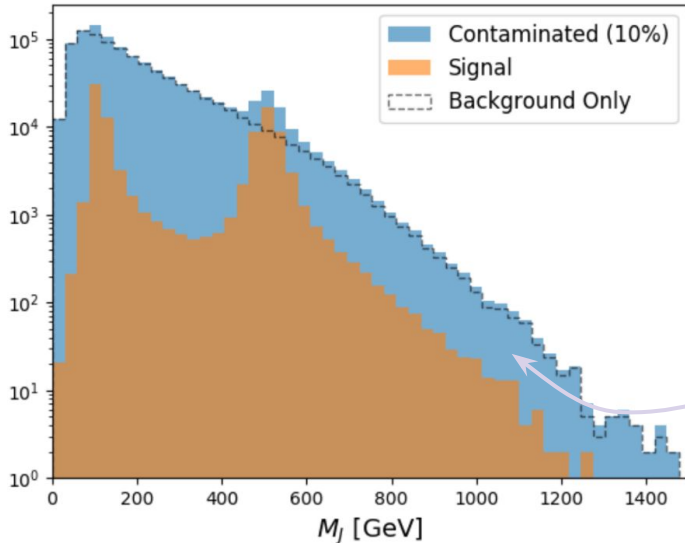
# Anomaly detection methods - Autoencoders



Autoencoders for LHC data analysis:

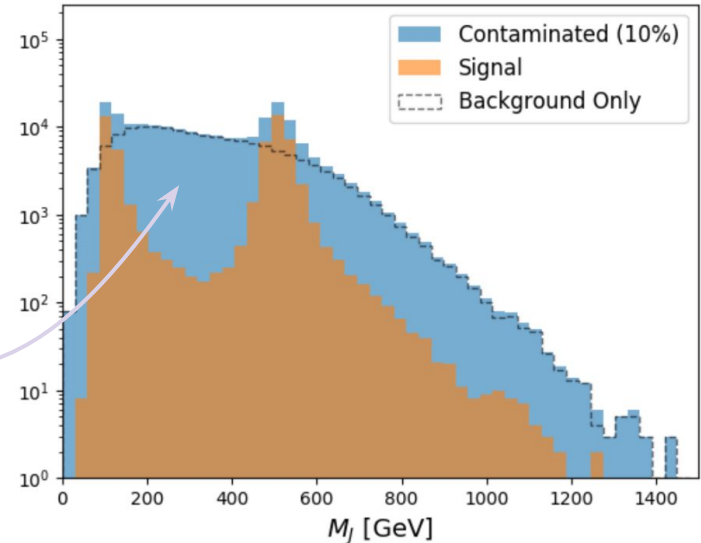
- » Quasi Anomalous Knowledge (QuAK)
- » Variational Recurrent Neural Network (VRNN)

2-Prong: Leading Jet Mass



Injected  
100 [GeV] and  
500 [GeV]  
jets

2-Prong: Leading Jet Mass, Anomaly Score > 0.65



# Ongoing efforts and recent result

---





# Experimental results

- CMS → working on an analysis combining several of these methods
  - No public results yet
- ATLAS → Two jet-based anomaly detection publications
  - Dijet search with weak supervision [[HDBS-2018-59](#)]
  - $Y \rightarrow XH \rightarrow qqbb$  [[ATLAS-CONF-2022-045](#)]



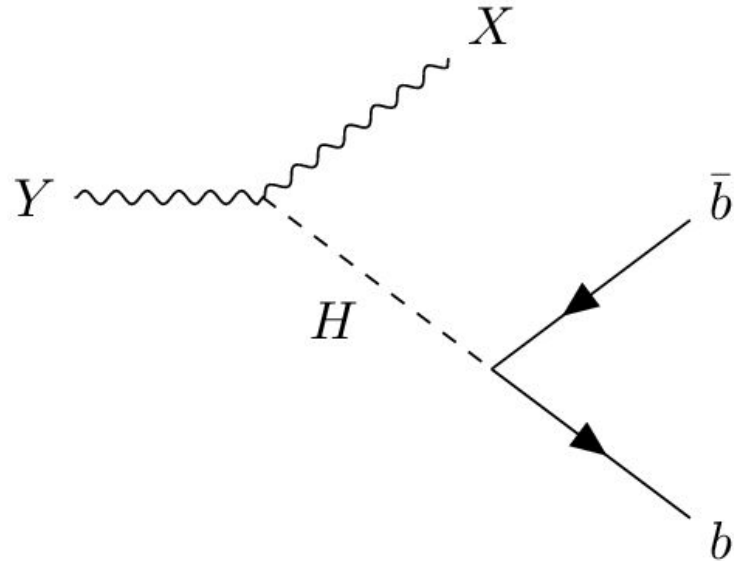
# Experimental results

- CMS → working on an analysis combining several of these methods
  - » No public results yet
- ATLAS → Two jet-based anomaly detection publications
  - » Dijet search with weak supervision [[HDBS-2018-59](#)]
  - »  $Y \rightarrow XH \rightarrow qqbb$  [[ATLAS-CONF-2022-045](#)]



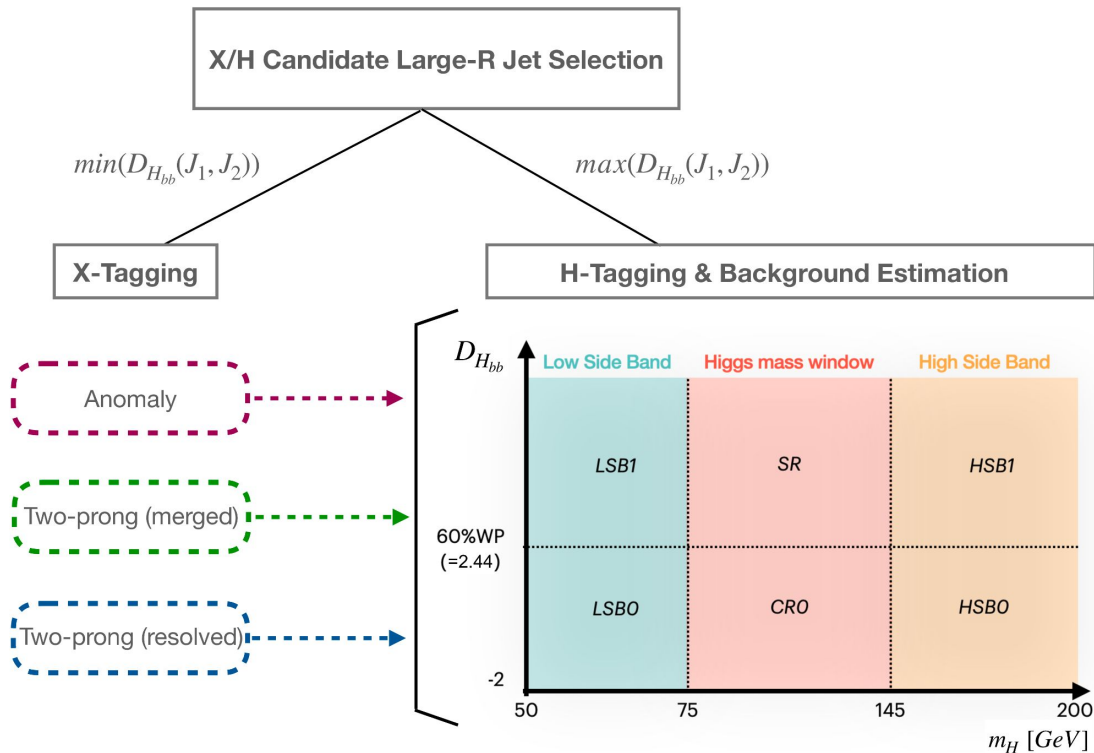
# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

- Search for new resonance
- $Y$  decaying to  $XH$
- Main feature:
  - Use a Variational Recurrent Neural Network ([VRNN](#)) to detect anomalous jets



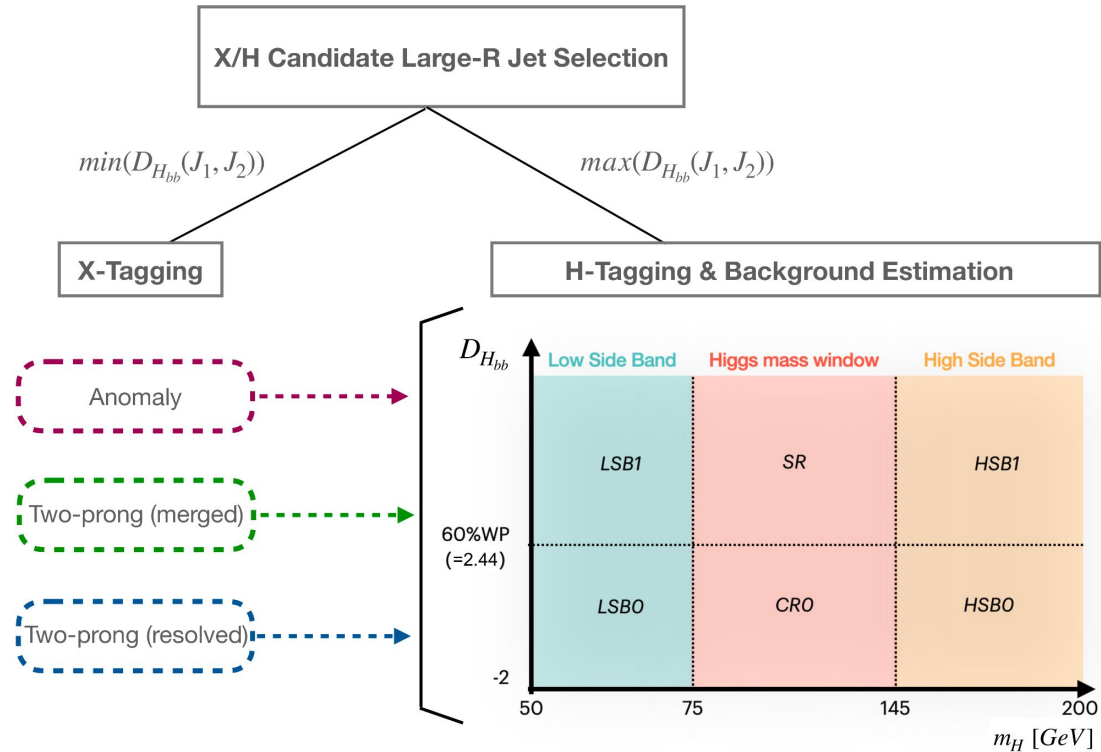
# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

- Select events with **two large jets**
- Use  **$H \rightarrow bb$**  jet to define **background estimation** regions
  - Use the **other one** for **anomaly detection**



# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

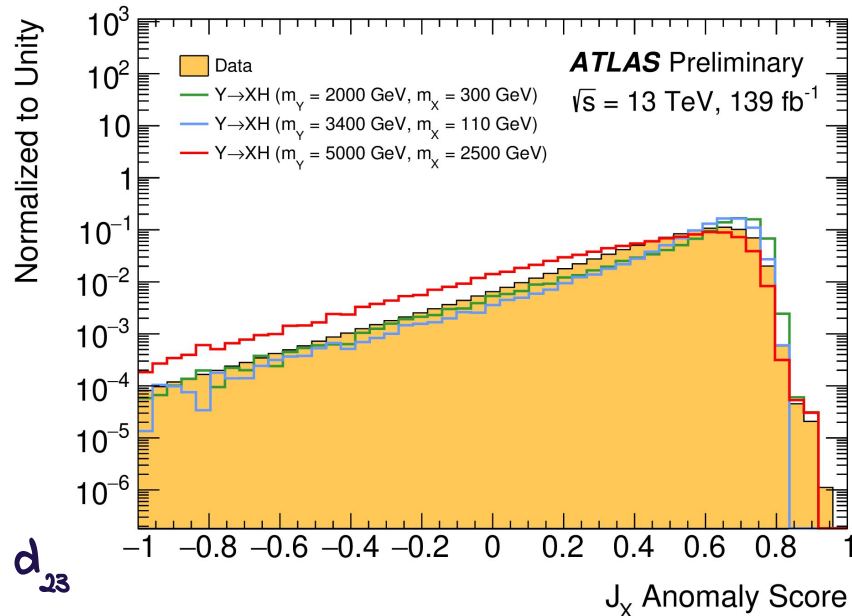
- **Three separate signal regions** using VRNN
  - **Anomaly**
  - **Two-prong merged**
  - **Two-prong resolved**
- **Bump-hunt in all SRs**
  - **Signal+Background fit in two-prong regions only**



# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

## VRNN inputs:

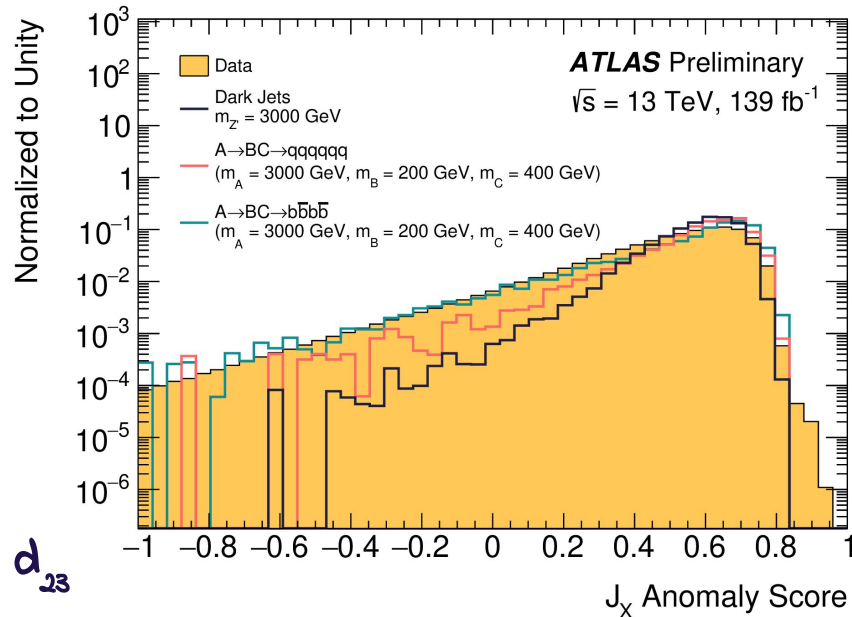
- Only jets from data
- $\leq 20$  jet-constituents from each
- Jet-substructure variables
  - ↪ Energy correlator ---  $\mathcal{D}_2$
  - ↪ N-subjettines ---  $\tau_{3,2}$
  - ↪  $k_t$  splitting scale ---  $d_{1,2}, d_{2,3}$
- No labels



# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

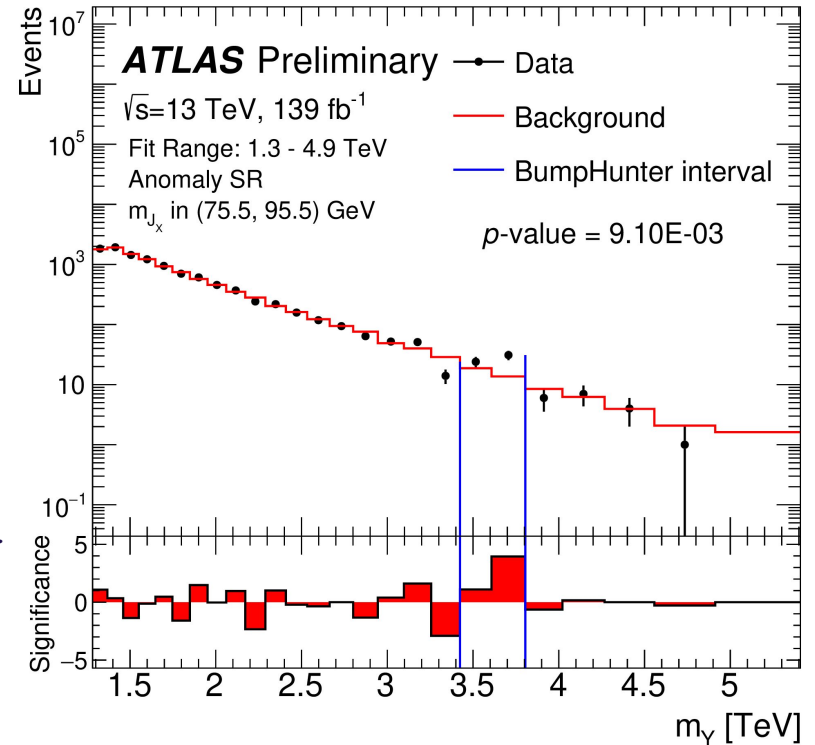
## VRNN inputs:

- Only jets from data
- $\leq 20$  jet-constituents from each
- Jet-substructure variables
  - ↪ Energy correlator ---  $\mathcal{D}_2$
  - ↪ N-subjettines ---  $\tau_{3,2}$
  - ↪  $k_t$  splitting scale ---  $d_{1,2}, d_{2,3}$
- No labels



# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

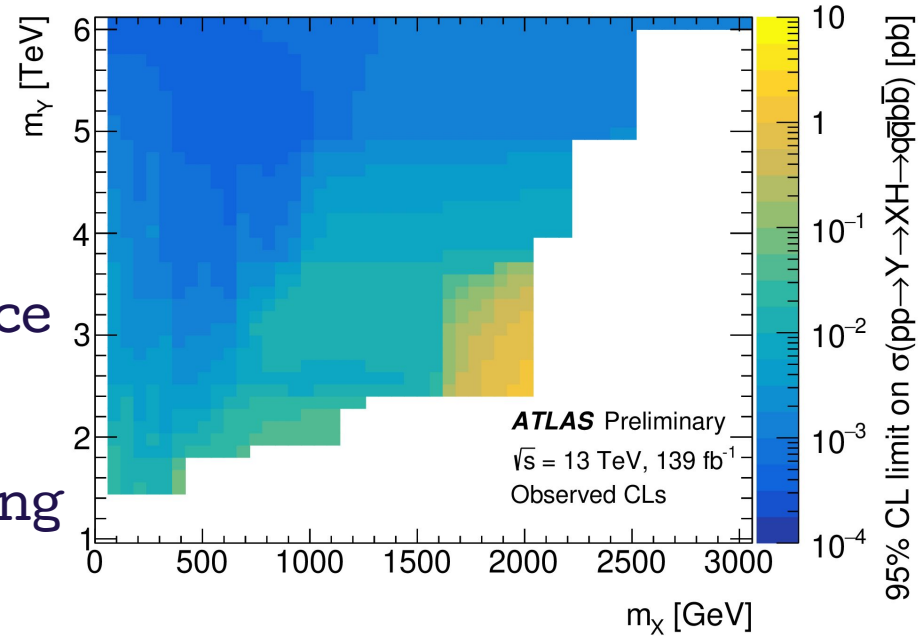
- **No significant excess** in
  - Anomaly signal region or
  - $1.47 \sigma$  global significance
  - Two-prong signal region
- **Limits set** on  $Y \rightarrow XH \rightarrow qqbb$  using **two-prong results**.





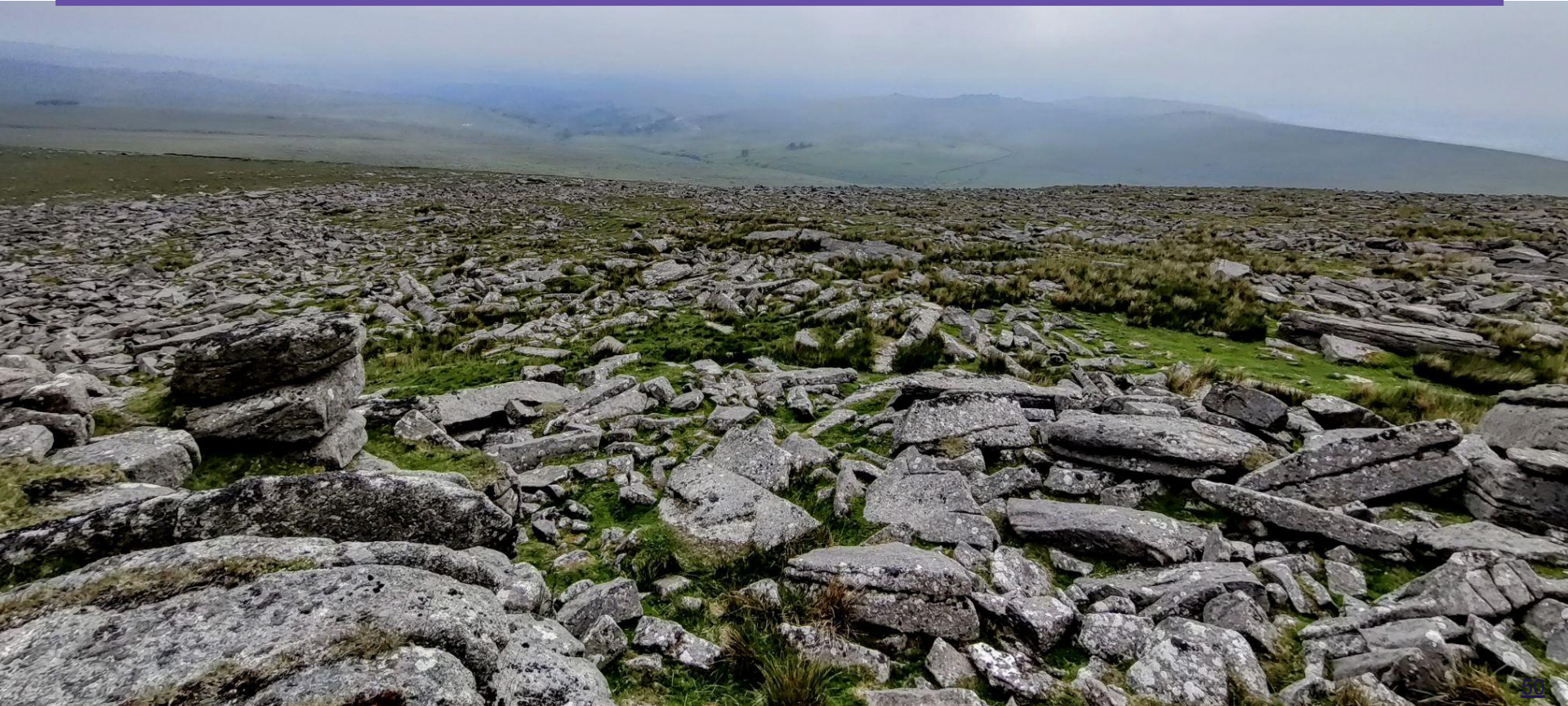
# ATLAS anomaly detection in $Y \rightarrow XH \rightarrow qqbb$

- **No significant excess** in
  - Anomaly signal region or
  - $1.47 \sigma$  global significance
  - Two-prong signal region
- **Limits set** on  $Y \rightarrow XH \rightarrow qqbb$  using **two-prong results.**



# Summary

---

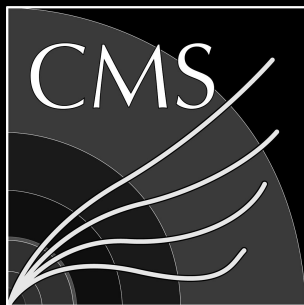


# Summary

---

- **Anomaly detection** is an exciting area for **model-independent searches** at the LHC
  - **Jets** are a **prime use case**
- These **state-of-the-art ML** methods are **already** being **implemented** in **real LHC data**
- A lot of work being done at the LHC experiments to implement these ideas
  - **Stay tuned** for new results!
- Still early stages
  - Both **search** and **statistical** analysis **strategies** are rapidly **evolving**

**iihe**  
BRUXELLES BRUSSEL



**ULB**



# Thanks!

**Santiago Paredes Saenz**

On behalf of the ATLAS and CMS collaborations

[santiago.paredes@cern.ch](mailto:santiago.paredes@cern.ch)

**ISMD 2022**

August 2022



# Event Selection

Parameter	Preselection requirements				
$m_{JJ}$ [GeV]	$> 1300$				
$p_T(J_1)$ [GeV]	$> 500$				
$m_J$ [GeV]	$m_{J_1} > 50 \parallel m_{J_2} > 50$				
$D_{Hbb}$	$> -2$				
	Signal regions				
	Merged	Resolved	Anomaly		
$m_H$ [GeV]	(75, 145)				
$D_{Hbb}$	$> 2.44$				
$D_2^{trk}$	$< 1.2$	$> 1.2$	-		
$ \Delta y_{j_1, j_2} $	-	$< 2.5$	-		
$p_T^{bal}$	-	$< 0.8$	-		
Anomaly Score	-	-	$> 0.5$		
	Background estimation regions				
	CR0	HSB0	HSB1	LSB0	LSB1
$m_H$ [GeV]	(75, 145)	(145, 200)		(65, 75)	
$D_{Hbb}$	$< 2.44$	$< 2.44$	$> 2.44$	$< 2.44$	$> 2.44$

# Anomaly detection methods: weak supervision

Several proposals for weak/un supervised ML @ LHC!

- »» [Tag N' Train](#)
- »» [CWoLa](#)
- »» [CATHODE](#)

