Discussion on Using Unsupervised ML for Anisotropy Analysis

Unsupervised ML in astronomy

- Unsupervised ML is already being used for many purposes.
- It is also being used to classify supernovae and galaxy types.
- While it produces results equivalent to conventional methods, there have been changes a few objects in classification.
- And it has already reached the point where applications of unsupervised machine learning are being reviewed. “S. Fotopoulou, A review of unsupervised learning in astronomy”

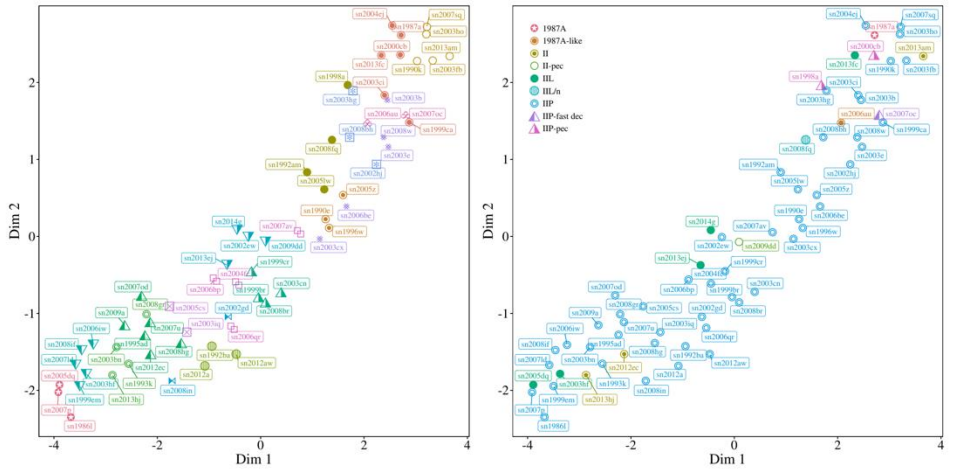
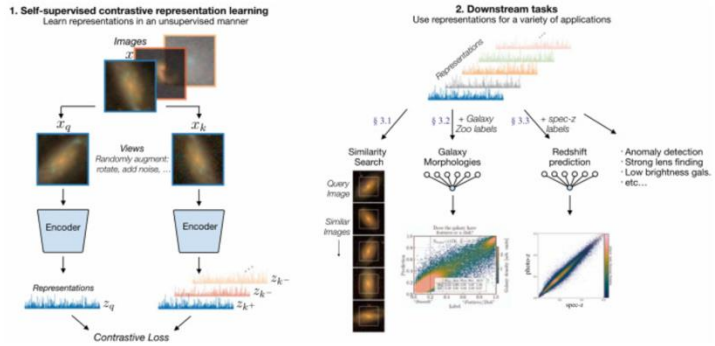


Figure 7: The two-dimensional embedding of SN spectra by UMAP at maximum light is depicted in two panels. The left panel displays our graph-based community groups, while the right panel depicts the standard classification from the literature. In both panels, each SN is coded with a unique colour and shape according to its respective classification, which was assigned independently of the UMAP embedding.

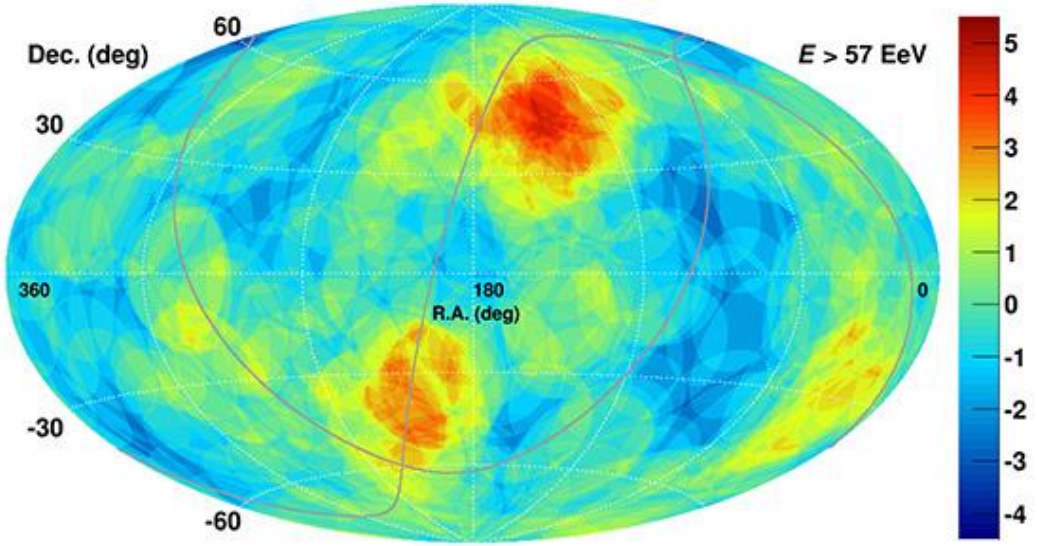


Download: Download high-res image (460KB)
Download: Download full-size image

Fig. 6. Application of contrastive self-supervised learning. Input features are augmented (rotation, flip, etc.), and the contrastive loss is trained to bring representations of the same object closer in the latent space.
Figure adopted from Hayat et al. (2021).

Anisotropy in energy of UHECR

- The energy anisotropy of UHECR was suggested over 10 years ago.
- Discussion continues, as do observations to confirm this.



Kawata et al., ICRC 2015

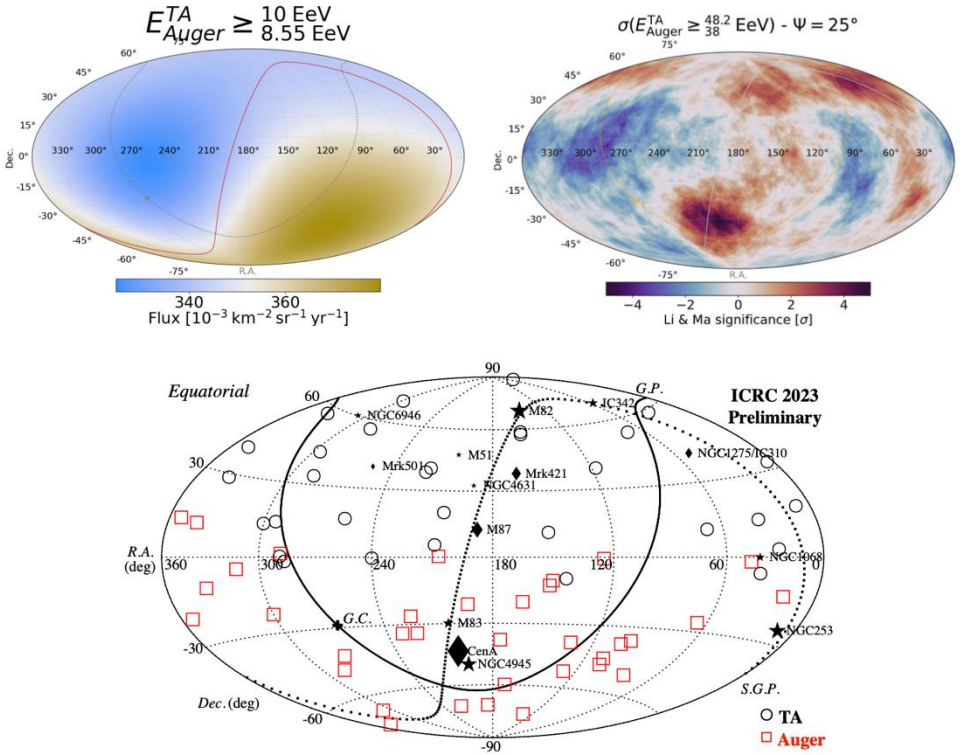


Figure 10: UHECR sky-maps around the "ankle", "cutoff", and above 100 EeV. The flux sky-map of 45° oversamplings with energies around the "ankle" region (top-left) and significance sky-map of 25° oversamplings around the "cutoff" region (top-right) in equatorial coordinates reported by the Auger-TA anisotropy working group [65]. The bottom figure indicates the arrival directions of UHECRs above 100 EeV measured by Auger and TA, together with nearby astronomical source candidates.

T. Fujii, ICRC2023

Possibility of Xmax anisotropy detection in HotSpot

- The background assumes a composition distribution based on Auger.
- Only events within the HotSpot are assumed to be purely composed of various particles.

From these results, does this mean that if events within the HotSpot are composed of light particles, there will be differences in the Xmax distribution inside and outside the HotSpot, even for current TA-SD events?

Table 4: Ratio of composition per energy

	Proton : CNO : Si : Iron(Background)
$10^{19.7\sim19.8}[eV]$	0.0 : 4.4 : 4.4 : 1.2
$10^{19.8\sim19.9}[eV]$	0.0 : 2.6 : 5.4 : 2.0
$10^{19.9\sim20.0}[eV]$	0.0 : 1.6 : 5.2 : 3.2
	Proton : CNO : Si : Iron(HotSpot)
$10^{19.7\sim19.8}[eV]$	7.6 : 1.1 : 1.1 : 0.2
$10^{19.8\sim19.9}[eV]$	7.6 : 0.6 : 1.3 : 0.5
$10^{19.9\sim20.0}[eV]$	7.6 : 0.4 : 1.2 : 0.8

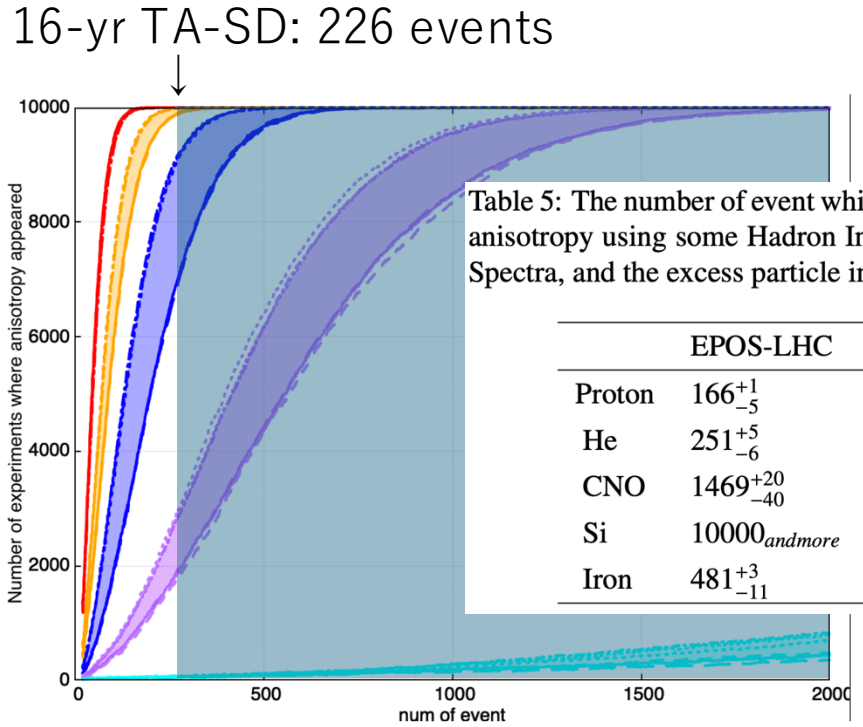


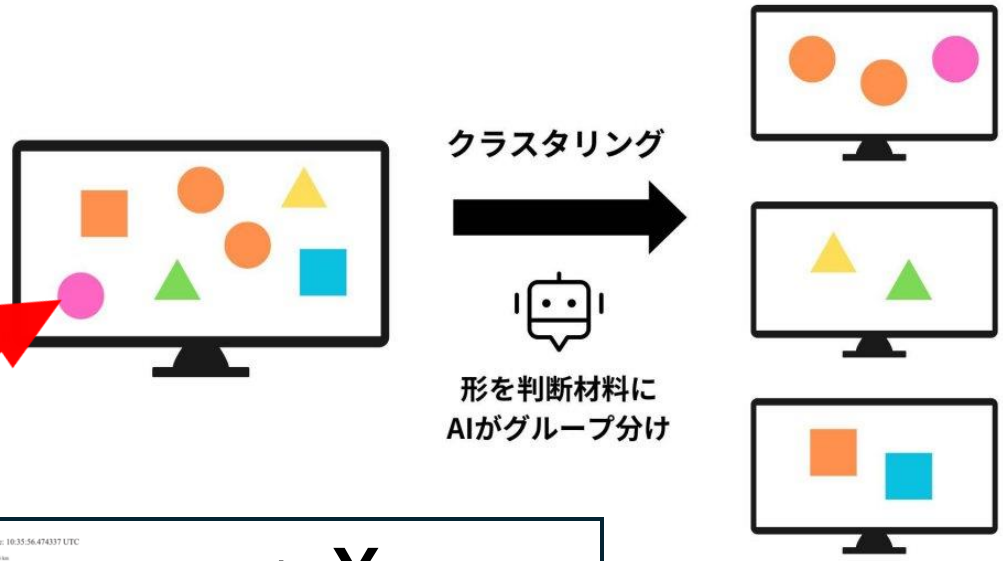
Table 5: The number of event which required to find Xmax anisotropy using some Hadron Interaction model, Energy Spectra, and the excess particle in the HotSpot region

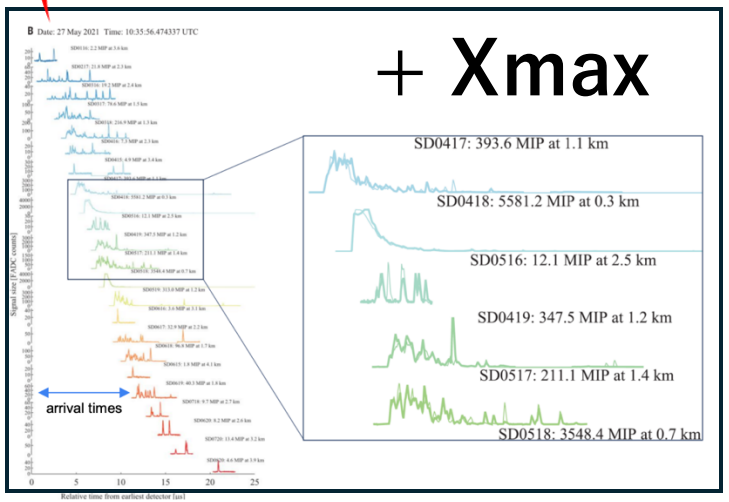
	EPOS-LHC	QGSJET II-04
Proton	166^{+1}_{-5}	158^{+2}_{-3}
He	251^{+5}_{-6}	323^{+6}_{-4}
CNO	1469^{+20}_{-40}	1986^{+14}_{-2}
Si	$10000_{andmore}$	$10000_{andmore}$
Iron	481^{+3}_{-11}	693^{+1}_{-23}

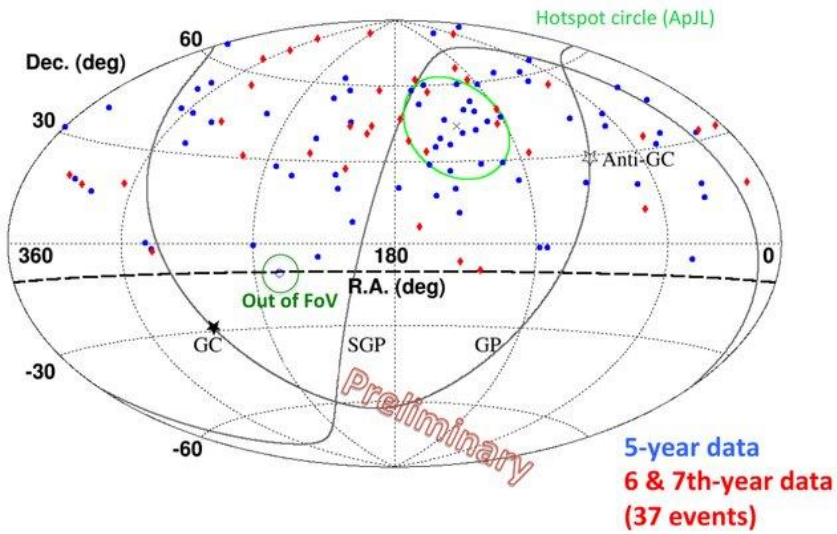
Figure 9: This is the Result when EPOS-LHC was used, the excess particles were He, CNO, Si and Iron particles, and the energy range was changed.Proton, He, CNO, Si, Iron as indicated as Red, Orange, Purple, Dark Blue,Blue respectivily

UHECR2022, R. Saito et, al

My idea.







Classify the raw SD data and "Xmax by ML" as a dataset using unsupervised ML.
Consider specifying the number of classifications (e.g., K-means).

The classification into a Skymap

Note

- The raw SD data also contains energy information, so it would be better to reduce the influence of energy.
- It may be a good to weight by Xmax?
- What can be done to avoid discovering features by chance?
    - ＊If the direction of arrival is incorporated into the training data in advance, it will be possible to evaluate by chance.

- This is still in the idea stage, but we hope to start using it in practice in spring 2026.