

Using modern machine learning methods on KASCADE data for outreach and education

Victoria Tokareva,
victoria.tokareva@kit.edu

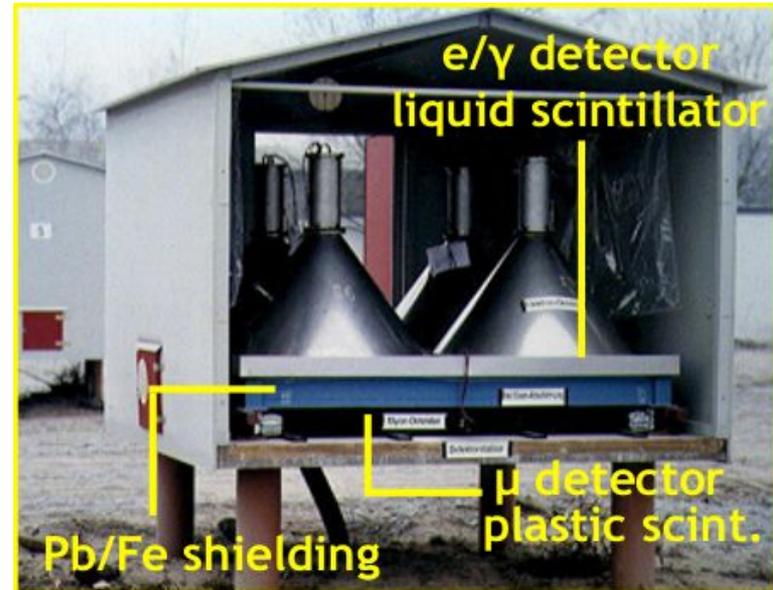
INSTITUT FÜR ASTROTEILCHENPHYSIK (IAP)

28-30 June 2021, Deep Learning in Computational Physics



KASCADE experiment

- Location: 110 m a.s.l., 49° N, 8° E, KIT-Campus North, Karlsruhe, Germany;
- Operation time: 1996 October – 2010 May
⇒ effective time ~ 4223.6 days;
- Area: 200 × 200 m² ;
- 252 scintillator detectors;
- E = 100 TeV – 80 PeV;
- Ne (> 5 MeV);
- Ntr μ (> 230 MeV, r = 40 – 200 m).



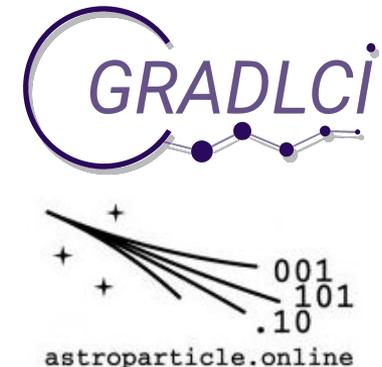
KASCADE Cosmic-Ray Data Center (KCDC)

- created in 2013
- only open-source technologies
- all-in-one: data center, archive, information and educational platform
- <http://kcdc.ikp.kit.edu>



German-Russian Astroparticle Data Life Cycle Initiative (GRADLCI)

- 2018 - this year
- collaboration between KASCADE and TAIGA researchers
- directions: KCDC extension, multimessenger astroparticle physics, data engineering and machine learning for astroparticle physics, outreach (via astroparticle.online)
- <https://gradlc-dc.ikp.kit.edu/>



Machine learning for KASCADE

ML Models

- DecisionTree
- Random Forest
- CNN

Hadronic interaction models

- QGS-jet4
- EPOS-LHC
- Sibyll-23c

Classification

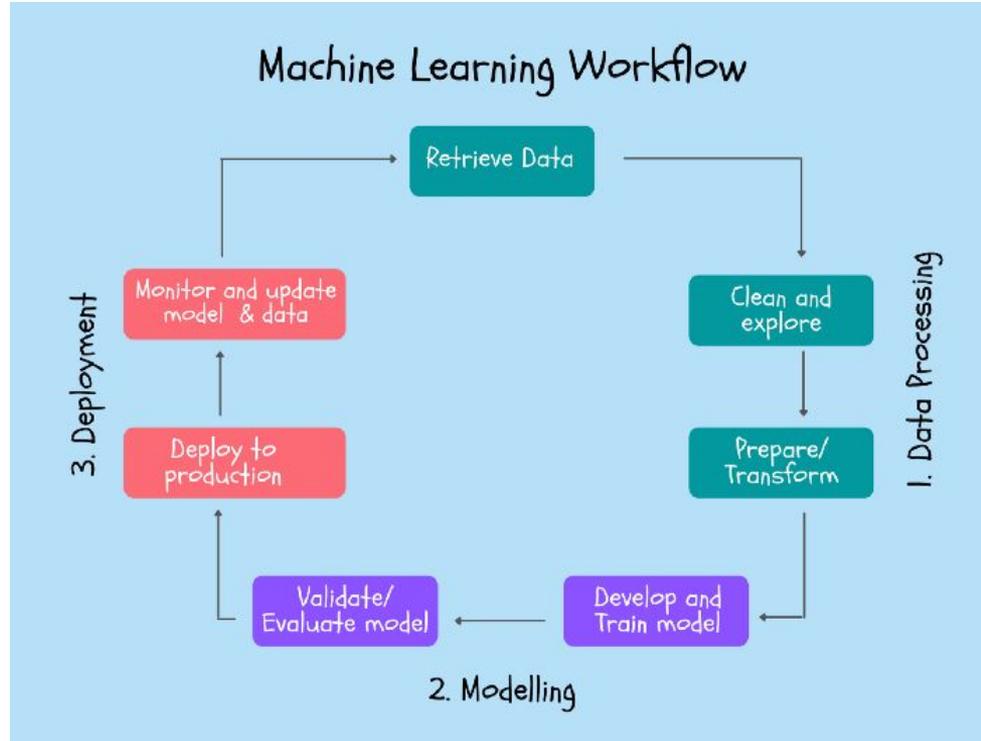
- binary
- multiclass (via particle mass reconstruction)

More in:

ICRC-2021, 13/07/2021, 18:00:

D. Kostunin, I. Plokhikh, M. Ahlers, V. Tokareva, V. Lenok, P. Bezyazeev, S. Golovachev, V. Sotnikov, R. Mullyadzhyanov, E. Sotnikova, New insights from old cosmic rays: A novel analysis of archival KASCADE data

Machine learning workflow



©credit: Arunn Thevapalan, towardsdatascience.com

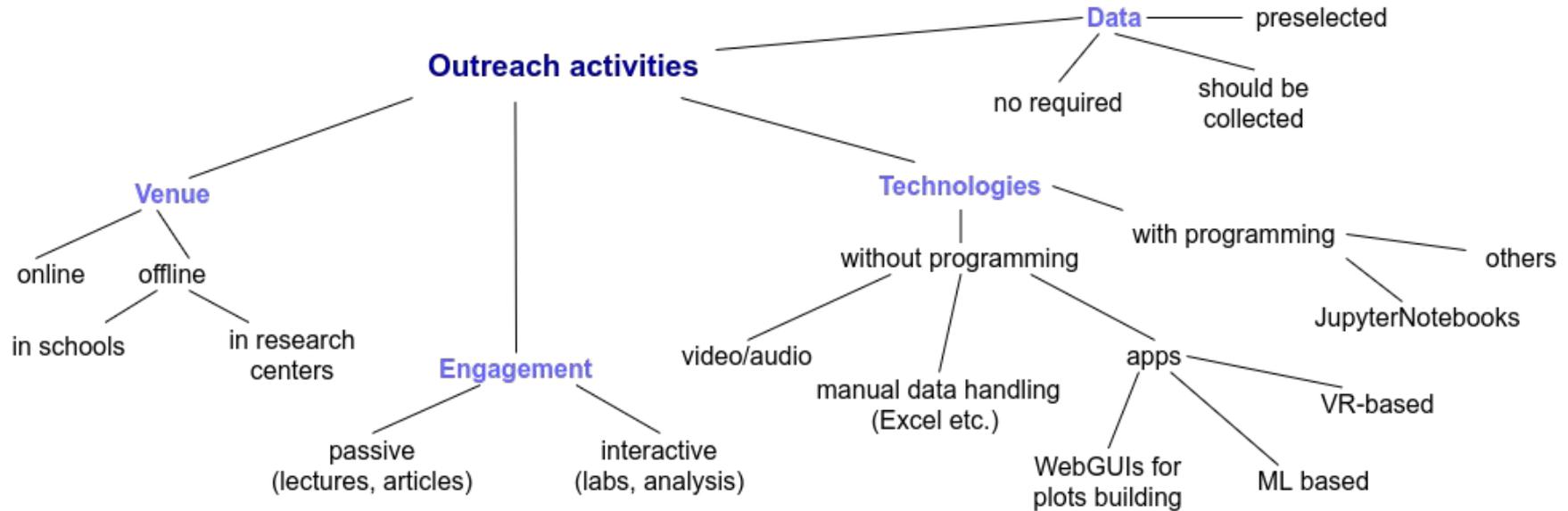
Data workflow engineering

- Data retrieval: KCDC, GRADLCI API
- Data exploration: JupyterHub@KCDC
- Data management: ASW S3
- Preprocessing: IAP computing cluster
- Model training and validation: IAP computing cluster, Google Collaboratory
- Model deploy: Docker, Streamlit
- Project management: JetBrains Space

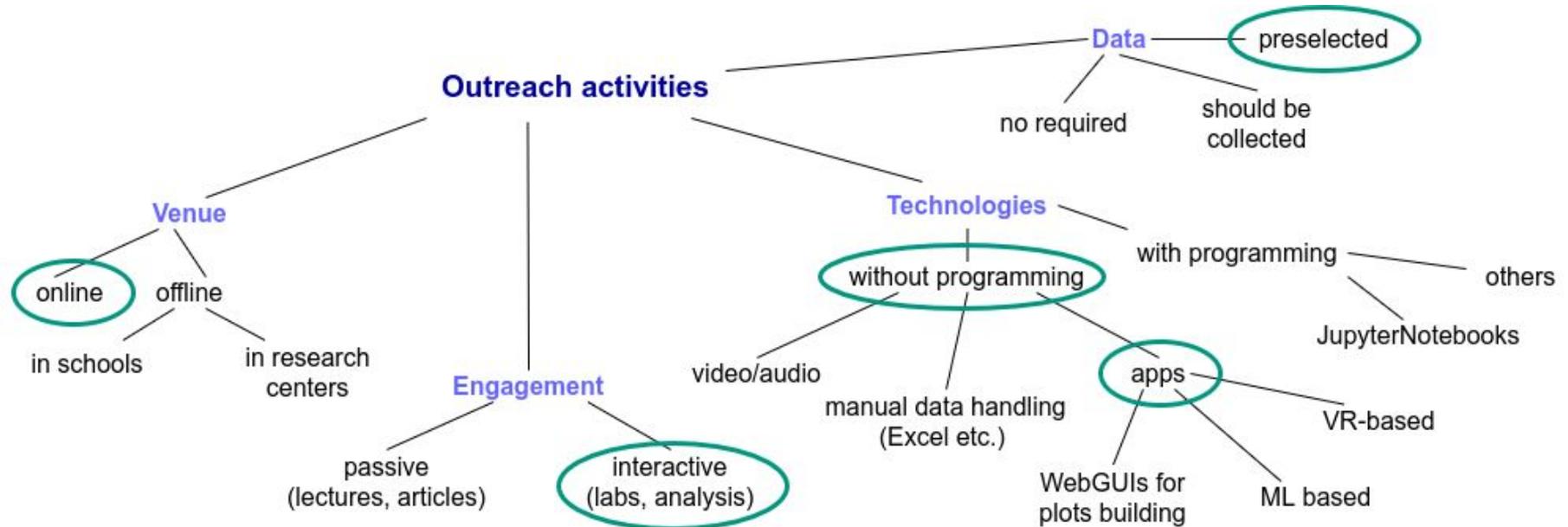
Neural networks for outreach

- Information technology is developing rapidly and finding its application in astroparticle physics
- However, to date, they are seldom used in the development of educational materials
- Online education:
 - Education accessible to everyone
 - Covid-19

Taxonomy of outreach activities in astroparticle physics



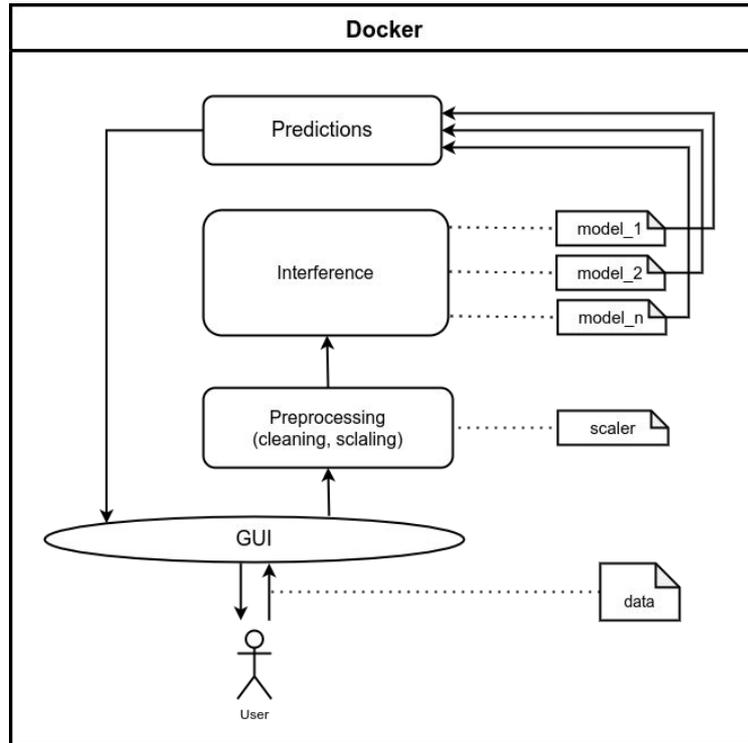
Taxonomy of outreach activities in astroparticle physics



Interactive data analysis applications

- VR based applications by IceCube
- Cazadores de rayos gamma (gamified JupyterNotebooks) by MAGIC
- Cosmic@Web by DESY (earlier example - Showers of knowledge)
- *Online tools of the Gravitational Wave Open Science Center:*
 - *Gravity Spy (citizen science)*
 - *Black Hole Hunter*
- *Bayesian Deep Learning for Galaxy Zoo DECaLS*
- *Astroparticle CNN Client for TAIGA by GRADCL Initiative*

KASCADE ML Application scheme



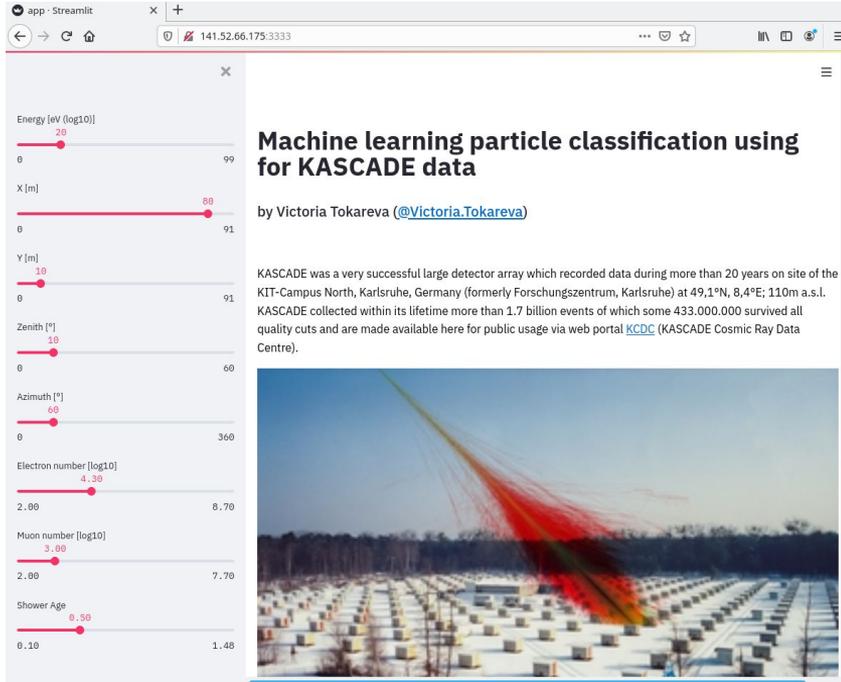
Implementation:

python3.8,
 pandas, numpy, sklearn, streamlit,
 html/css,
 pipenv,
 Docker

Testbed:

Singularity, IAP local network

KASCADE ML Application demo GUI



app : Streamlit

141.52.66.175:3333

Energy [eV (log10)] 29

X [m] 88

Y [m] 10

Zenith [°] 10

Azimuth [°] 60

Electron number [log10] 4.30

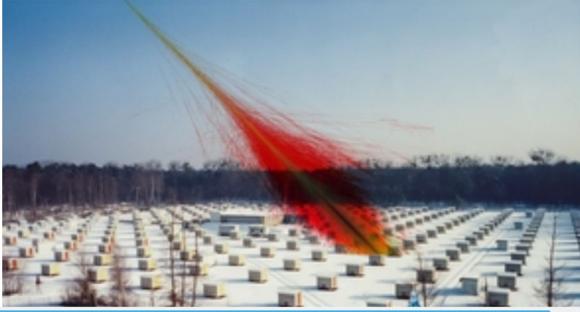
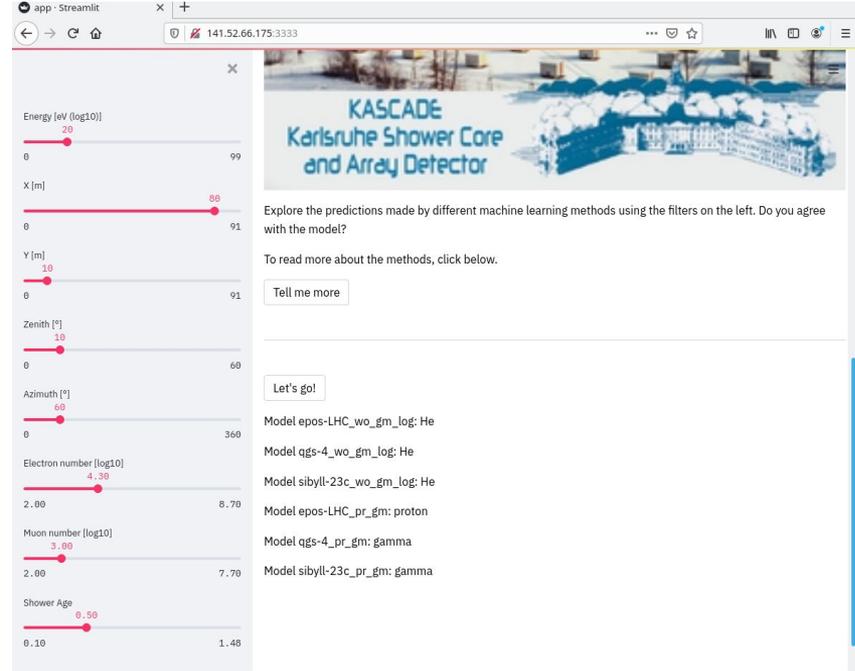
Muon number [log10] 3.80

Shower Age 8.50

Machine learning particle classification using for KASCADE data

by Victoria Tokareva (@Victoria.Tokareva)

KASCADE was a very successful large detector array which recorded data during more than 20 years on site of the KIT-Campus North, Karlsruhe, Germany (formerly Forschungszentrum, Karlsruhe) at 49,1°N, 8,4°E; 110m a.s.l. KASCADE collected within its lifetime more than 1.7 billion events of which some 433.000.000 survived all quality cuts and are made available here for public usage via web portal [KADC](#) (KASCADE Cosmic Ray Data Centre).

app : Streamlit

141.52.66.175:3333

KASCADE Karlsruhe Shower Core and Array Detector

Explore the predictions made by different machine learning methods using the filters on the left. Do you agree with the model?

To read more about the methods, click below.

Model epos-LHC_wo_gm_log: He

Model qgs-4_wo_gm_log: He

Model sibyll-23c_wo_gm_log: He

Model epos-LHC_pr_gm: proton

Model qgs-4_pr_gm: gamma

Model sibyll-23c_pr_gm: gamma

Energy [eV (log10)] 29

X [m] 88

Y [m] 10

Zenith [°] 10

Azimuth [°] 60

Electron number [log10] 4.30

Muon number [log10] 3.80

Shower Age 8.50

Possible improvements

- Add CNN model(s)
- Add gamification
- Improve interface
- Allow users to load their own ML models
- Publish outside of KIT inner network

In this talk we:

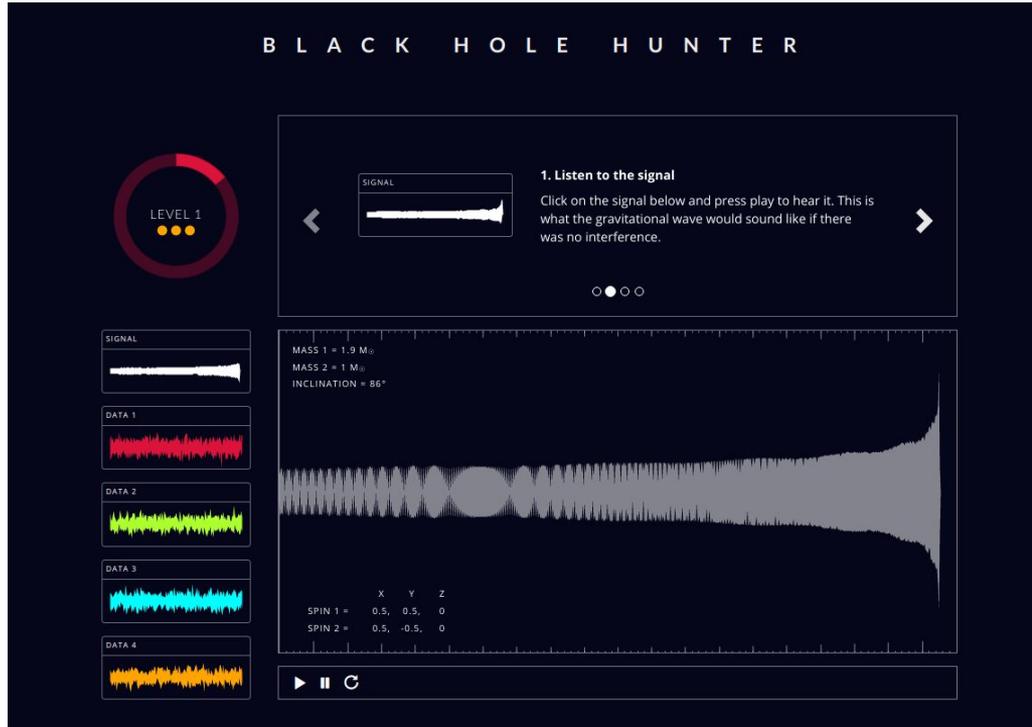
- Gave an overview of machine learning workflow and discussed it using machine learning activities for KASCADE experiment as an example
- Discussed usage of machine learning and other topical IT technologies for outreach in astroparticle physics
- Introduced a concept of ML-based outreach application based on KASCADE data
- Shared the application schema, interface and details of the implementation
- Shown how to deploy machine learning models in a way one can use them for educational purposes

Thank you for your attention!

Questions? Contact: victoria.tokareva@kit.edu

Backup slides

Example: black holes hunters



BLACK HOLE HUNTER

LEVEL 1

SIGNAL

1. Listen to the signal

Click on the signal below and press play to hear it. This is what the gravitational wave would sound like if there was no interference.

DATA 1

DATA 2

DATA 3

DATA 4

MASS 1 = 1.9 M_{\odot}
MASS 2 = 1 M_{\odot}
INCLINATION = 86°

	X	Y	Z
SPIN 1 =	0.5,	0.5,	0
SPIN 2 =	0.5,	-0.5,	0

Successful gamification of analysis task through:

- Attractive name and interface
- Levels, lives and points
- clear game scenario

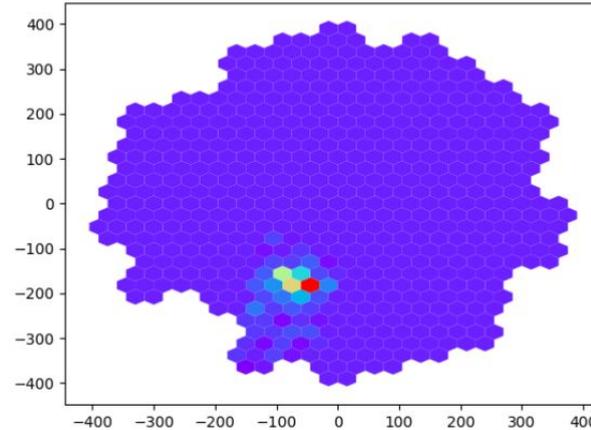
Example: TAIGA CNN client for astropartilce.online

Let's check yourself and determine which type of the particle is it?

5/50

Score: right 1, missed 2, total 2

Your previous answer is correct
This is a gamma with probability: 1.00



Back



Proton



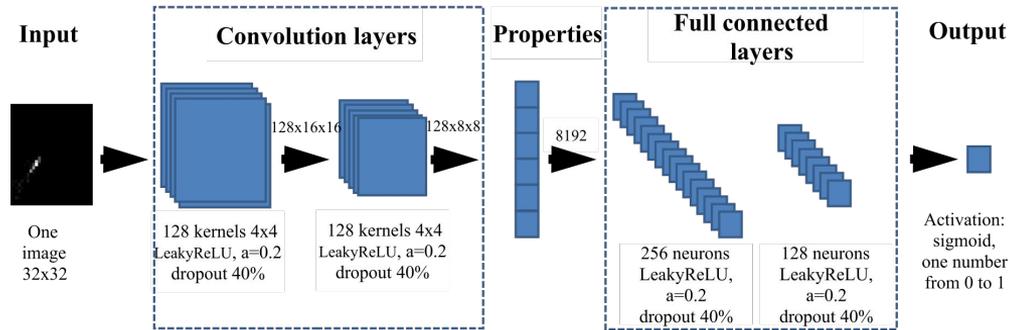
Gamma

Next

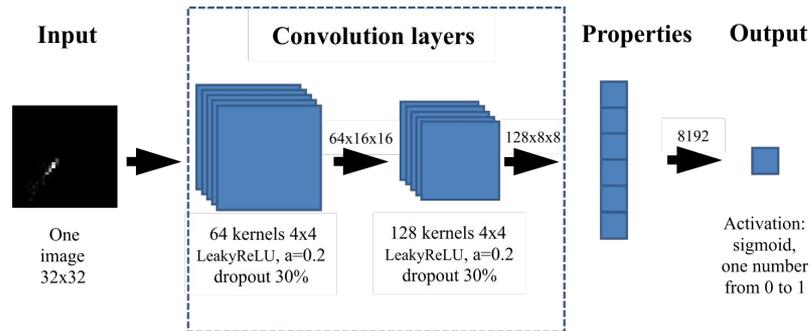
Finish

Hint: In comparison with hadron showers gamma-ray ones have more elliptic shape, less width and major axis pointed to the source. Axes of gamma-induced images are pointed to the center of the camera because the telescope is pointed at a known source that is placed in the center of the telescope's camera.

Example: TAIGA CNN client for astropartilce.online - 1



Proton discriminator architecture



Gamma discriminator architecture