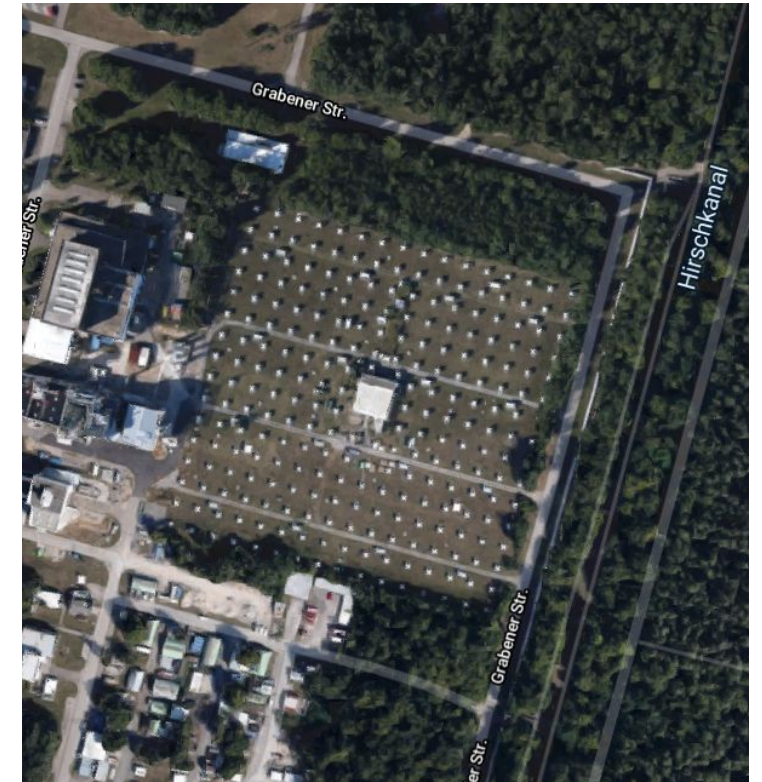
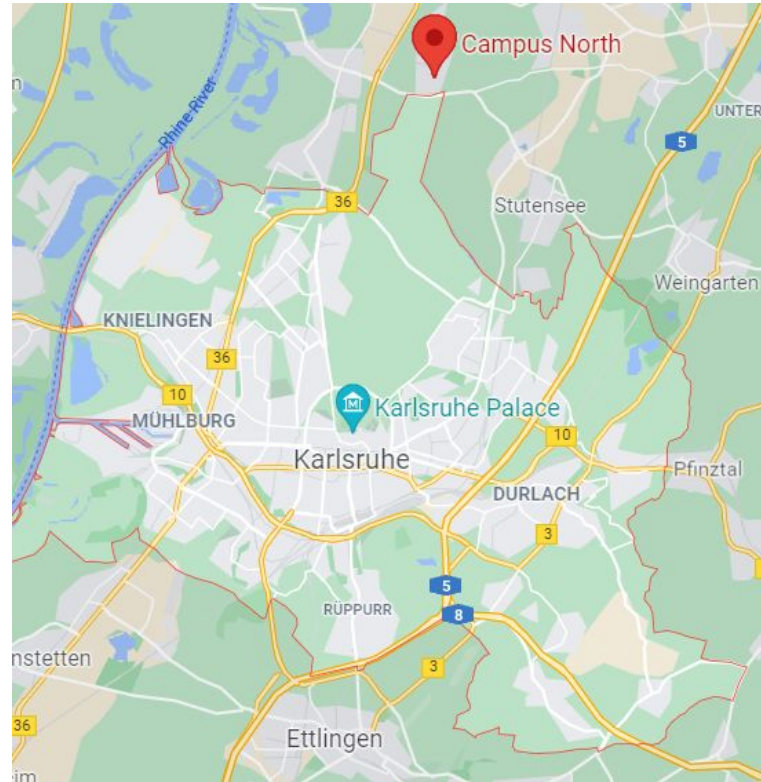


Towards mass composition study with KASCADE using deep learning

Speaker **Daniil Reutsky**

Co-authors	P. Bezyazeev	I. Plokhikh	V. Sotnikov
	S. Golovachev	N. Petrov	V. Tokareva
	D. Kostunin	M. Tsobenko	O. Shchegolev
	V. Lenok		



KASCADE (KArllsruhe Shower Core and Array DEtector)

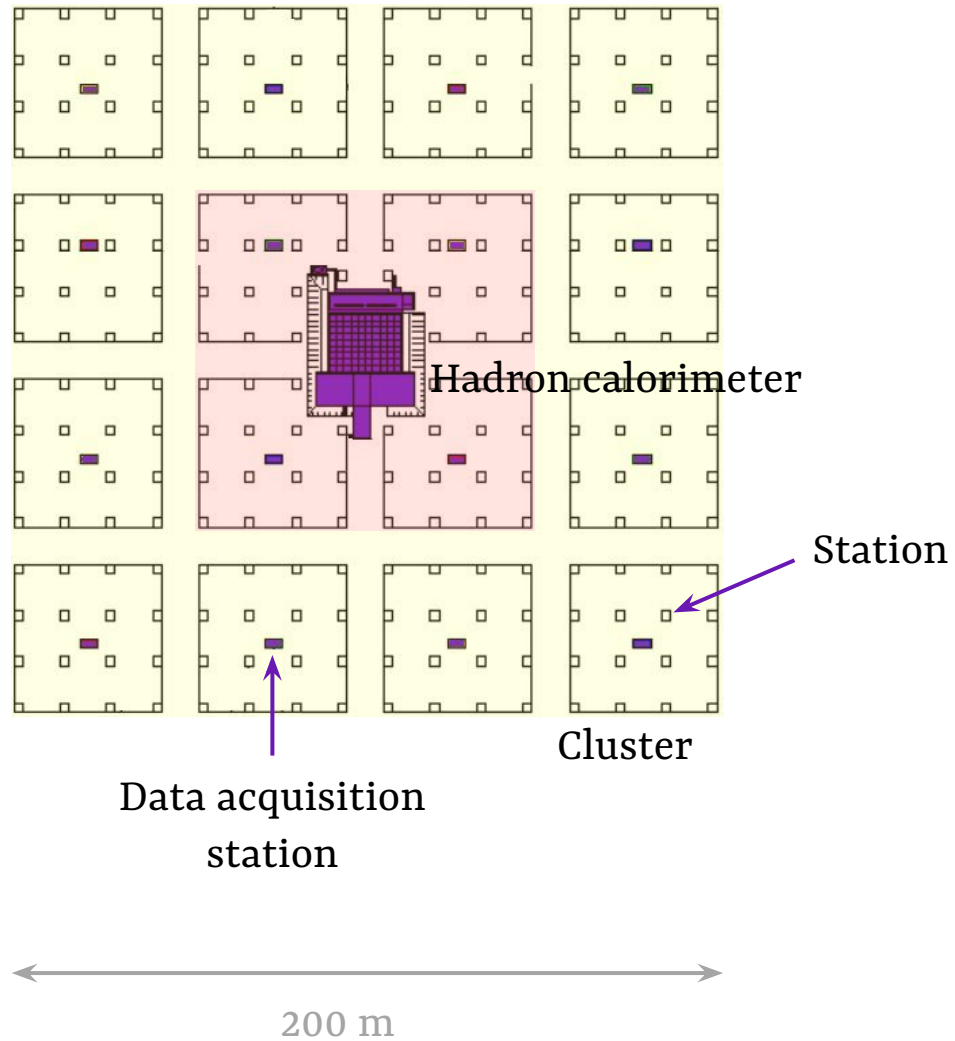
Location Karlsruhe Institut für
Technologie

Years in operation 1996...2013

Events recorded 433 M

A 16×16 grid of scintillating detectors
measuring e/γ , μ energies

Schematic view



Type-1 stations

detect **e/γ** and **muon** signals

Type-2 stations


detect **only e/γ** signals

Event is recorded when ≥ 1 cluster detects a signal $>$ certain threshold

Run is a group of events


Online data center

kcdc.iap.kit.edu



KIT
Karlsruhe Institute of Technology

KIT | IAP | HOME | Data Privacy | Impressum | admin | login



KASCADE Cosmic Ray Data Centre (KCDC) / Open β

KCDC Homepage

KCDC Motivation

KCDC Regulations

► Information

► Announcements

FAQs

► User Account

► Data Shops

► Simulations

Spectra

Materials

► Publications

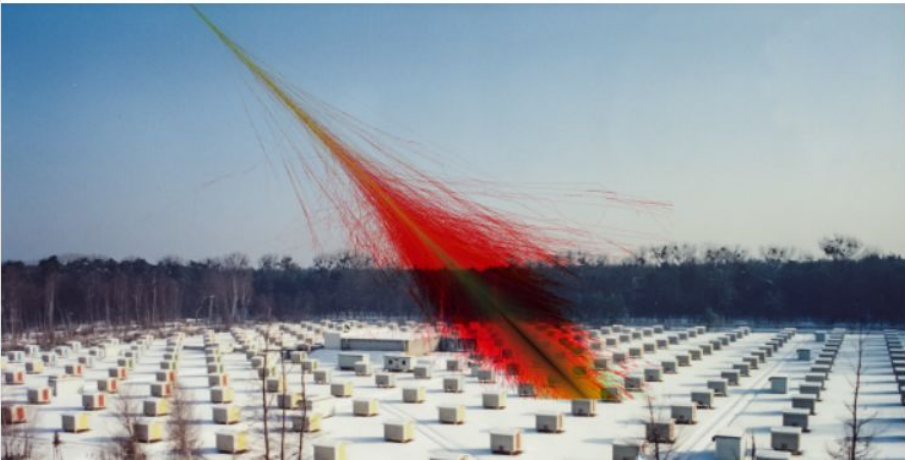
► Report a Bug

► Education/Lehre

KCDC Partners

Welcome to kcdc

The aim of the project **KCDC** (**KASCADE Cosmic Ray Data Centre**) is the installation and establishment of a public data centre for high-energy astroparticle physics based on the data of the KASCADE experiment. 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 three DataShops. The first, called 'KASCADE', was introduced in 2013 and has been continuously extended since then. It contains data from four KASCADE-Grande detectors, which were analysed separately and from the LOPES radio antennas. The second, called COMBINED, is based on a subsample of the KASCADE-Grande data, but was evaluated in a joint analysis of the KASCADE and GRANDE detector arrays. Newly added in 2021 was the MAKET-ANI DataShop, which contains data from the MAKET-ANI air shower experiment at Mount Aragats, Armenia.



Institute for Astroparticle Physics
(IAP)
KIT Campus North

Address:
Institute for Astroparticle Physics
Karlsruhe Institute of Technology
Hermann-v.Helmholtz-Platz 1
D-76344 Eggenstein-
Leopoldshafen

Postal Address:
Institute for Astroparticle Physics
Karlsruhe Institute of Technology
Postbox 3640
D-76021 Karlsruhe

Phone: +49/721/608-23546
Fax: +49/721/608-23548

E-Mail:
► [iap-kcdc\[at\]lists.kit.edu](mailto:iap-kcdc[at]lists.kit.edu)

DOI:
► doi.org/10.17616/R3TS4P

Downloads
► [KCDC Materials](#)

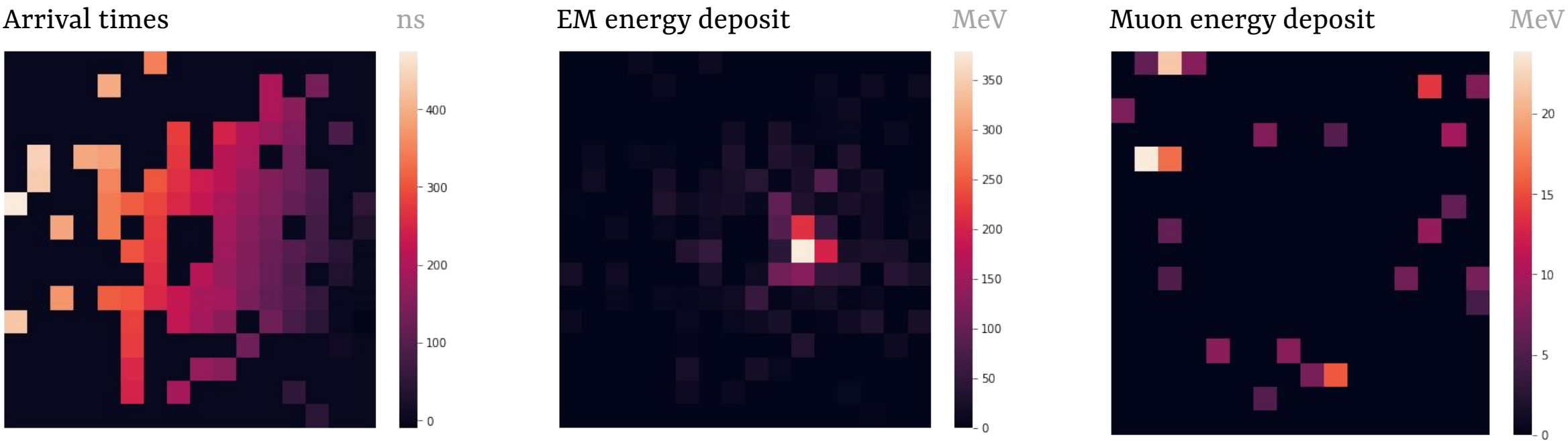
A.Haungs et al; *Eur. Phys. J. C* (2018) 78:741;

“The KASCADE Cosmic ray Data Centre KCDC: granting open access to astroparticle physics research data”;

(doi: [10.1140/epjc/s10052-018-6221-2](https://doi.org/10.1140/epjc/s10052-018-6221-2))

Event example

Experimental features



Reconstructed features

Primary particle lg E	Shower core center (x, y)		Arrival direction (zenith, azimuth)		lg N _e	lg N _μ	Age	
15,31	34,66 m	-3,11 m	42,64°	120,9°	4,44	4,06	1,15	34,70 m

Approach

Input: Event

3×16×16 experimental features

9 reconstructed features



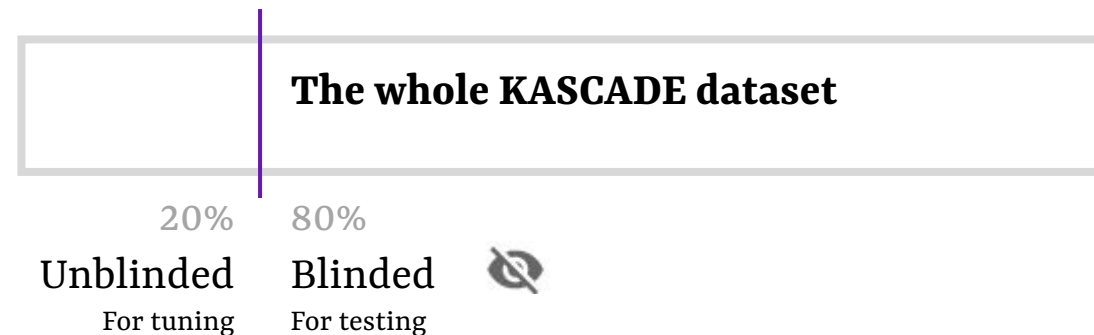
Target: Primary particle type

Categorical feature (p, He, C, Si, Fe)

Some of our models

- Random Forest classifier (baseline)
- CNN classifier
 - Validation on full unblinded dataset
 - Validation with data quality cuts

Semi-blind analysis



Training step

CORSIKA simulations (in this presentation EPOS-LHC only)

Validation step

Checking out predicted particles spectra. Testing hypotheses

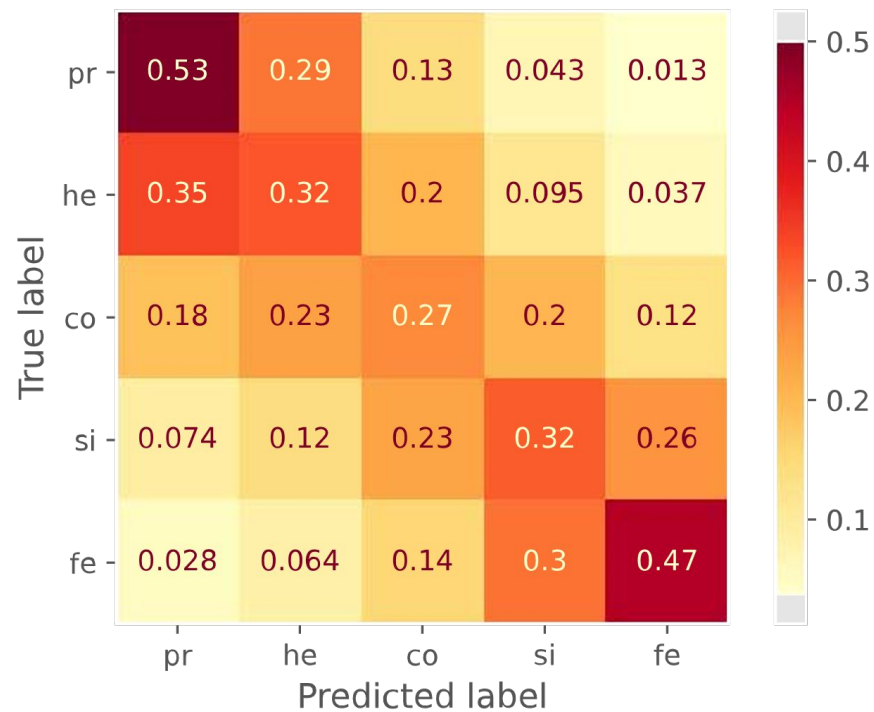
Testing step

Revealing the blinded part

Random Forest classifier performance

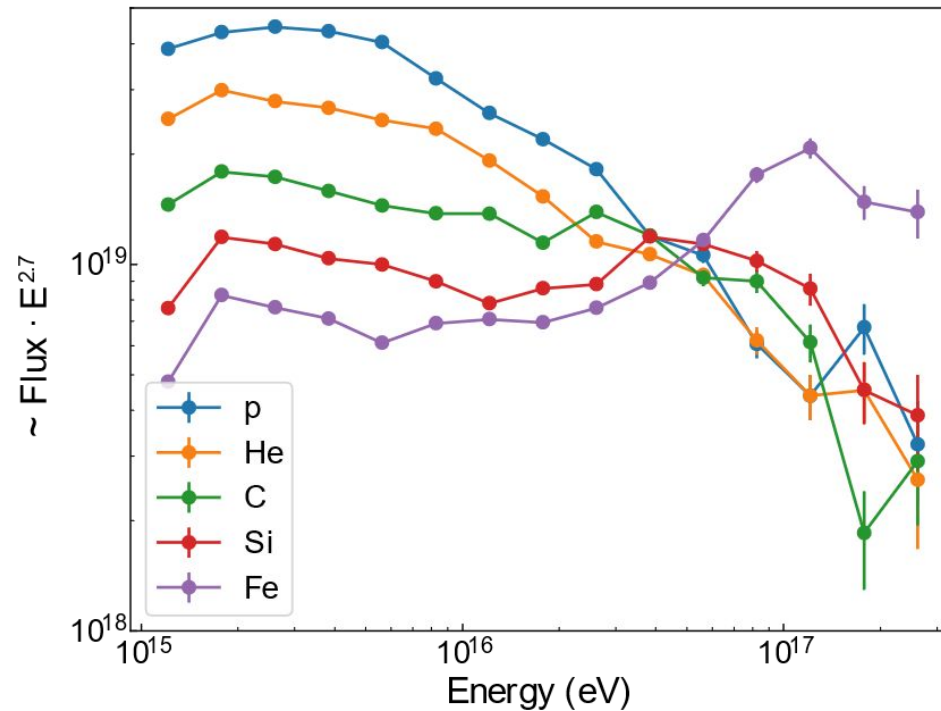
Confusion matrix

Simulated data (EPOS-LHC)

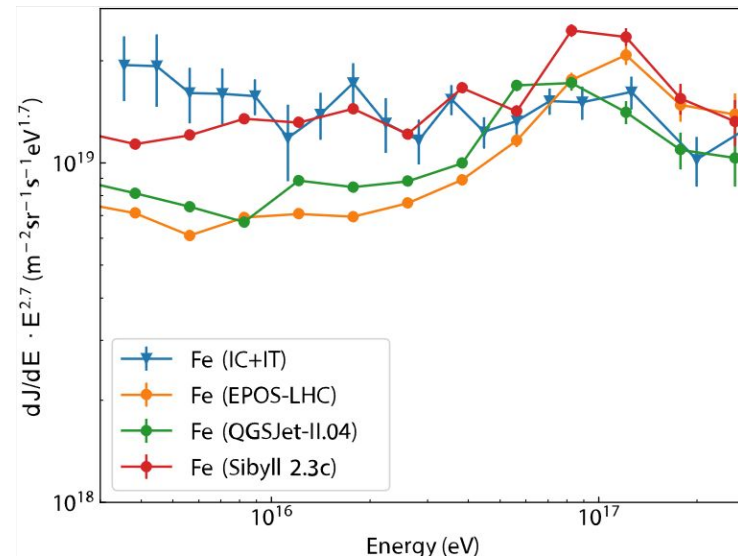
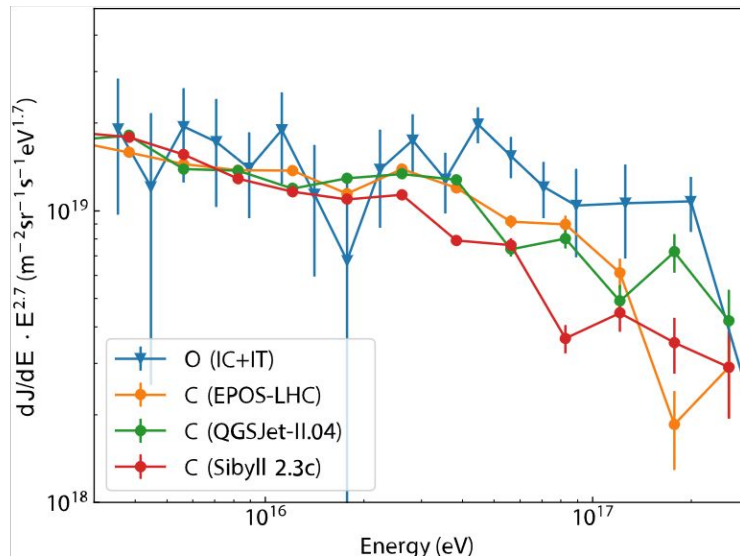
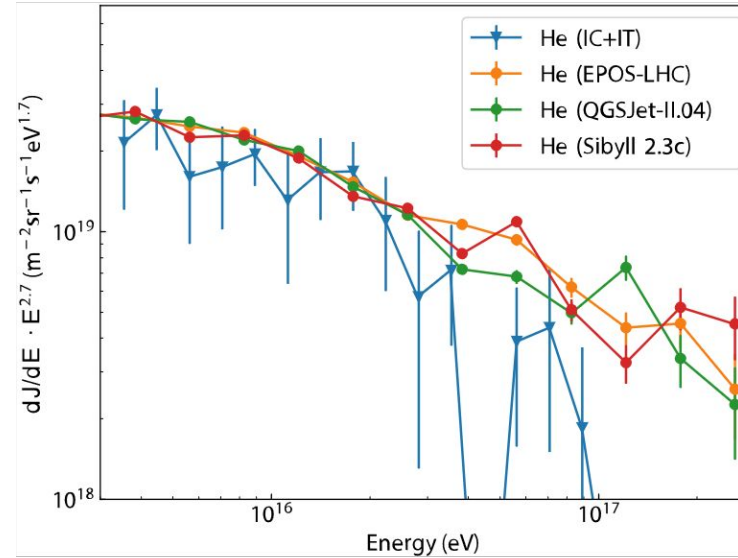
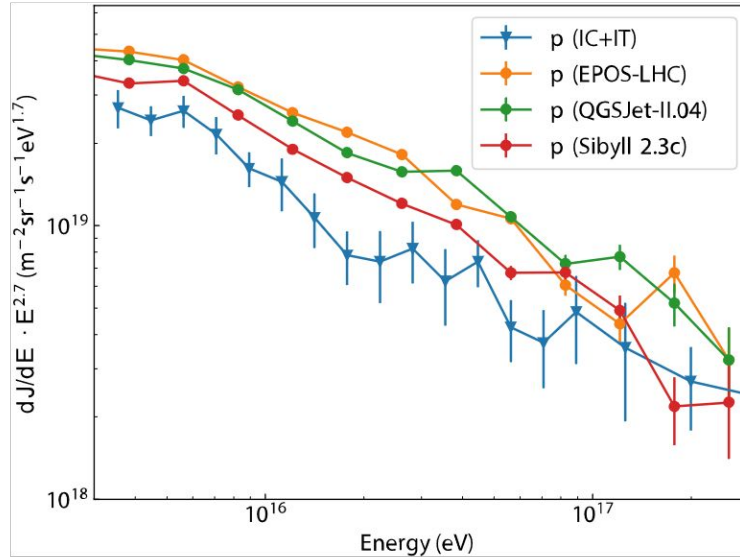


Spectra

Experimental data



Random Forest compared with IceCube collaboration



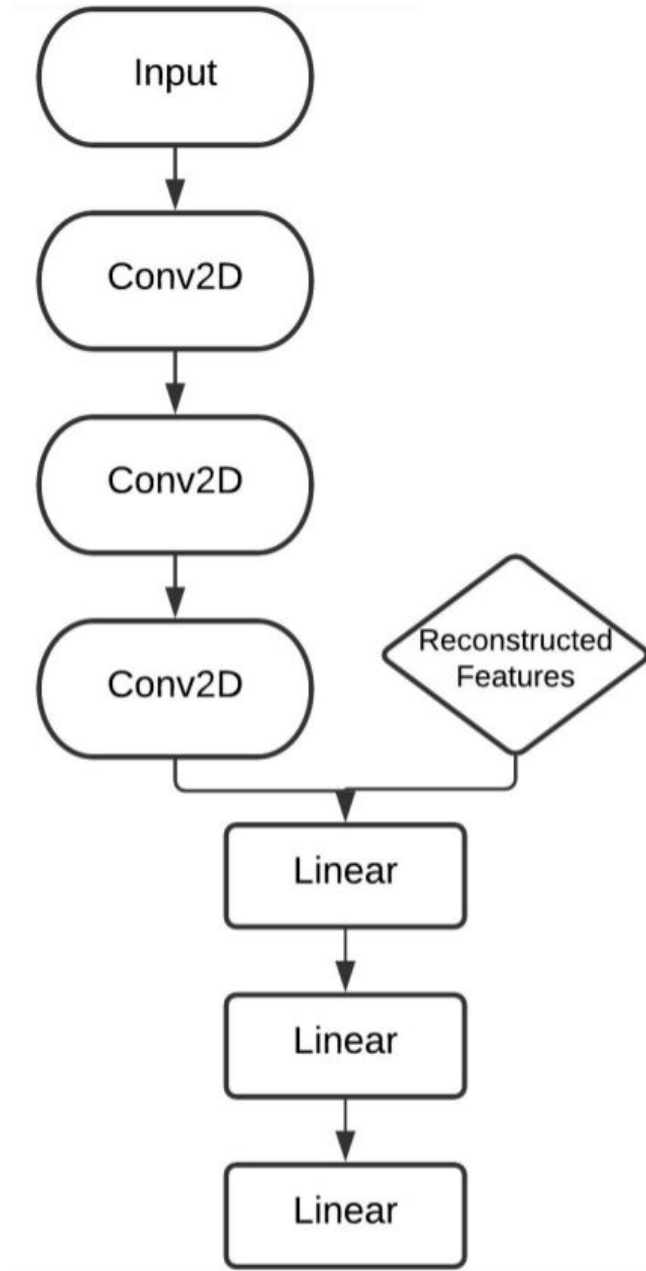
IceCube+IceTop reconstruction

Phys. Rev. D 100, 082002 (2019)

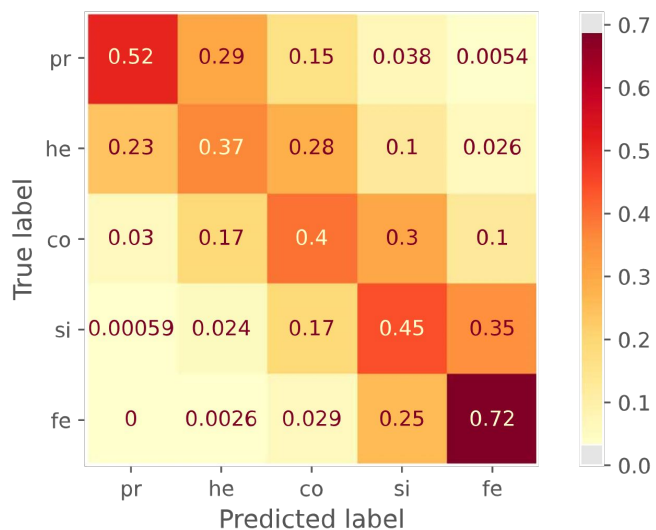
- Sibyll 2.1 hadronic model
- 4 mass groups
- ML approach
- Same energy range

Comparison published in
PoS ICRC2021 (2021) 319

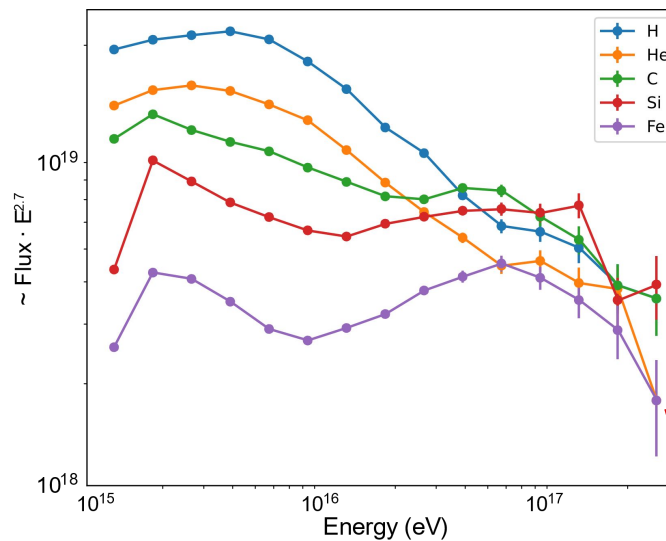
CNN architecture



CNN performance and comparison to Random Forest

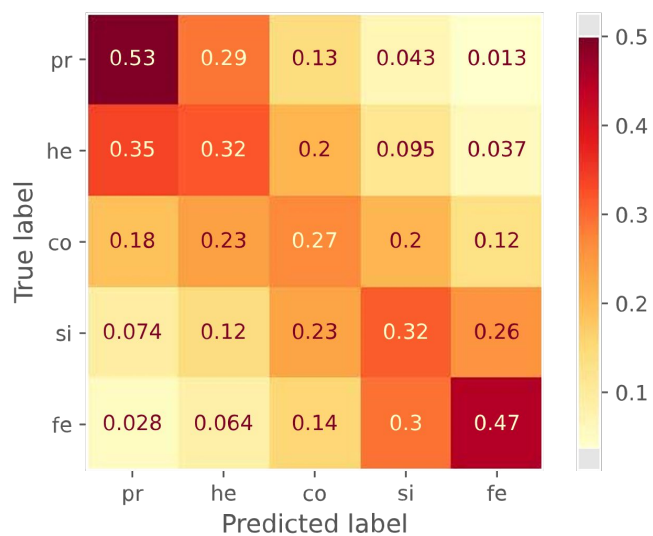


Confusion matrix for
CNN

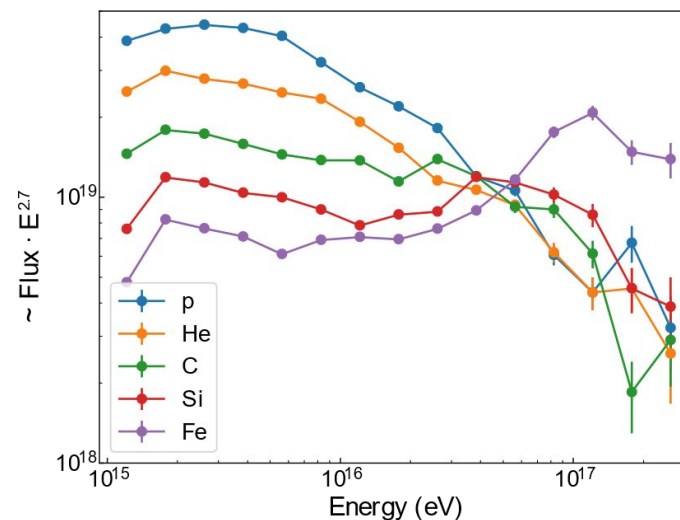


Spectra for CNN

EXEMPLARY SPECTRA
NOT FOR SCIENTIFIC USE



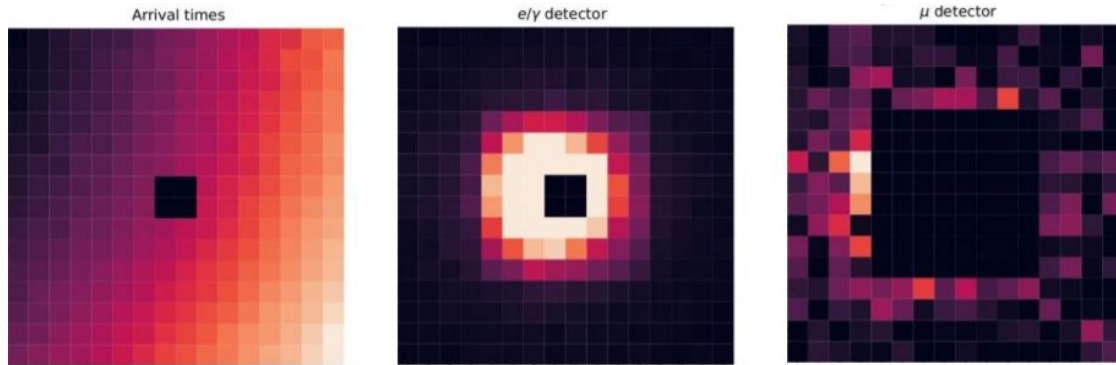
Confusion matrix for
Random Forest



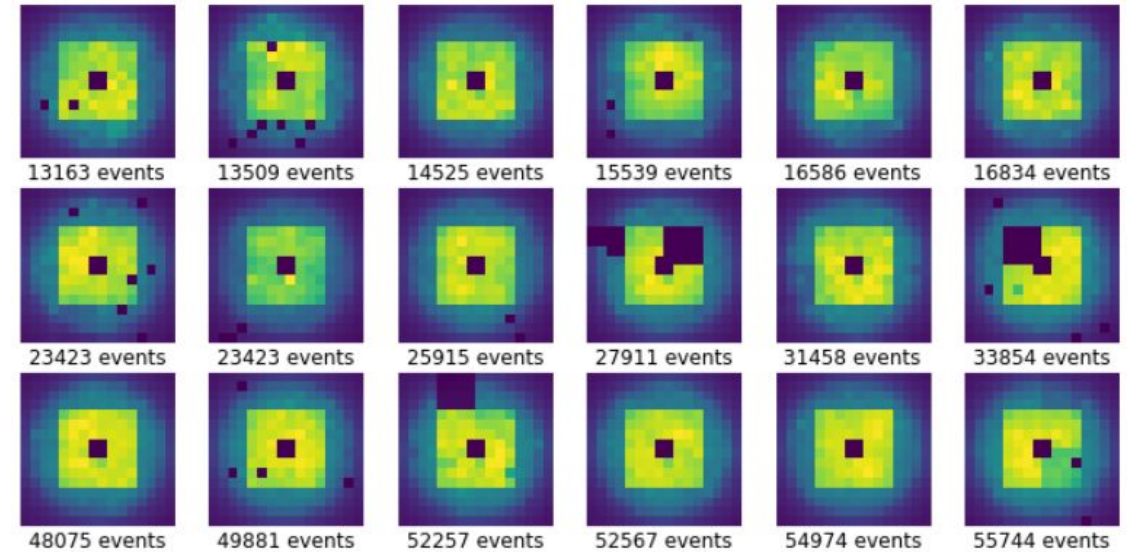
Spectra for Random Forest

Motivation behind quality cuts

In a simulated event
all detectors have 100% uptime



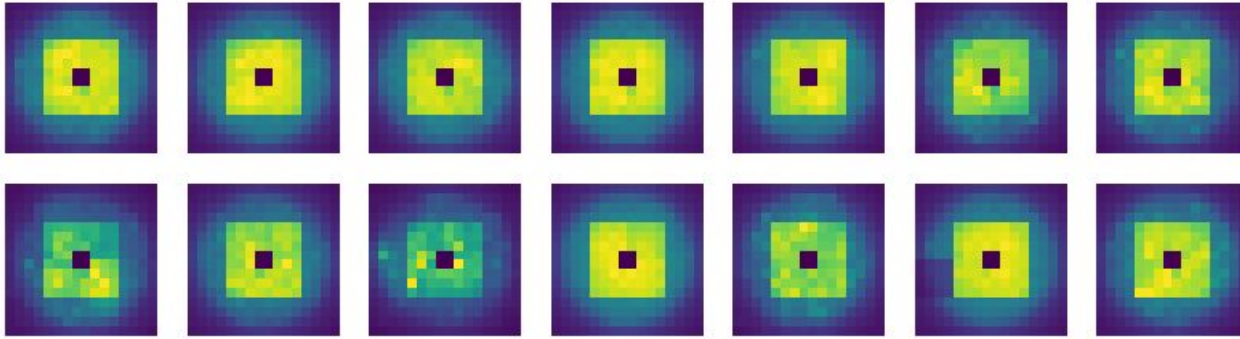
In a real event
some detectors might go down



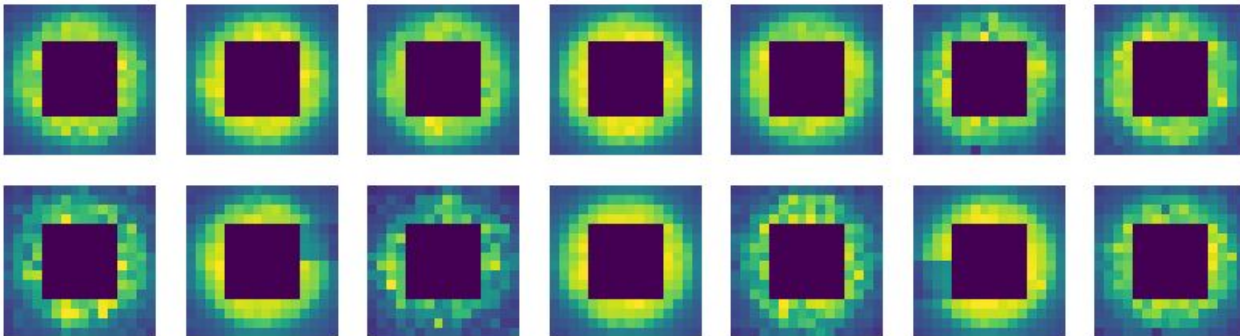
*Each square shows sum of EM energy deposits
for some run*

Solution: drop out unsatisfactory runs

Some runs which passed our cuts

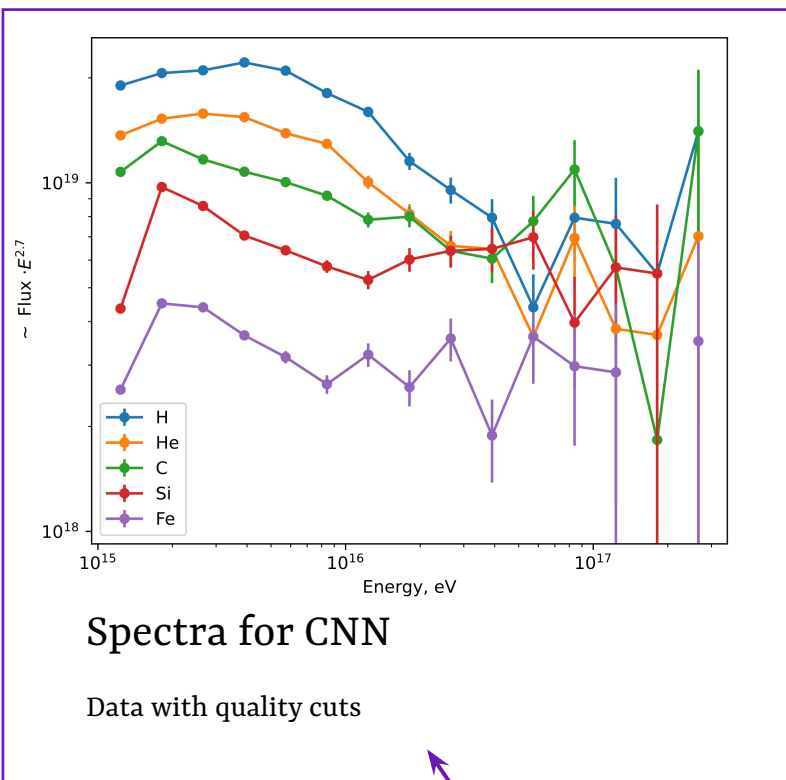


*Each square shows sum of EM energy deposits
for some run*

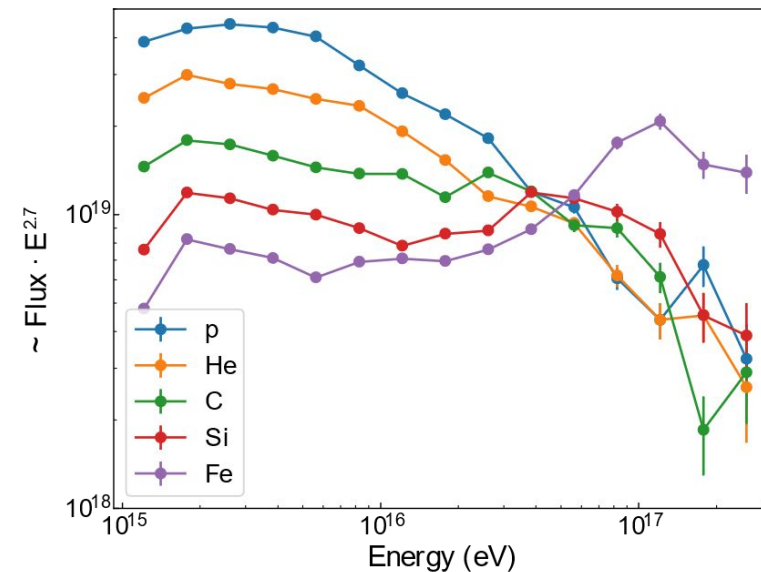
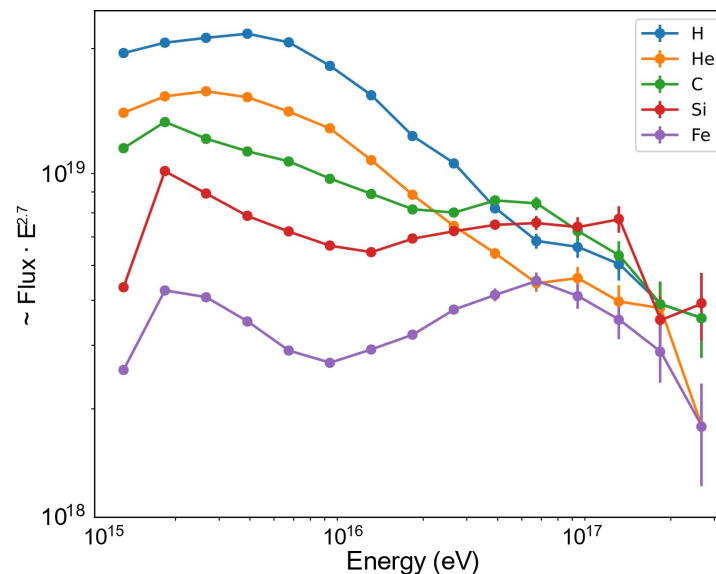


Same for muons energy

CNN performance with quality cuts



1/3 of the full dataset



EXEMPLARY SPECTRA, NOT FOR SCIENTIFIC USE

Conclusion

- CNN is a promising architecture for KASCADE mass composition analysis
- CNN is sensitive to irregularities in the data as expected
- Application of quality cuts is important for results
- Sophisticated cuts are required to keep sufficient amount of data
- Mass composition study with CNN serves as validation for gamma/hadron classifier
(see talk by M. Tsobenko).