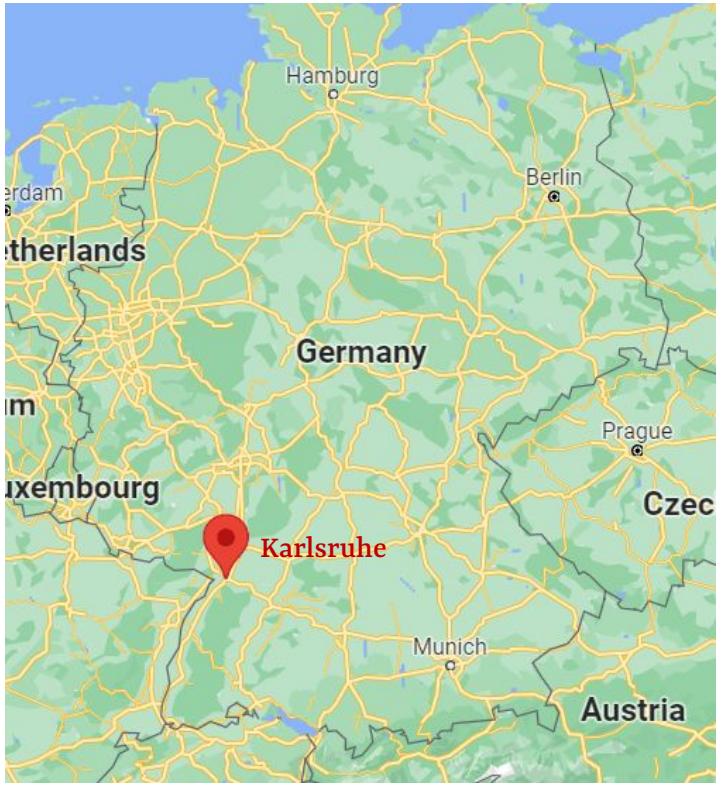
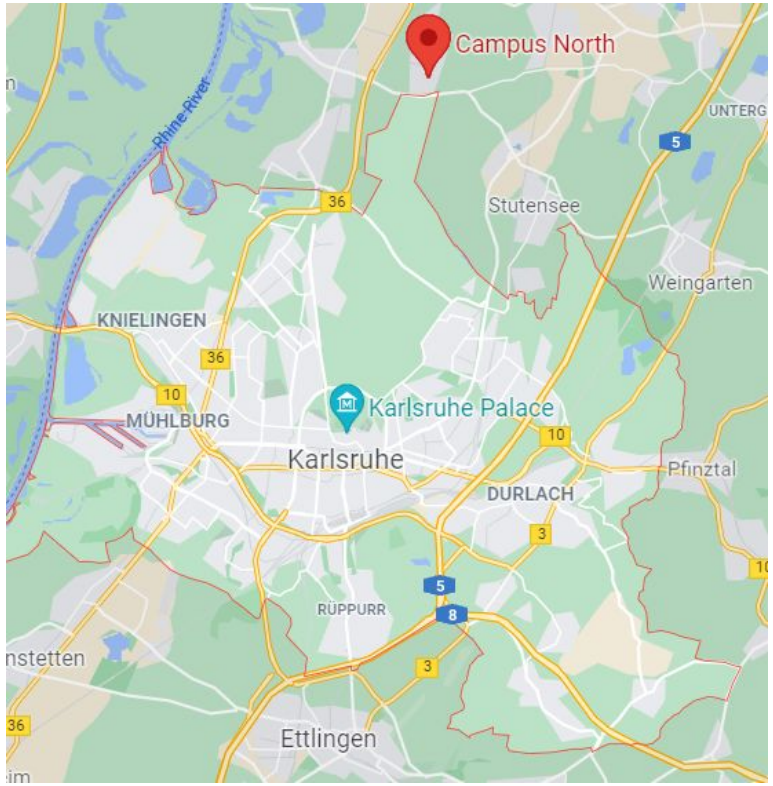
Towards mass composition study with KASCADE using deep learning

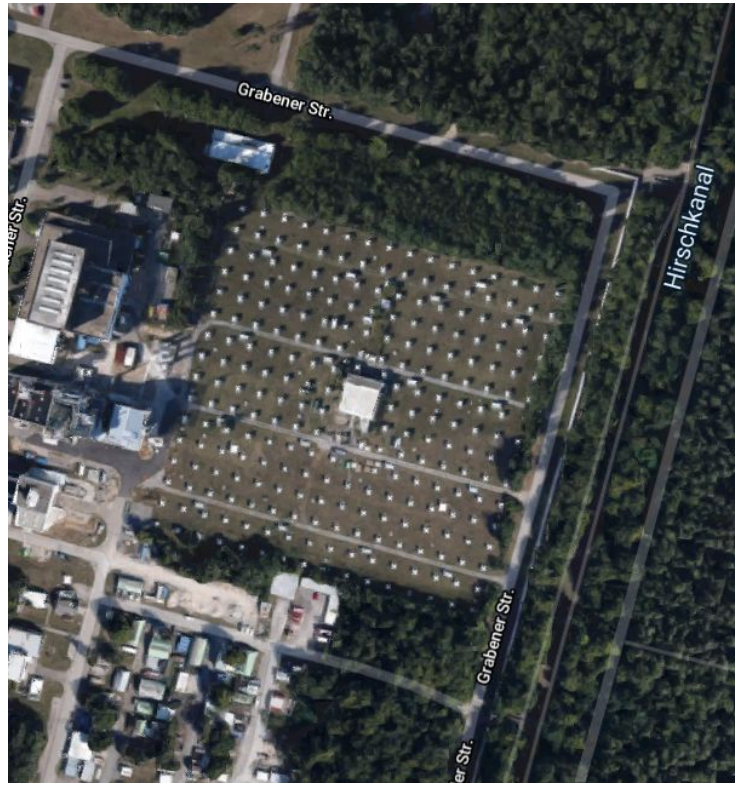
Speaker **Daniil Reutsky**

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

Supported by the Russian Science Foundation (RSF)





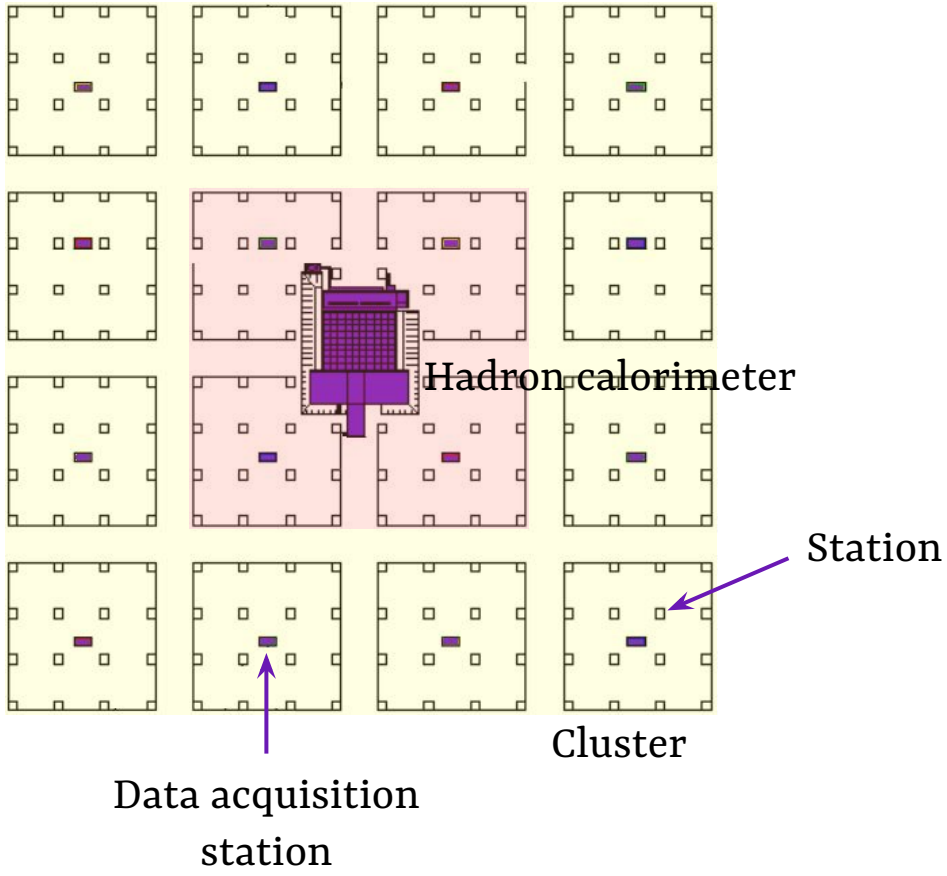


KASCADE (**KA**rlsruhe **S**hower **C**ore and **A**rray **DE**tector)

Location	Karlsruher 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

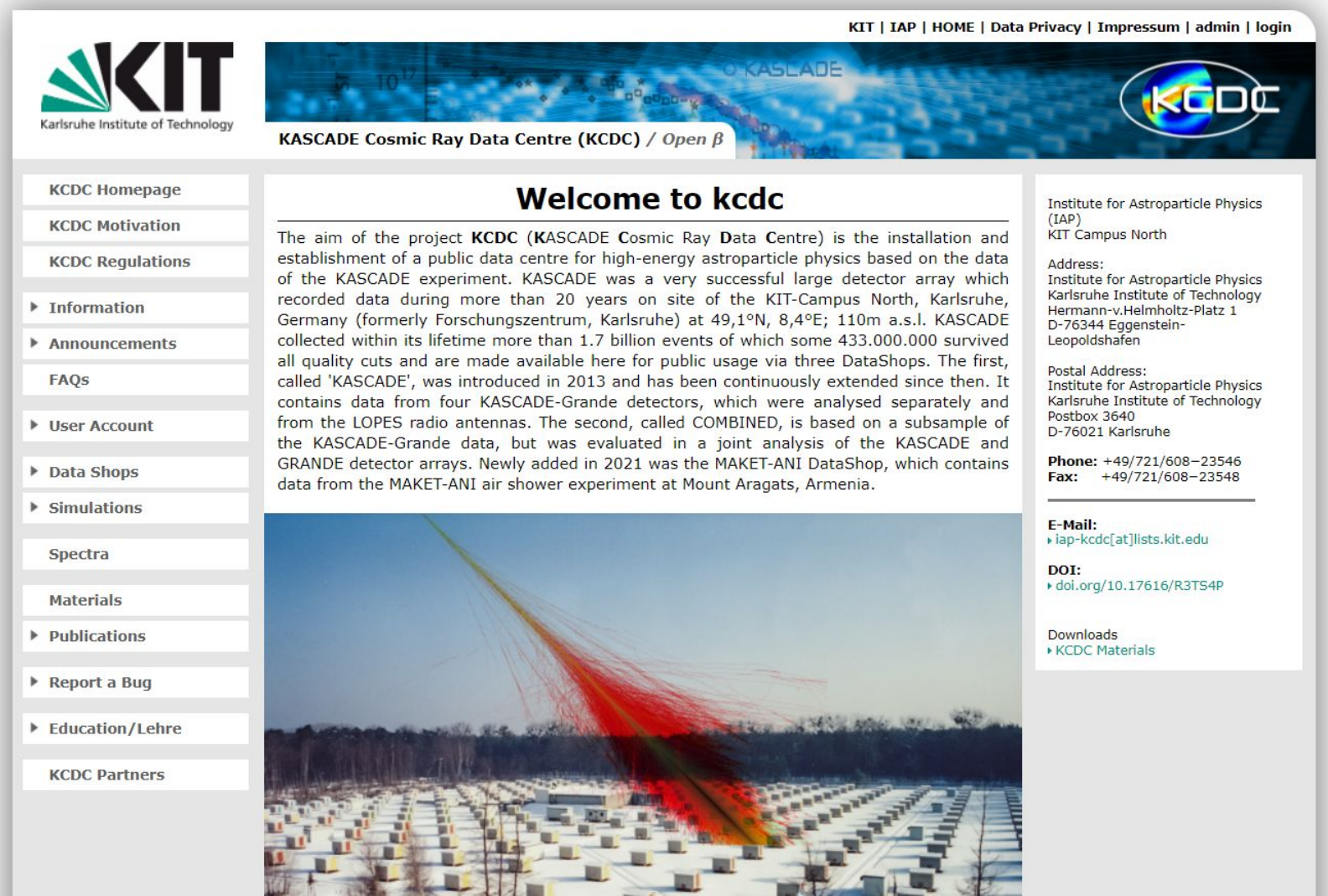
Type-2 stations
detect **only e/γ** signals

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

Run is a group of events

Online data center

kcdc.iap.kit.edu



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

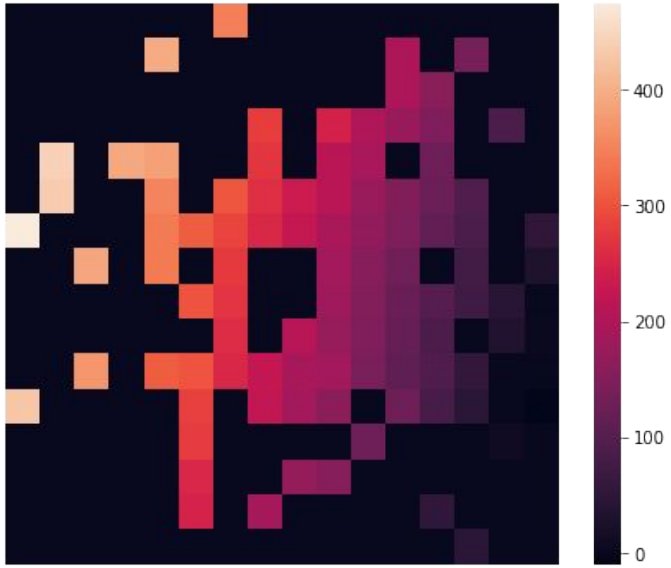
"The KASCADE Cosmic ray Data Centre KCDC: granting open access to astroparticle physics research data";

(doi: 10.1140/epjc/s10052-018-6221-2)

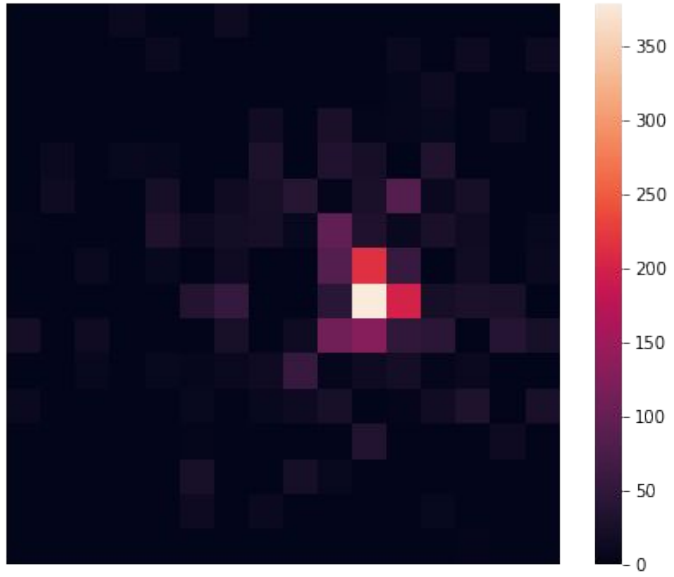
Event example

Experimental features

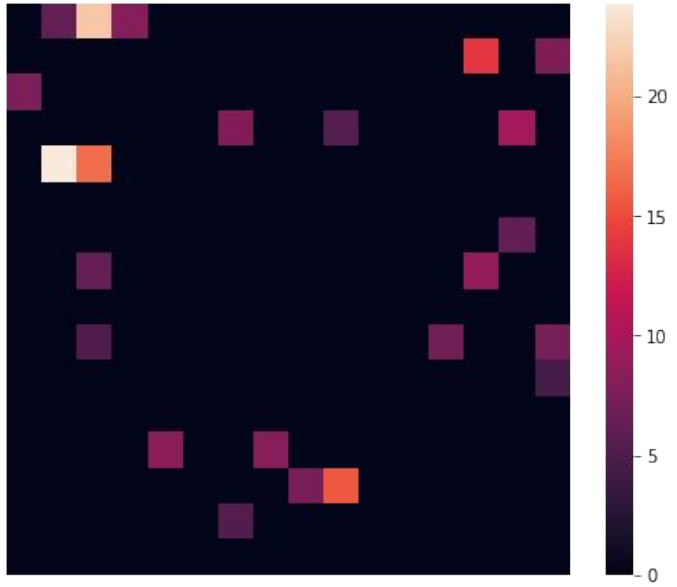
Arrival times ns



EM energy deposit MeV



Muon energy deposit MeV



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

Predict

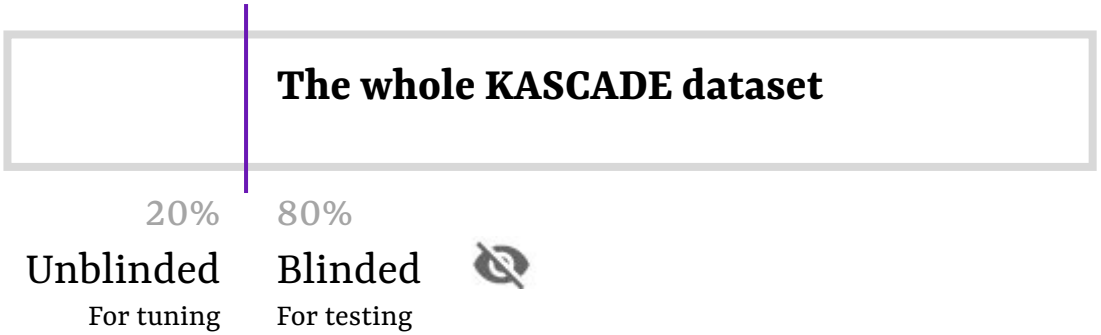
Target: Primary particle type

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

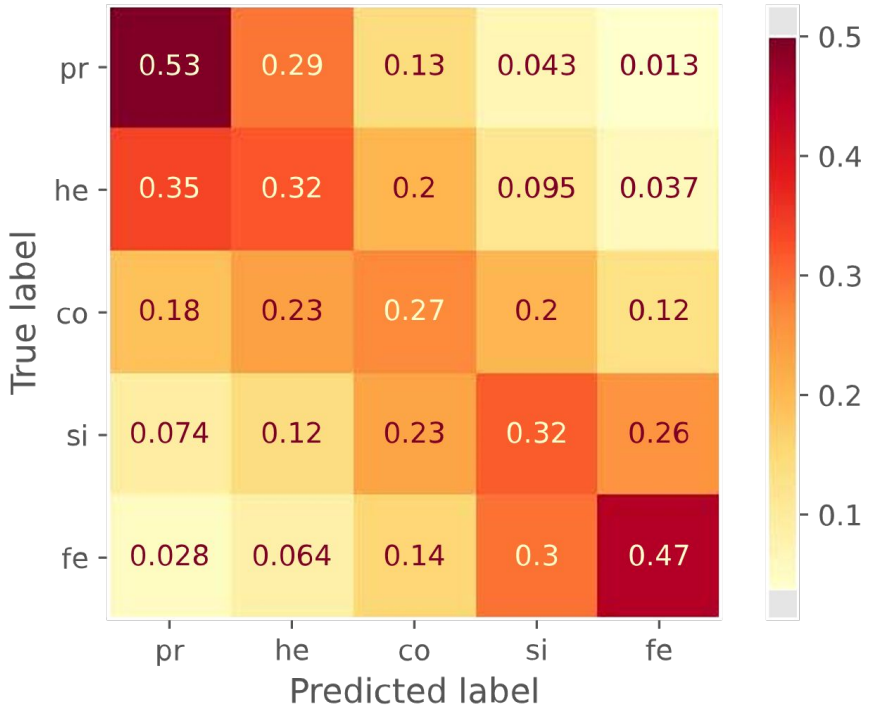


Training step	CORSIKA simulations (in this presentation EPOS-LHC only)
Validation step	Checking out predicted particles spectra. Testing hypotheses
Testing step	Revealing the blinded part

Random Forest classifier performance

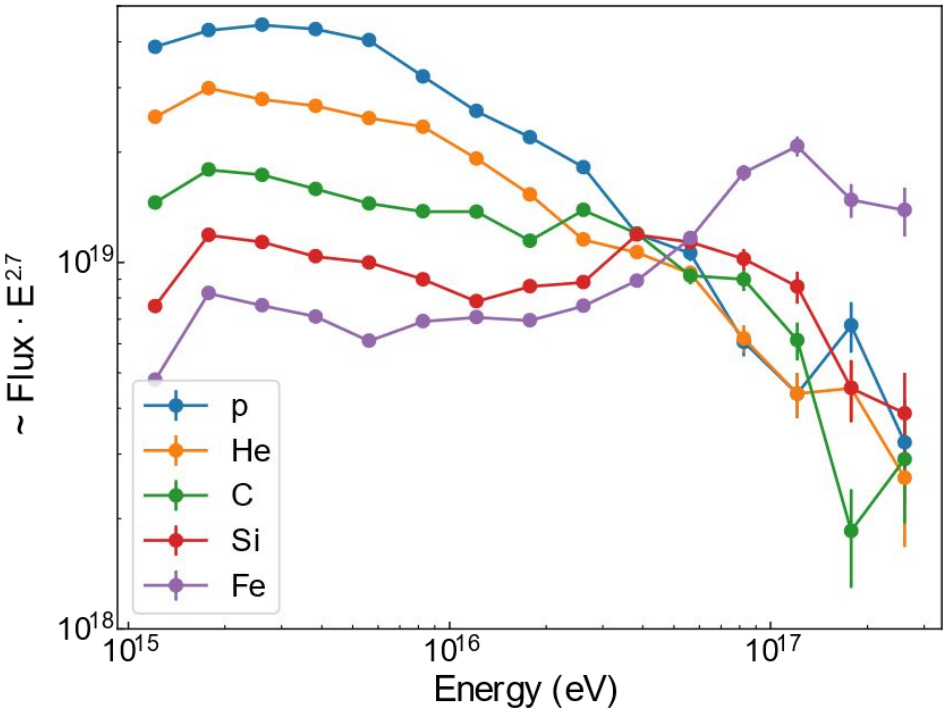
Confusion matrix

Simulated data (EPOS-LHC)



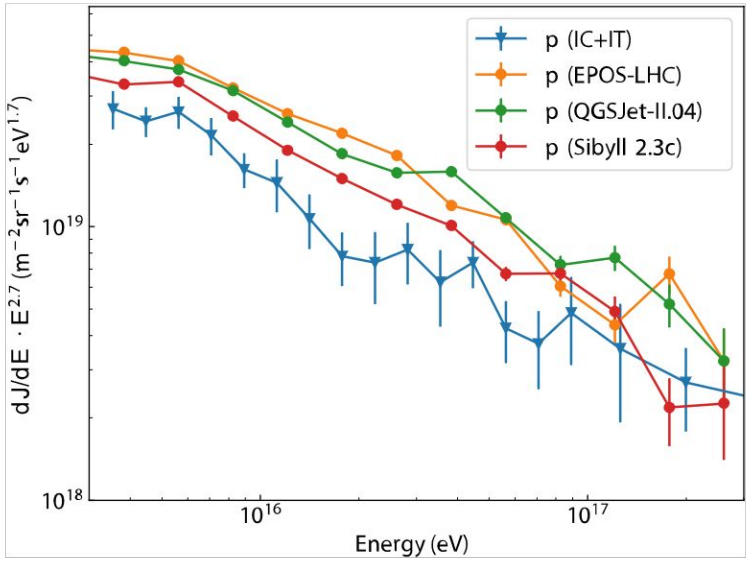
Spectra

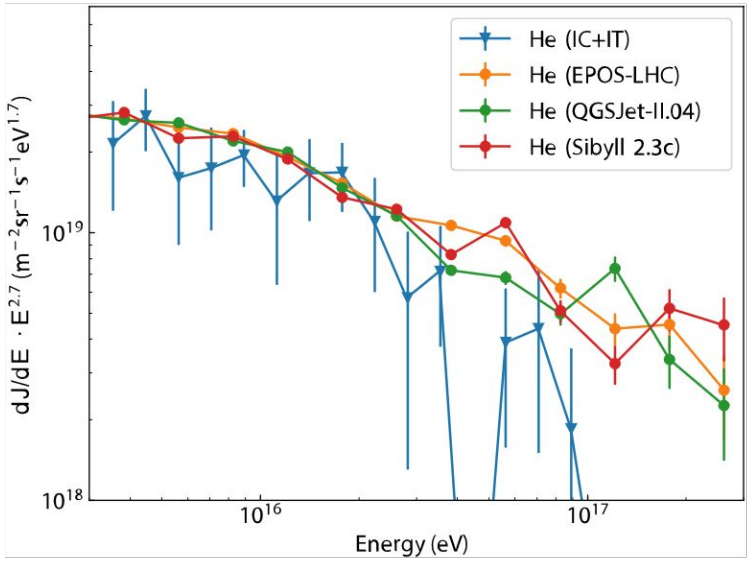
Experimental data

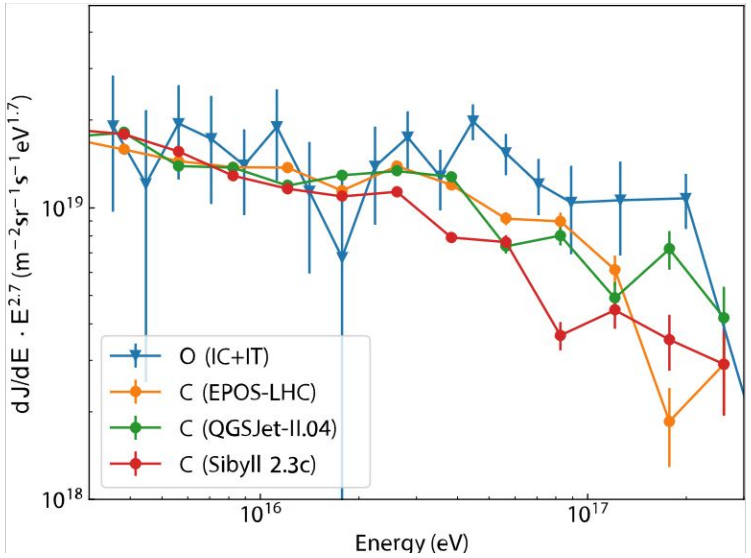


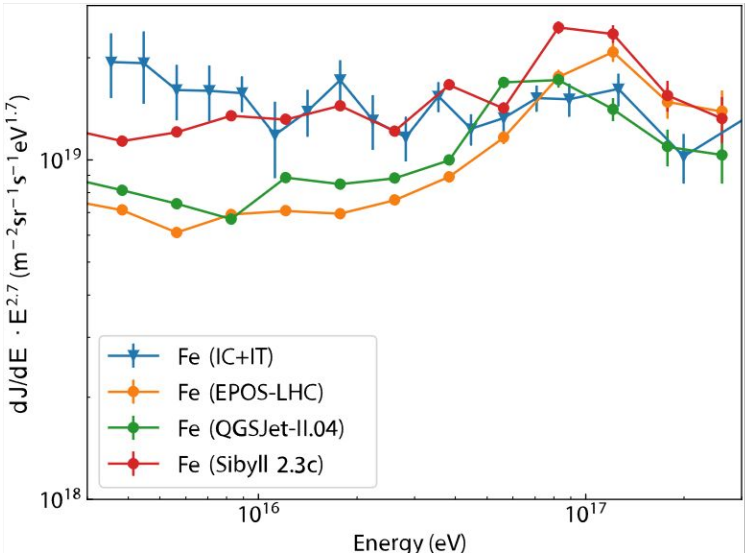
PoS ICRC2021 (2021) 319

Random Forest compared with IceCube collaboration







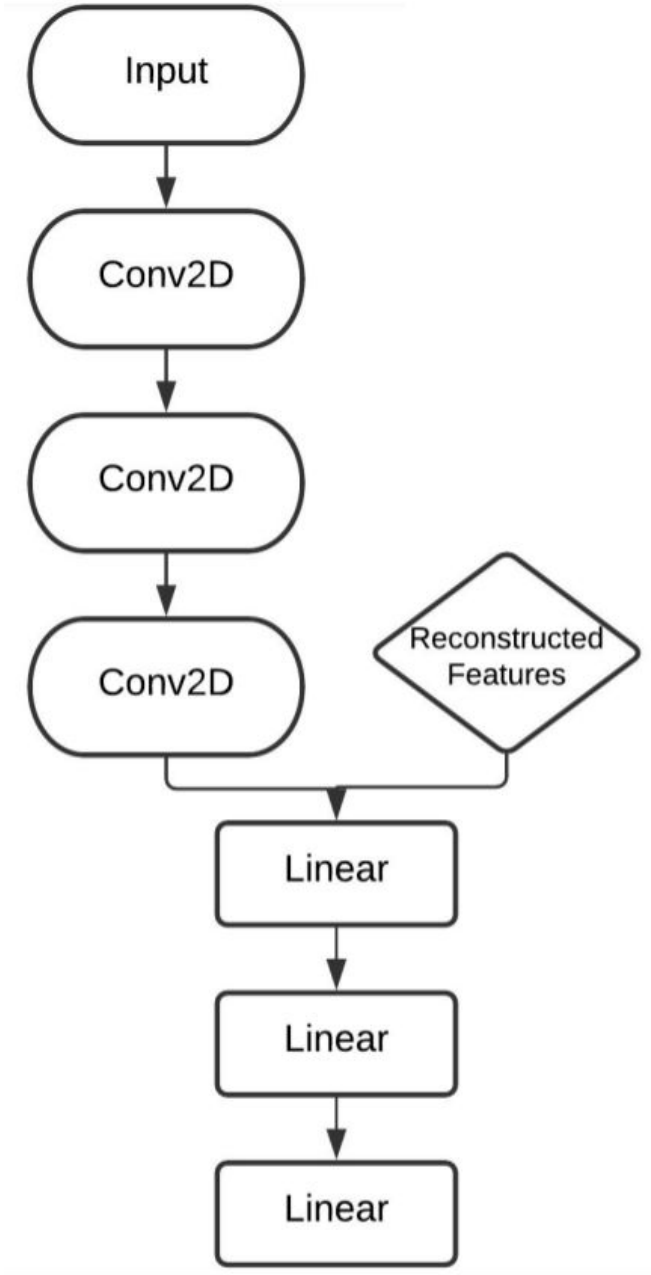


IceCube+IceTop reconstruction
Phys. Rev. D 100, 082002 (2019)

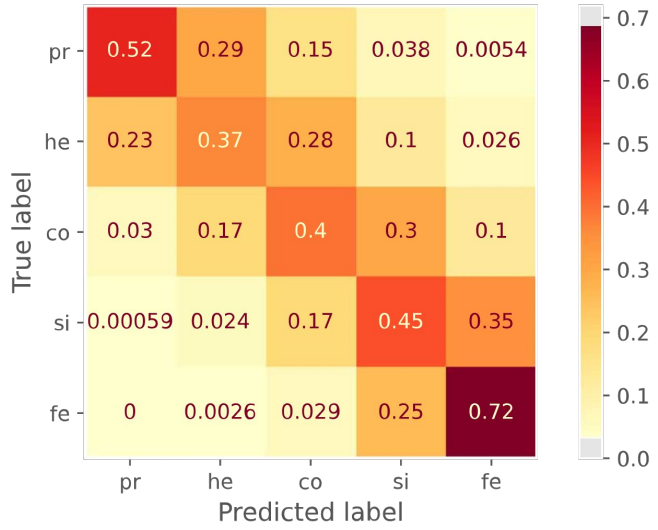
- Sibyll 2.1 hadronic model
- 4 mass groups
- ML approach
- Same energy range

Comparison published in
PoS ICRC2021 (2021) 319

CNN architecture

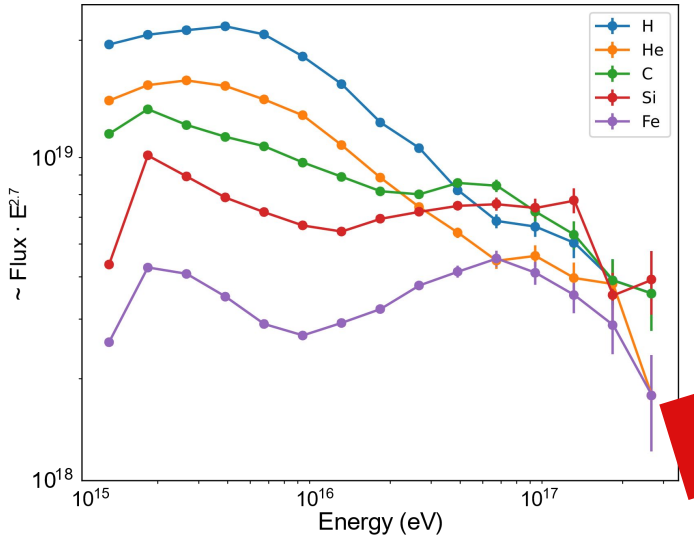


CNN performance and comparison to Random Forest



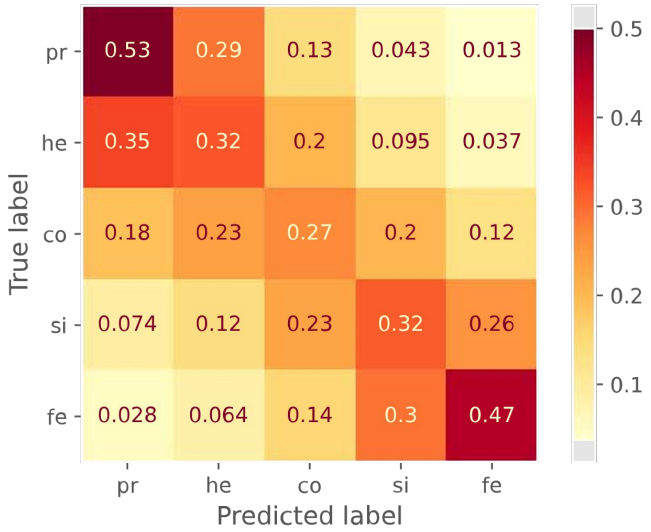
Confusion matrix for

CNN



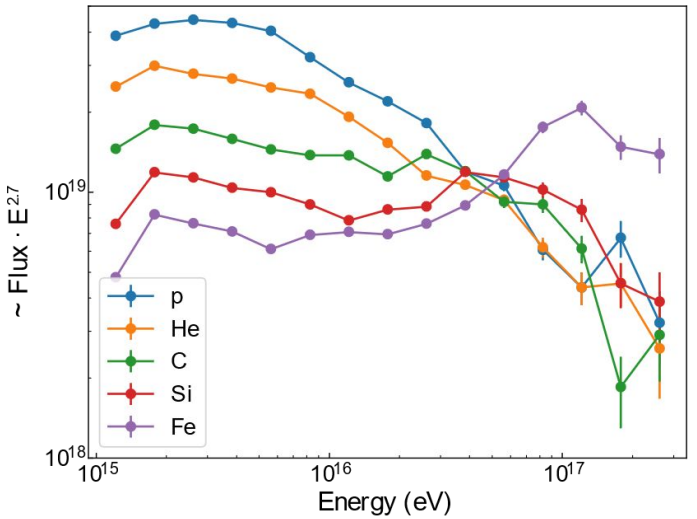
Spectra for CNN

EXEMPLARY SPECTRA
NOT FOR SCIENTIFIC USE



Confusion matrix for

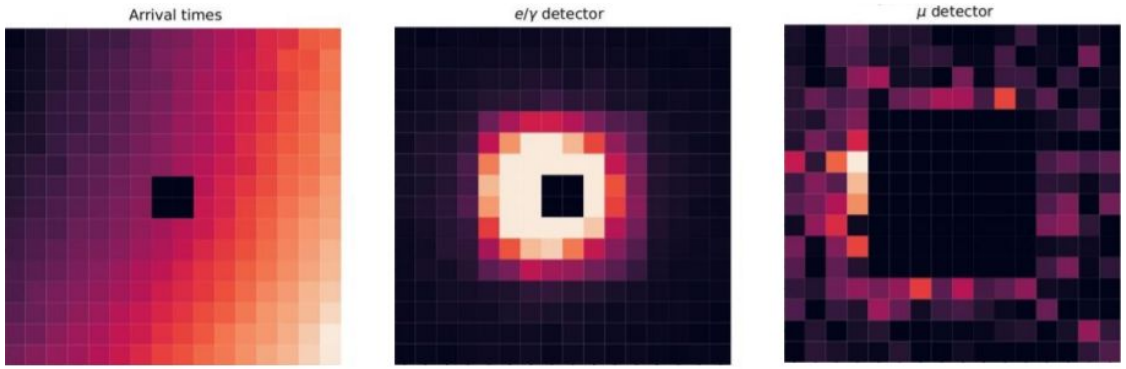
Random Forest



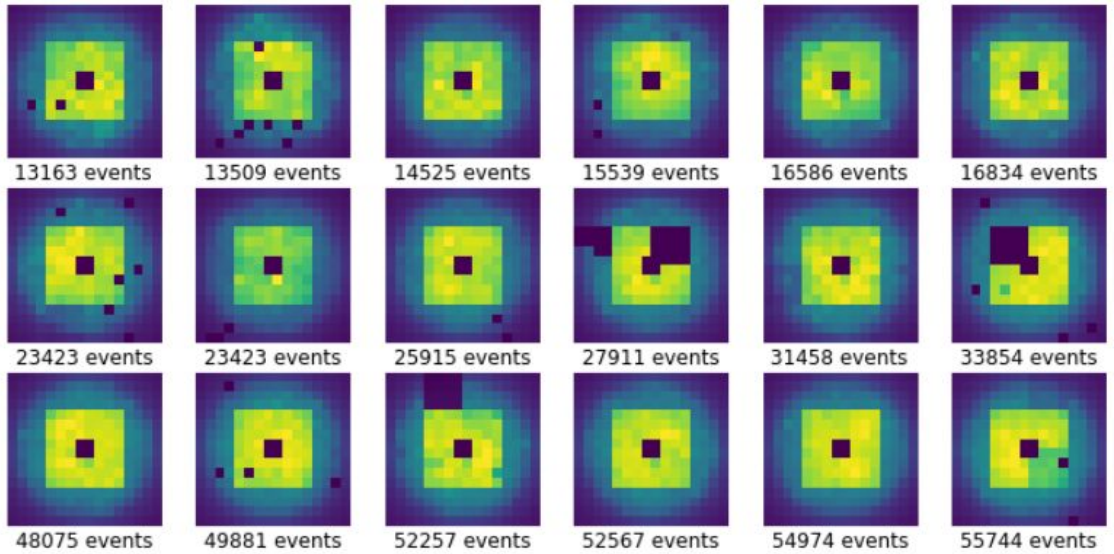
Spectra for Random Forest

Motivation behind quality cuts

In a simulated event
all detectors have 100% uptime



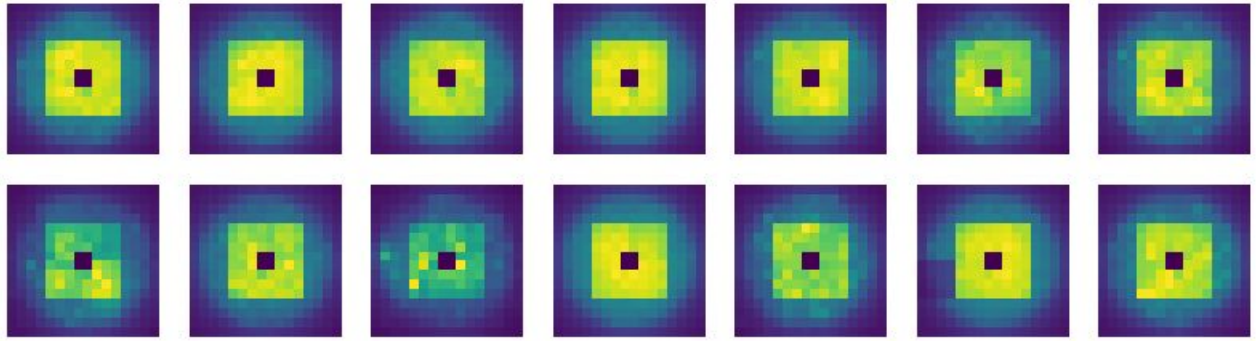
In a real event
some detectors might go down



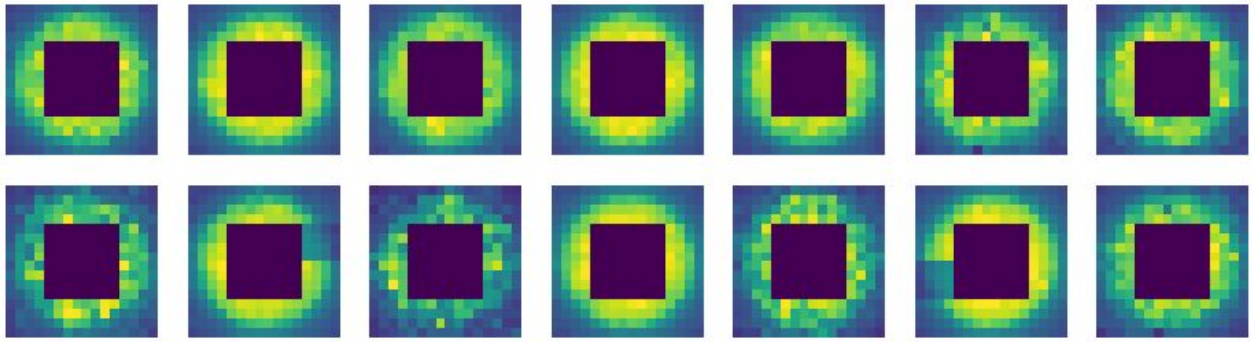
Each square shows sum of EM energy deposits for some run

Solution: drop out unsatisfactory runs

Some runs which passed our cuts

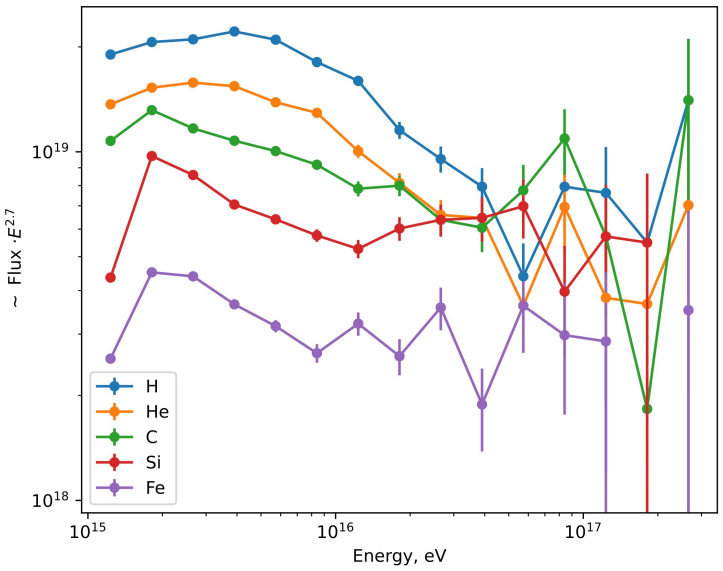


Each square shows sum of EM energy deposits for some run



Same for muons energy

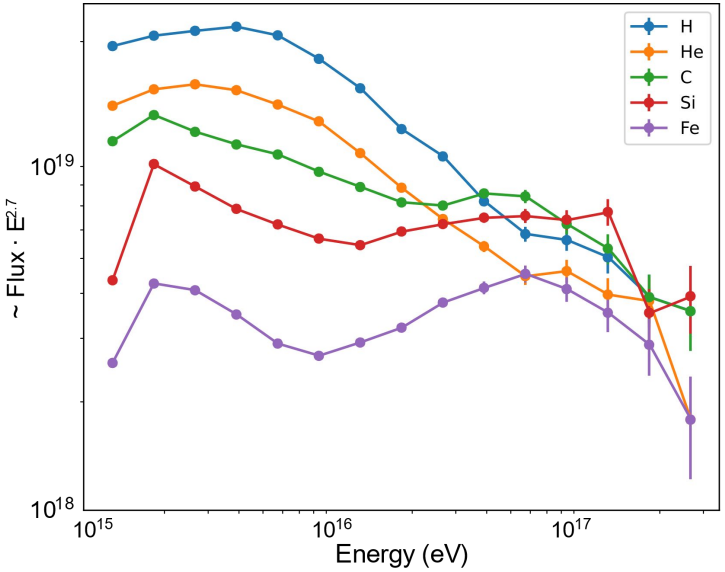
CNN performance with quality cuts



Spectra for CNN

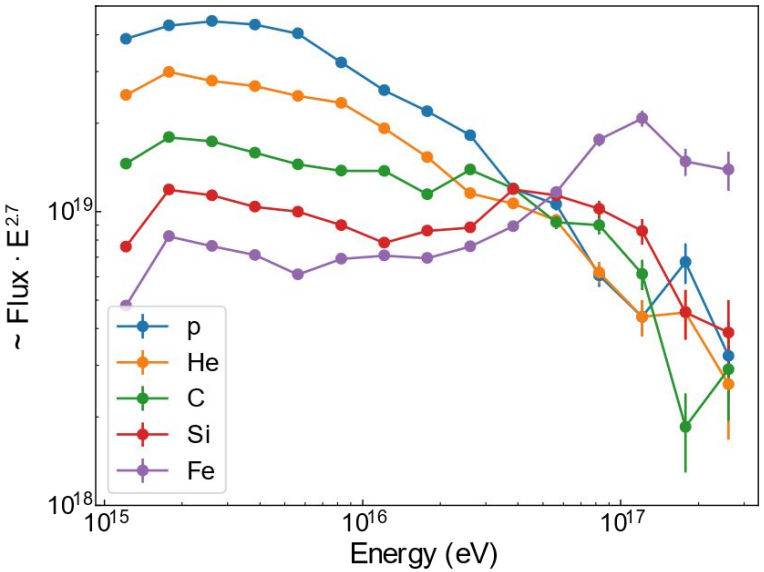
Data with quality cuts

1/3 of the full dataset



Spectra for CNN

No quality cuts



Spectra for RF

No quality cuts

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).