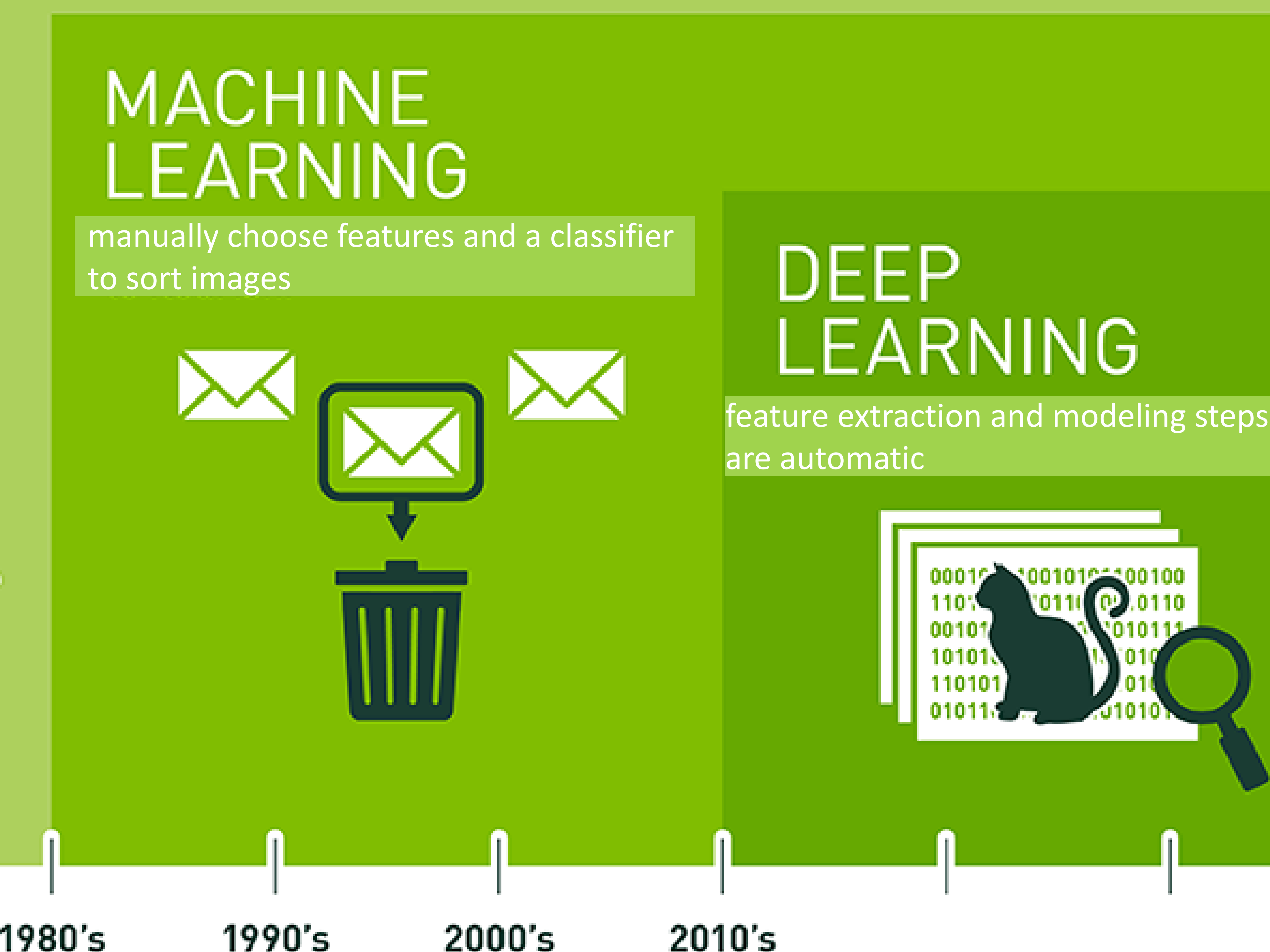
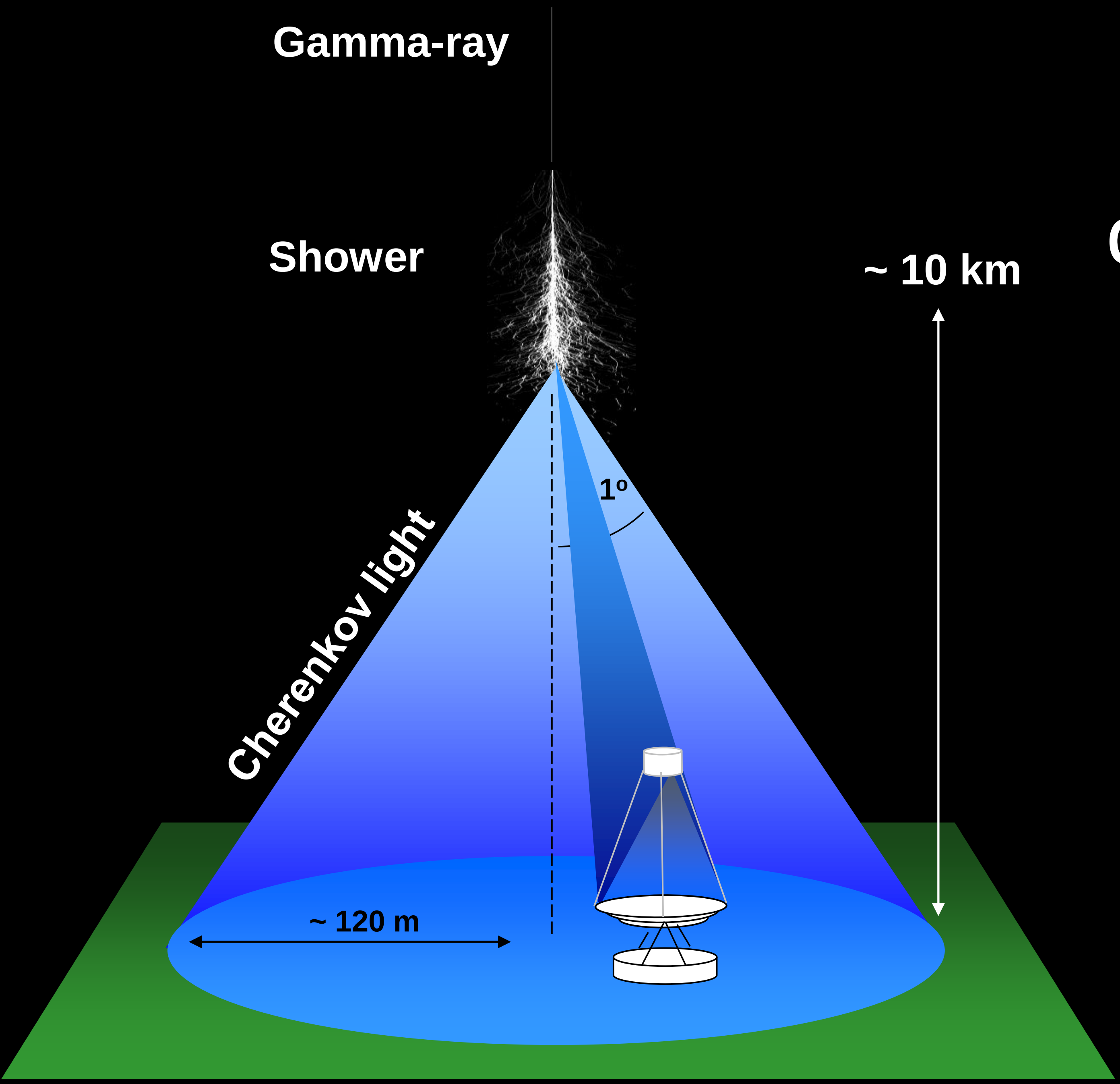
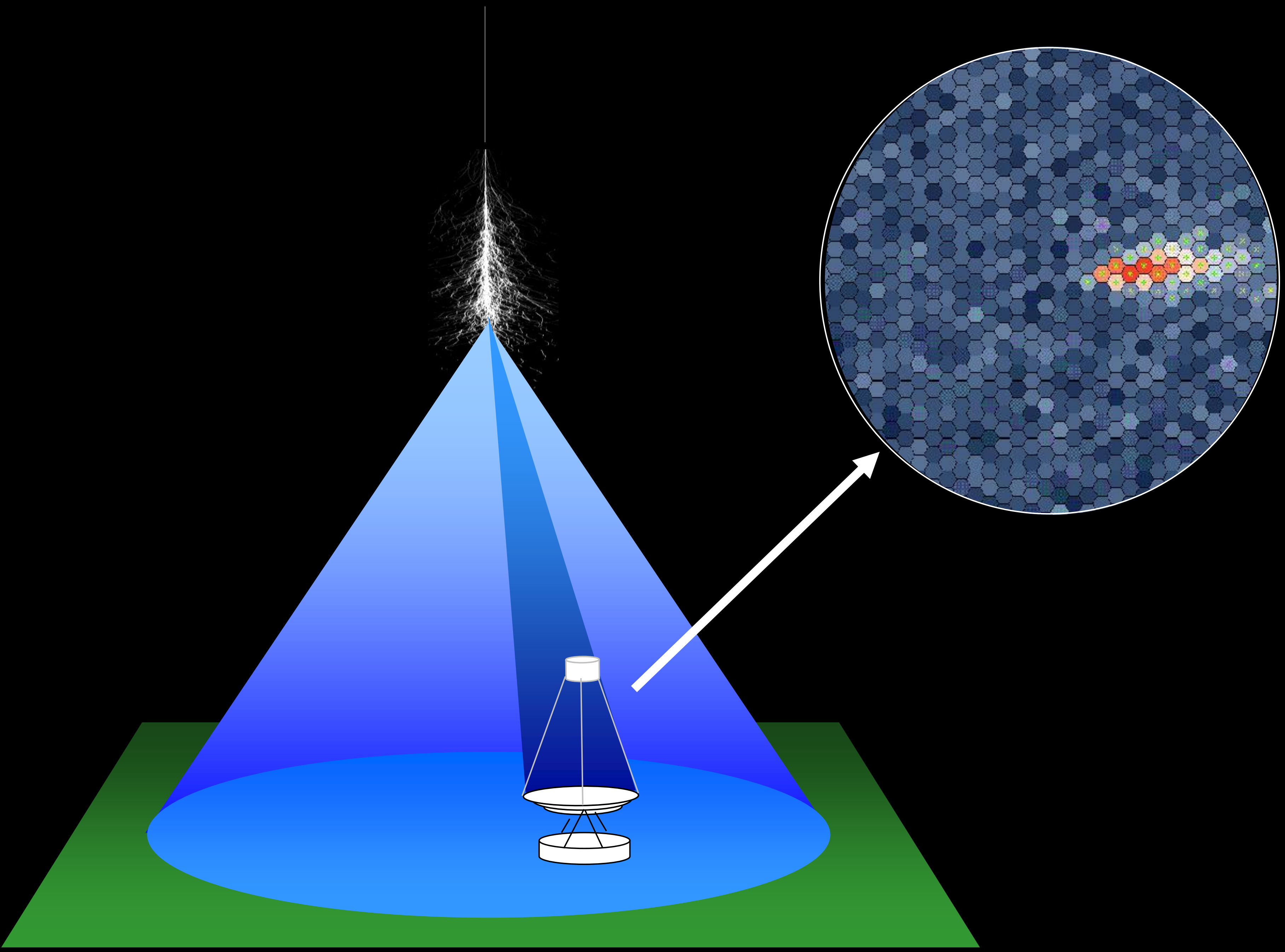
Deep Learning for IACT data analysis

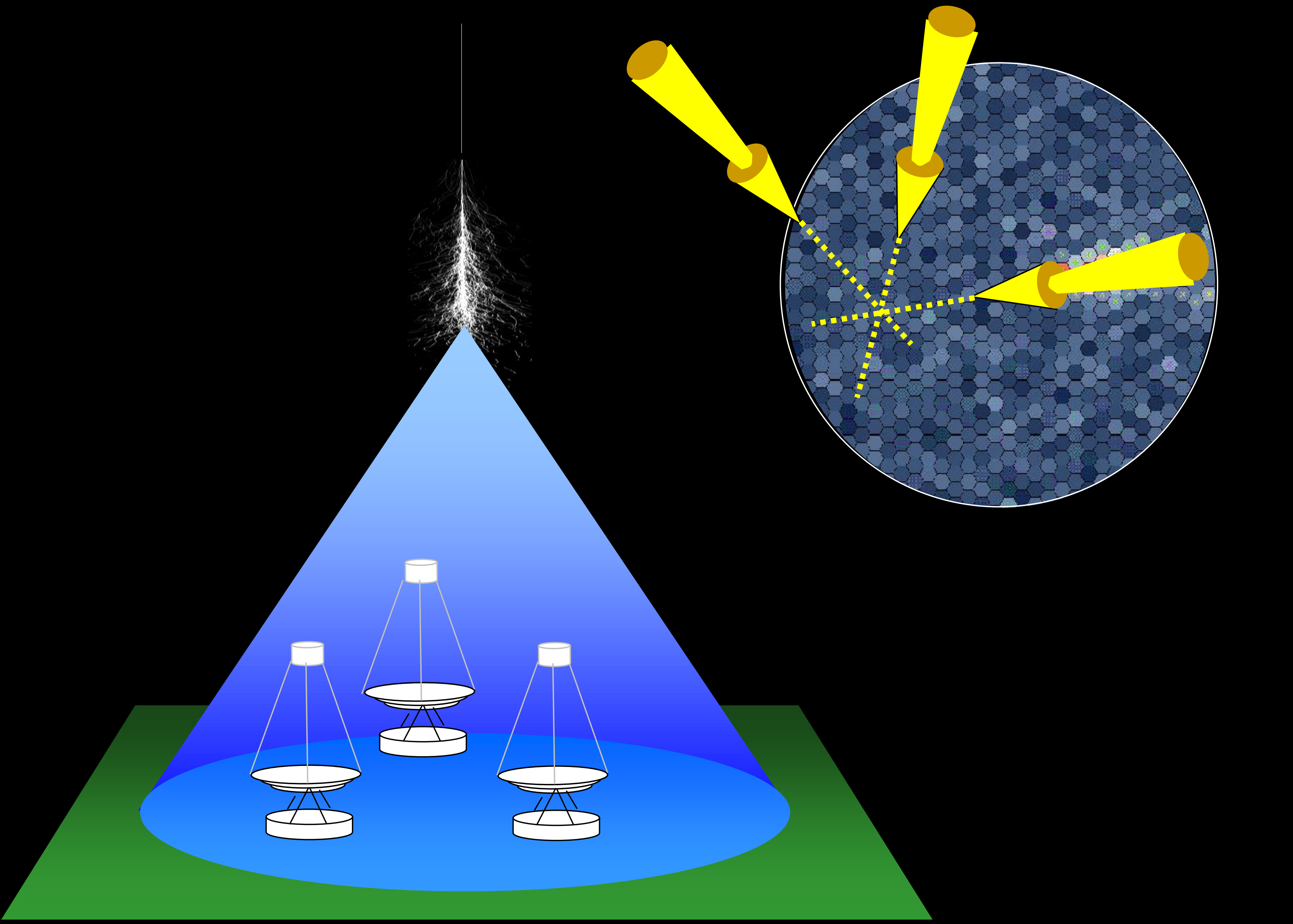
Evgeny Postnikov, SINP MSU
evgeny.post@gmail.com


MACHINE LEARNING
manually choose features and a classifier to sort images
DEEP LEARNING
feature extraction and modeling steps are automatic
1980's
1990's
2000's
2010's

IACT data taking

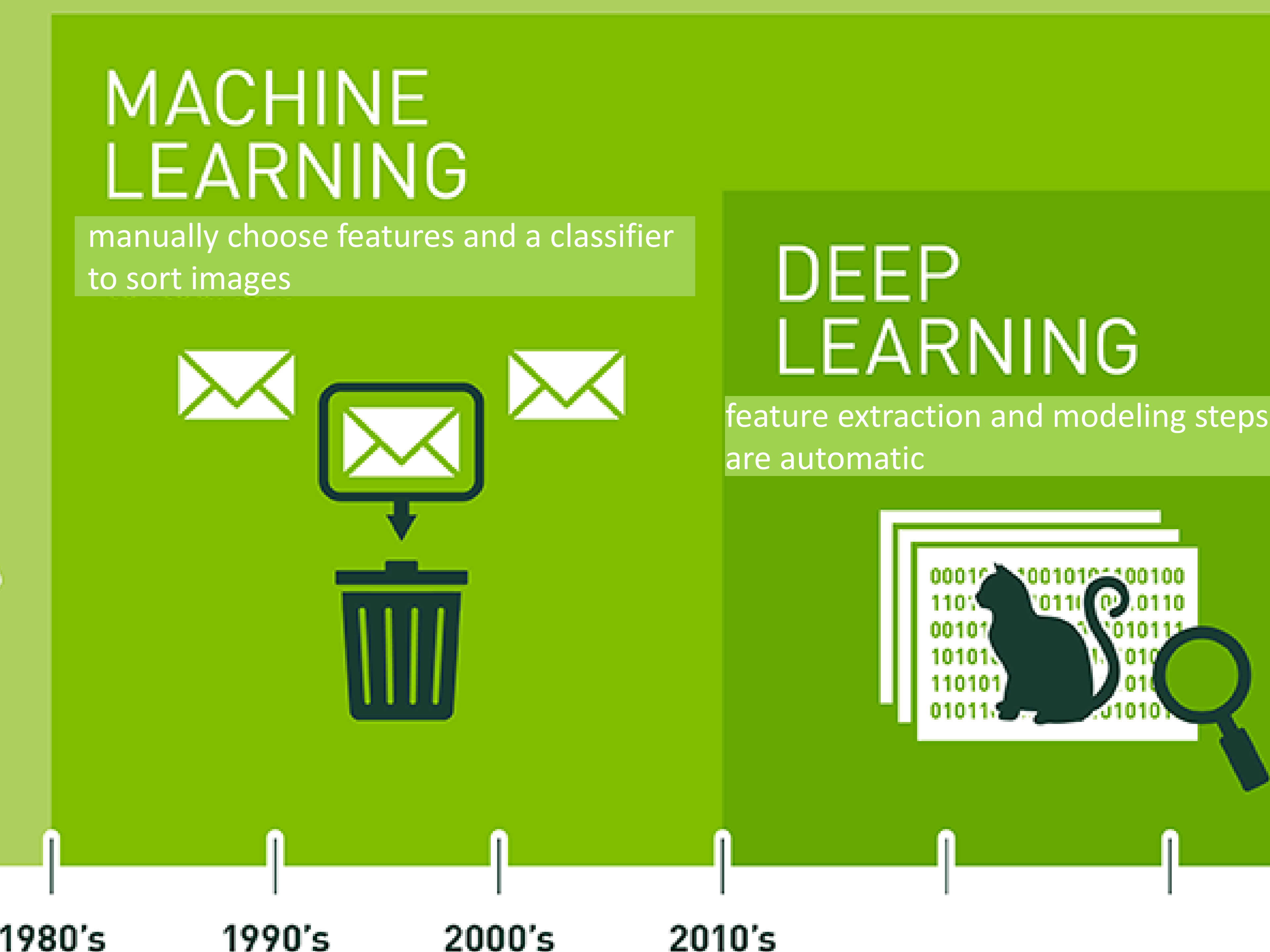






ML for primary particle identification on IACT data

- Linear discriminant analysis [MAGIC: NIMA 516 (2004) 511]
- Kernel probability density estimation (Gaussian kernel, nearest neighbours)
  [HESS: arXiv astro-ph/0611882
  MAGIC: NIMA 516 (2004) 511, 2007ICRC 1473]
- Decision trees
- ANN [Whipple: 1991ICRC 472 & 496, 1995ICRC 772
  HEGRA: arXiv astro-ph/0101318
  TACTIC: arXiv 1301.0064 & 0904.4096, astro-ph/0509012 & 0109476 & 9905312
  HESS: PoS(ICRC2015)837 & 1022
  MAGIC: arXiv astro-ph/0503539, NIMA 516 (2004) 511]
- CNN [VERITAS: PoS(ICRC2017)826, arXiv 1611.09832
  CTA: PoS(ICRC2017)809]
- Support vector machines [MAGIC: NIMA 516 (2004) 511, arXiv cs/0602083]
- Ensemble learning (random forest, boosting (AdaBoost), bagging)
  [HESS: arXiv 0904.1136, 1104.5359
  VERITAS: arXiv 1701.06928
  MAGIC: arXiv 0709.3719, NIMA 516 (2004) 511, 2007ICRC 1473]

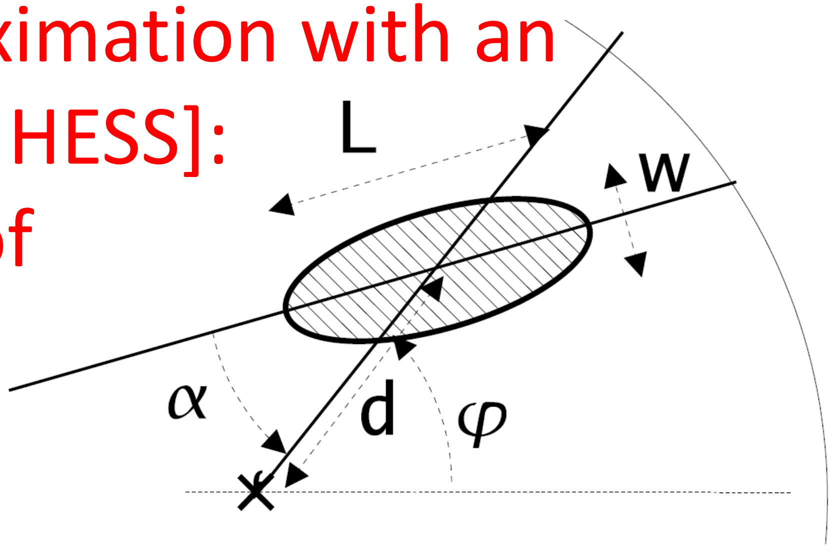

MACHINE LEARNING
manually choose features and a classifier to sort images
DEEP LEARNING
feature extraction and modeling steps are automatic
1980's
1990's
2000's
2010's

Manually choose features to sort images

- Linear discriminant analysis [MAGIC: NIMA 516 (2004) 511]
- Kernel probability density estimation (Gaussian kernel, nearest neighbours)
  [HESS: arXiv astro-ph/0611882
  MAGIC: NIMA 516 (2004) 511, 2007ICRC 1473]
- Decision trees
- ANN [Whipple: 1991ICRC 472 & 496, 1995ICRC 772
  HEGRA: arXiv astro-ph/0101318
  TACTIC: arXiv 1301.0064 & 0904.4096, astro-ph/0509012 & 0109476 & 9905312
  HESS: PoS(ICRC2015)837 & 1022
  MAGIC: arXiv astro-ph/0503539, NIMA 516 (2004) 511]
- CNN [VERITAS: PoS(ICRC2017)826, arXiv 1611.09832
  CTA: PoS(ICRC2017)809]
- Support vector machines [MAGIC: NIMA 516 (2004) 511, arXiv cs/0602083]
- Ensemble learning (random forest, boosting (AdaBoost), bagging)
  [HESS: arXiv 0904.1136, 1104.5359
  VERITAS: arXiv 1701.06928
  MAGIC: arXiv 0709.3719, NIMA 516 (2004) 511, 2007ICRC 1473]

These manually chosen features to sort images can be:

•Geometrical parameters of images (dimensions and orientation of an image approximation with an elliptical shape, a variant [CAT, HESS]: 3-dimensional generalization of geometrical parameters)



Transformation of geometrical parameters (PCA, wavelets, multifractal parameters, pseudo-Zernike moments

Automatically extract features to sort images

- Linear discriminant analysis [MAGIC: NIMA 516 (2004) 511]
- Kernel probability density estimation (Gaussian kernel, nearest neighbours)
  [HESS: arXiv astro-ph/0611882
  MAGIC: NIMA 516 (2004) 511, 2007ICRC 1473]
- Decision trees
- ANN [Whipple: 1991ICRC 472 & 496, 1995ICRC 772
  HEGRA: arXiv astro-ph/0101318
  TACTIC: arXiv 1301.0064 & 0904.4096, astro-ph/0509012 & 0109476 & 9905312
  HESS: PoS(ICRC2015)837 & 1022
  MAGIC: arXiv astro-ph/0503539, NIMA 516 (2004) 511]
- CNN [VERITAS: PoS(ICRC2017)826, arXiv 1611.09832
  CTA: PoS(ICRC2017)809]
- Support vector machines [MAGIC: NIMA 516 (2004) 511, arXiv cs/0602083]
- Ensemble learning (random forest, boosting (AdaBoost), bagging)
  [HESS: arXiv 0904.1136, 1104.5359
  VERITAS: arXiv 1701.06928
  MAGIC: arXiv 0709.3719, NIMA 516 (2004) 511, 2007ICRC 1473]

Quality comparison of primary particle identification

ML in general vs non-ML techniques:

- Some of the methods under study show to be superior to direct cuts in parameters, used so far in the analysis of all Cherenkov telescope data
- The RF method often shows superior performance in comparison with traditional semi-empirical techniques
- The classical Hillas parameters based technique is shown to be robust and efficient, but more elaborate techniques can improve the sensitivity of the analysis. A comparison of the different analysis techniques shows that they use different information for gamma-hadron separation, and that it is possible to combine their qualities
- We show the stability of the [BDT] method and its capabilities to yield an improved background reduction compared to the H.E.S.S. Standard Analysis

ANN:

- properly-trained artificial neural network are found to provide more efficient primary characterization scheme than the one based on the use of Hillas or fractal parameters
- It is shown by a neural-net approach that these parameters, when used in suitable combinations, can bring about a proper segregation of the two event types, even with modest sized data samples of progenitor particles

Quality comparison of primary particle identification

Classification trees, kernel, and nearest-neighbour methods are very close to each other.

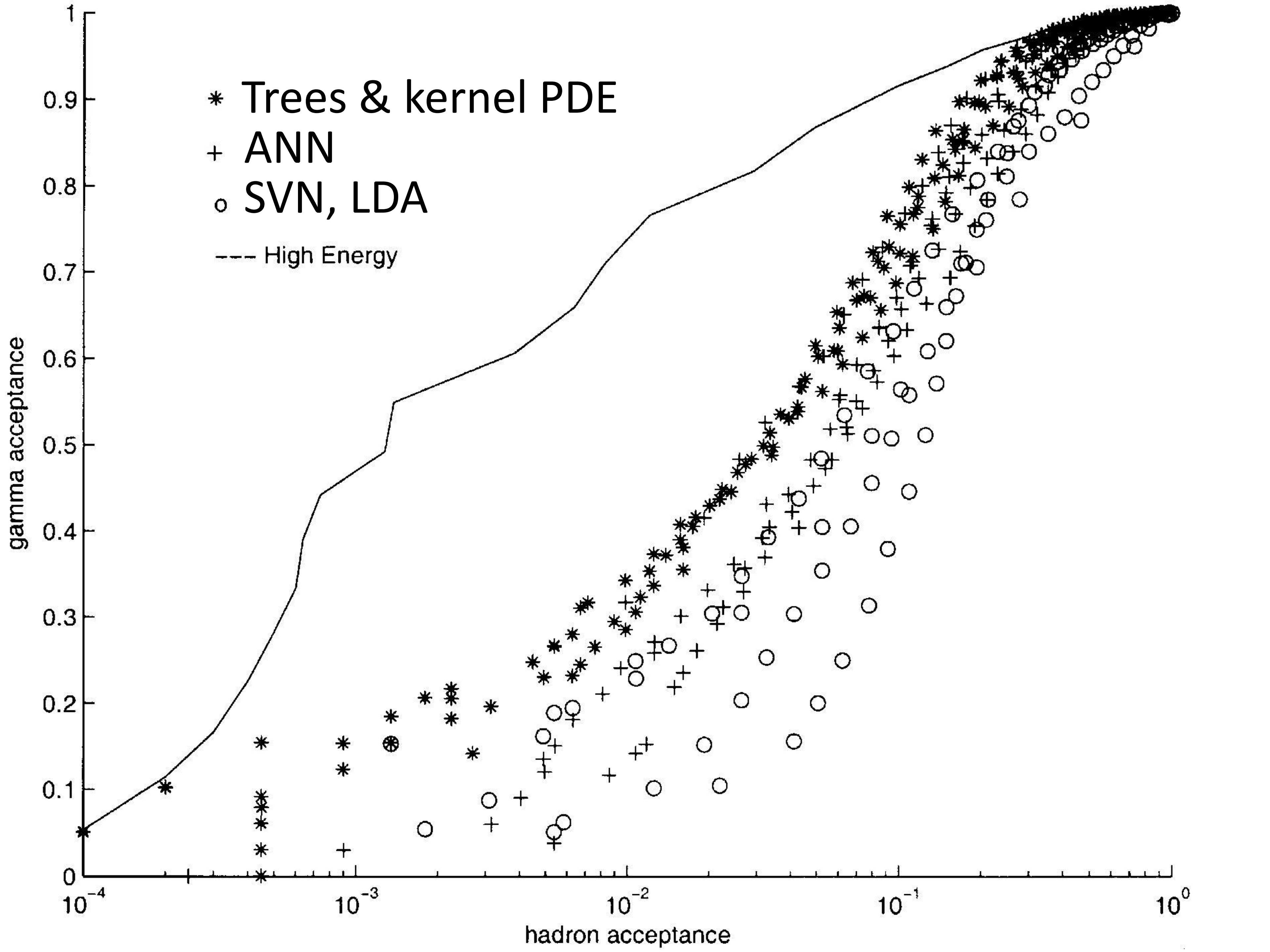
ANN give results over a wide range, between the very best and rather mediocre. This effect is not fully understood, and probably show that under the ANN umbrella name, very different methods exist, and that ANN methods need deeper understanding, they certainly cannot be used off the shelf. They also seem more sensitive to highly correlated parameters in the input.

Direct selection, LDA and SVM can be considered inferior.

[*NIMA 516 (2004) 511*]

RF training and classification are fast. A comparable analysis technique like ANN demands substantially more computer time for training.

[*arXiv 0709.3719*]



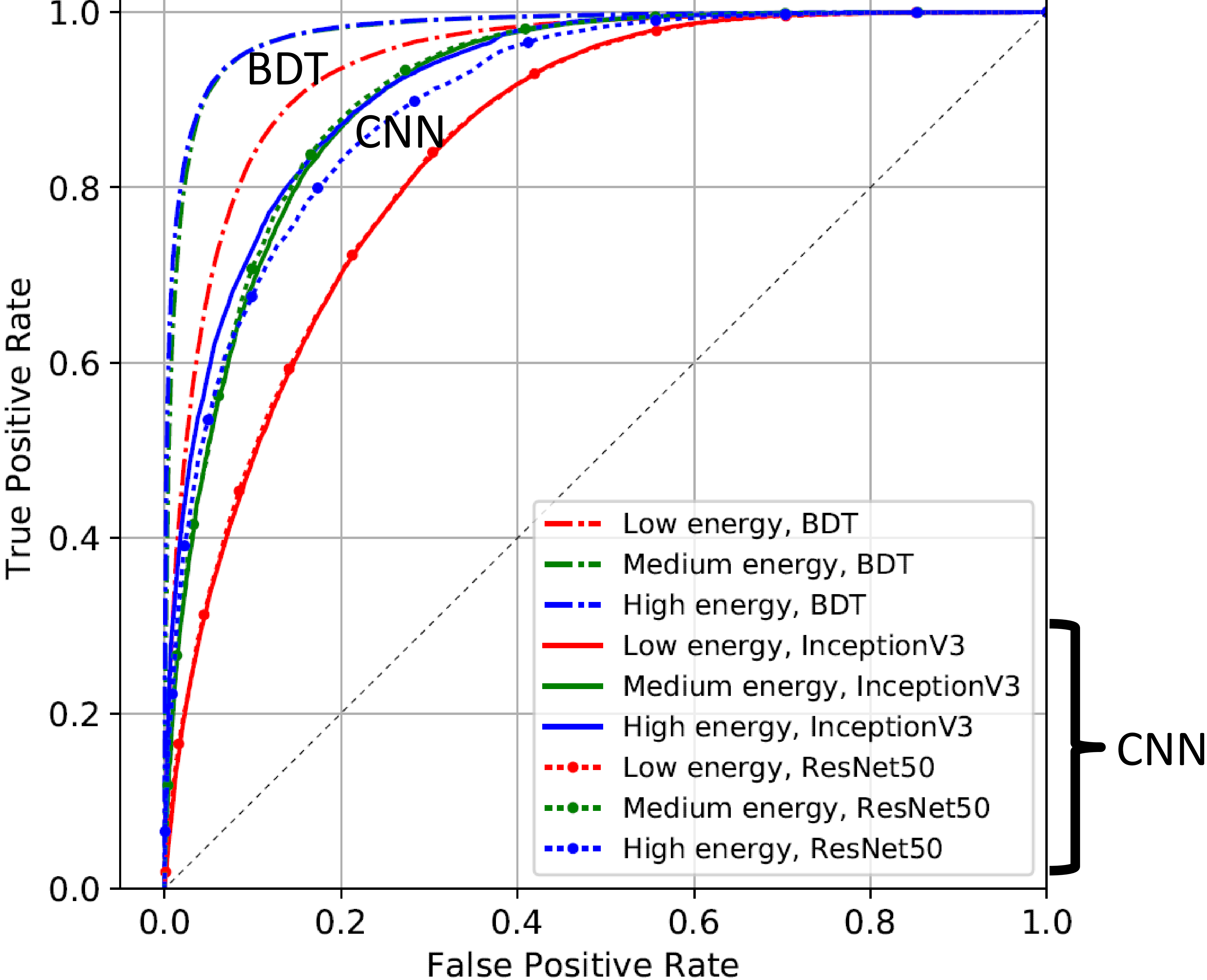
Trees & kernel PDE
ANN
SVN, LDA
High Energy
gamma acceptance
hadron acceptance
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

Quality comparison of primary particle identification

CNN are capable of classifying simulated IACT images without any prior parametrization nor any assumption on the nature of the images themselves.

The accuracy of the tested models is energy dependent, ranging from 81.4% in the low energy range to 91.6% in the high energy range for the *Inception V3* architecture.

[*PoS(ICRC2017)809*]



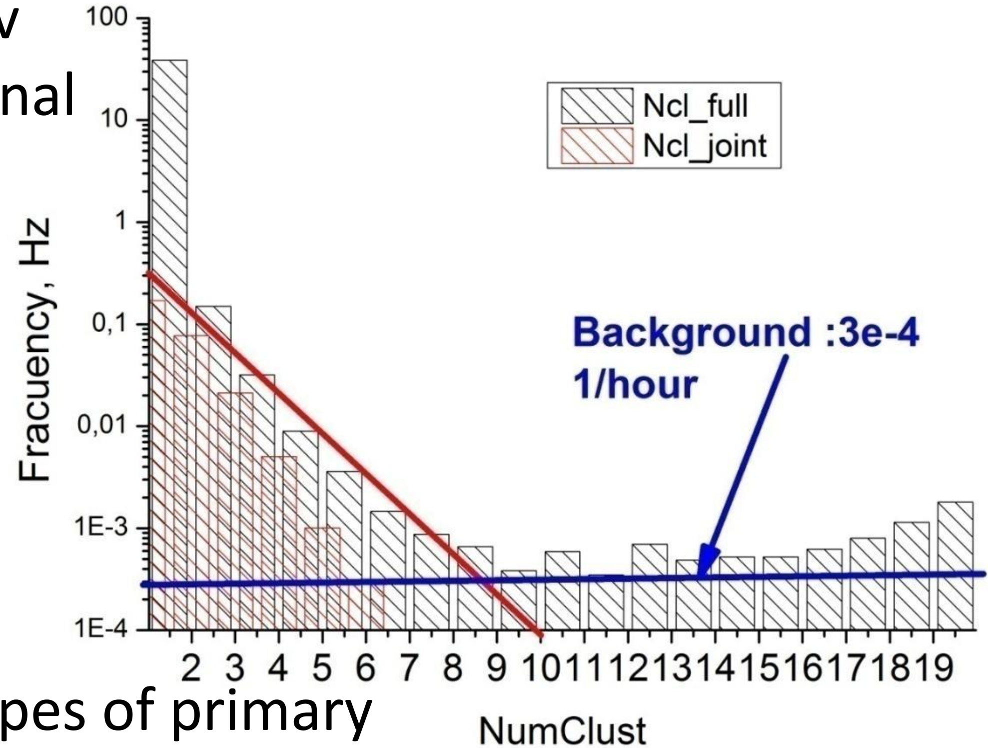
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
BDT
CNN
Low energy, BDT
Medium energy, BDT
High energy, BDT
Low energy, InceptionV3
Medium energy, InceptionV3
High energy, InceptionV3
Low energy, ResNet50
Medium energy, ResNet50
High energy, ResNet50
CNN

ML in IACT data analysis software

- MAGIC – MARS (the MAGIC Analysis and Reconstruction Software). In 2013 Random Forest classification is implemented [2013ICRC 0773]
- HESS – HAP (HESS Analysis Package). In 2015 Boosted Decision Trees classification (via AdaBoost) is included in a French part of the package, HAP-Fr [PoS(ICRC2015)837]

Other possible classification tasks on IACT data – no ML yet

•Discrimination between Cherenkov signal and night sky background signal (presently to be solved as an 'Image cleaning' procedure)



•Some special 'events' such as 'snow events' in TAIGA (presently to be solved empirically)

•Discrimination between various types of primary nuclei (protons, helium, oxygen, iron etc. – presently not to be solved by IACTs in general – with only very few exceptions). "Other promising applications for CNNs to IACT data analysis include performing more difficult background discrimination tasks such as gamma/electron separation or cosmic-ray composition studies" [PoS(ICRC2017)809].

ML for reconstruction & estimation on IACT data

- Primary energy estimation [MAGIC: arXiv 0709.3719]

- Other promising applications for CNNs to IACT data analysis include using similar methods to those presented here [PoS(ICRC2017)809] to do energy and angular reconstruction

Backup slides

- Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope, Nucl. Instrum. Meth. A 516, 511-528, 2004.
- Implementation of the Random Forest Method for the Imaging Atmospheric Cherenkov Telescope MAGIC, arXiv 0709.3719, 2007.
- Classification Methods for MAGIC Telescope Images on a Pixel-by-pixel base, Proceedings of the 30th International Cosmic Ray Conference. July 3 - 11, 2007, Mérida, Yucatán, Mexico. Volume 3, p.1473-1476.
- First results on characterization of Cerenkov images through combined use of Hillas, fractal and wavelet parameters, arXiv astro-ph 9905312, 1999.
- Image and Non-Image Parameters of Atmospheric Cherenkov Events: a comparative study of their -ray/hadron classification potential in UHE regime, arXiv astro-ph 0109476, 2001.
- Particle Identification by Multifractal Parameters in γ-Astronomy with the HEGRA-ˇCerenkov-Telescopes, arXiv astro-ph 0101318, 2001.
- A Third Level Trigger Programmable on FPGA for the Gamma/Hadron Separation in a Cherenkov Telescope Using Pseudo-Zernike Moments and the SVM Classifier, arXiv astro-ph 0602083, 2006.
- Neural Networks for Gamma-Hadron Separation in MAGIC, arXiv astro-ph 0503539, 2005.
- Analysis methods for Atmospheric Cerenkov Telescopes, arXiv astro-ph 0607247, 2006.
- A high performance likelihood reconstruction of gamma-rays for Imaging Atmospheric Cherenkov Telescopes, arXiv 0907.2610, 2009.
- Gamma-Hadron Separation in Very-High-Energy γ-ray astronomy using a multivariate analysis method, arXiv 0904.1136, 2009.
- Study on Cosmic Ray Background Rejection with a 30 m Stand-Alone IACT using Non-parametric Multivariate Methods in a sub-100 GeV Energy Range, arXiv astro-ph 0611882, 2006.

- Implementation of the Random Forest Method for the Imaging Atmospheric Cherenkov Telescope MAGIC, arXiv 0709.3719, 2007.

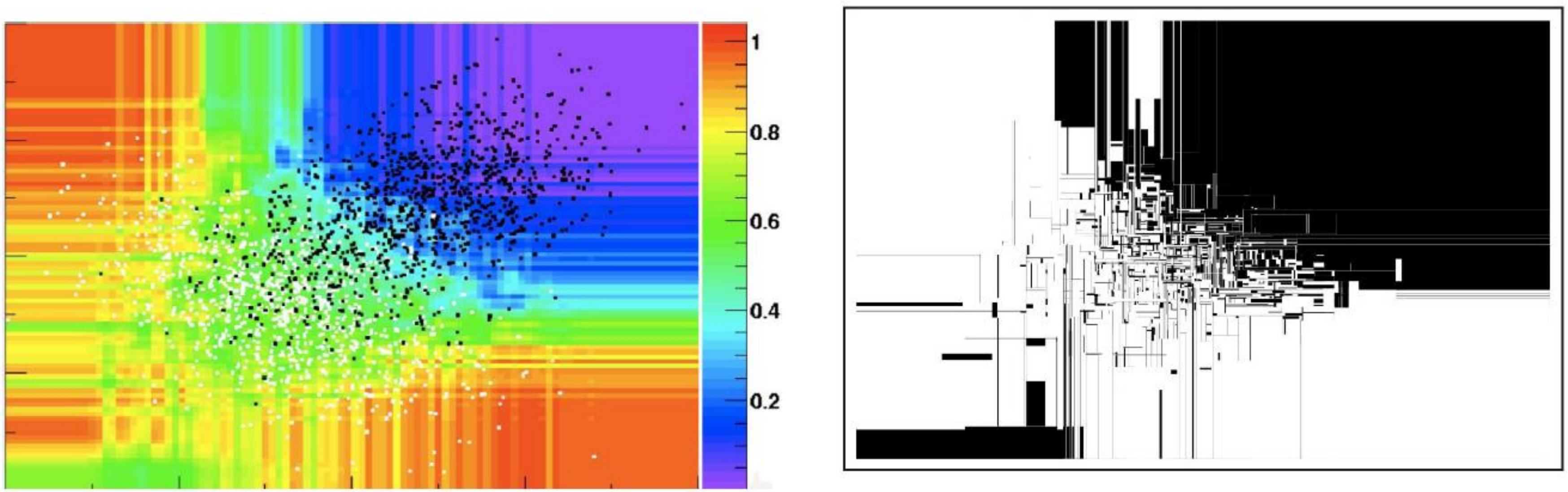


Fig. 1. *Left: Illustration of the RF method for a simple 2-dimensional model case. The black and white points are the observed points in class gamma and hadrons, respectively. They are distributed according to two different, but overlapping 2-dimensional Gaussians. The result of separation in terms of hadronness is shown in colour. Right: The result of using a single tree on the same data gives no probability measure like hadronness, but only y/n answers. Its performance is inadequate.*

Towards a common analysis framework for gamma-ray astronomy, arXiv 1307.5560, 2013

…developed a common analysis framework for gamma-ray astronomy data that can potentially be used for the analysis of any type of event data. So far, interfaces have been implemented to support the analysis of COMPTEL, Fermi-LAT and IACT data. Several applications illustrate the current capabilities of the framework, including joint multi-instrument spectral analyses and morphology studies. The basic building blocks of the framework are now implemented and tested; future work will be dedicated to expand the support to additional gamma-ray telescopes, and to enrich the existing interfaces for more complex analyses.

Demonstrating the potential of this framework for the future CTA observatory, allowing the scientific analysis of the observatory's data itself, and enabling the joint analysis of CTA data with data from other instruments, such as those from the Fermi-LAT telescope.

Still need to demonstrate that GammaLib and *ctools* can cope with the complex particle background that is encountered in VHE astronomy (so far, tests have only been done on the Crab nebula, which is a bright point source for which the particle background modeling is less important).

This will be achieved by applying these tools to existing data from H.E.S.S. and the other active VHE experiments. It's planned to implement also the conventional analysis methods in GammaLib and *ctools*, enabling cross-checking with results obtained by the existing analysis chains.

GammaLib and *ctools* are open source community tools. The software can be freely downloaded from http://sourceforge.net/projects/gammalib and http://cta.irap.omp.eu/ctools,
and everybody is welcome to join the development team for making the product even better.

Машинное обучение

- Кластеризация

Используется в основном при анализе карт звездного неба в гамма-спектре для поиска новых гамма-источников; самым популярным алгоритмом является Gaussian Mixture Models

Машинное обучение

- Кластеризация

An Einstein@home blind search for gamma-ray pulsars

The 13 new gamma-ray pulsars were discovered by searching 118 unassociated Fermi-LAT sources from the third LAT source catalog, selected using the Gaussian Mixture Model machine learning algorithm on the basis of their gamma-ray emission properties being suggestive of pulsar magnetospheric emission.

Clustering of gamma-ray burst types in the Fermi-GBM catalogue: evidence for photosphere and synchrotron emissions during the prompt phase

We use Gaussian Mixture Models to cluster gamma-ray bursts according to their parameters. We find five distinct clusters.