01.05.2024
Kathrin Nippel

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

TTK Institute for Theoretical Particle Physics and Cosmology

RWTH AACHEN UNIVERSITY

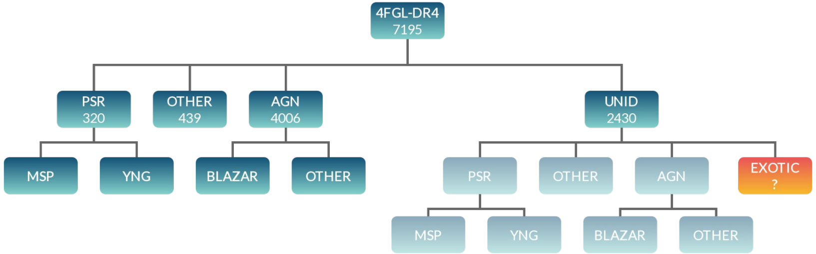
Searches for Exotic Objects among Fermi-LAT γ-Ray Sources with Weakly Supervised Machine Learning

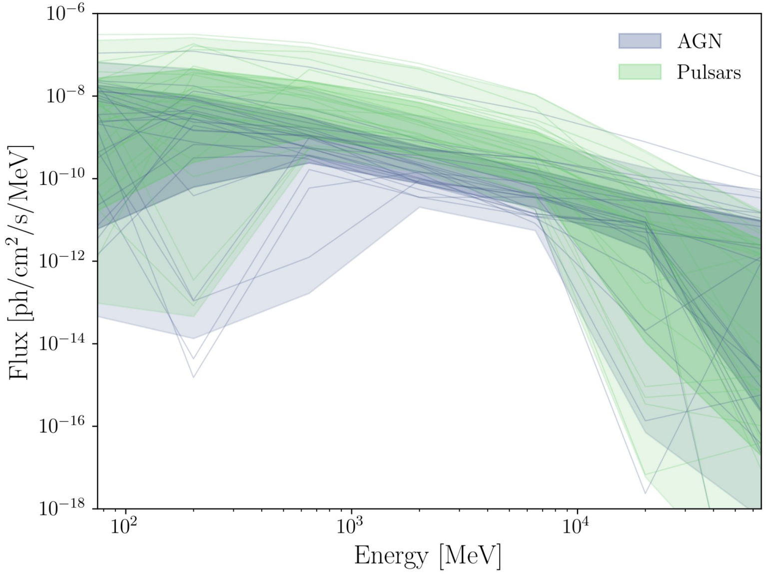
Michael Krämer[1], *Silvia Manconi*[2], ***Kathrin Nippel***[1]

[1] TTK RWTH Aachen, [2]LAPTh, CNRS

γ-Ray Observation with *Fermi*-LAT

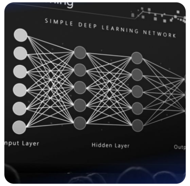
- Wide field-of-view detector
- Covering the high-energy range of γ-ray energies
  - between about 20 MeV and up to a few TeV
- Observations since 2008, still ongoing
- Publicly available source catalog





Goal: Find **anomalies**/exotic sources among observed data

4FGL-DR3: S. Abdollahi *et al* 2022 *ApJS* 260 53



Classifier Setup
for γ-Ray Classification

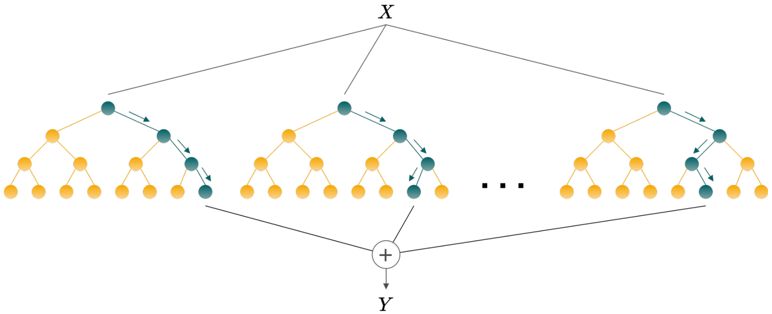
State-of-the-art γ-Ray classification:

- FFNN:
  - e.g.: AGN vs. Pulsar - *Finke et al. (2020) [2012.05251]*
  - ➜ Shows that flux spectrum of Fermi-LAT data is suitable for successful classification
  - Many other applications, e.g. multiclass, . . .

- Bayesian NN:
  - Blazar Classification - *Butter et al. (2022) [2112,01403]*
  - Dark Matter Subhalos vs Astrophysical Sources - *Butter et al. (+KN) (2023) [2303.07362]*
  - ➜ Provides robust classification by additional consideration of classification uncertainty

This work: **Boosted Decision Tree Classifier**

```
GradientBoostingClassifier(n_estimators = 200,
                           max_depth = 4,
                           learning_rate = 0.1)
```



Inspiration: *'Back To The Roots: Tree-Based Algorithms for Weakly Supervised Anomaly Detection'*,
- Finke et al. (2023) [2309.13111]

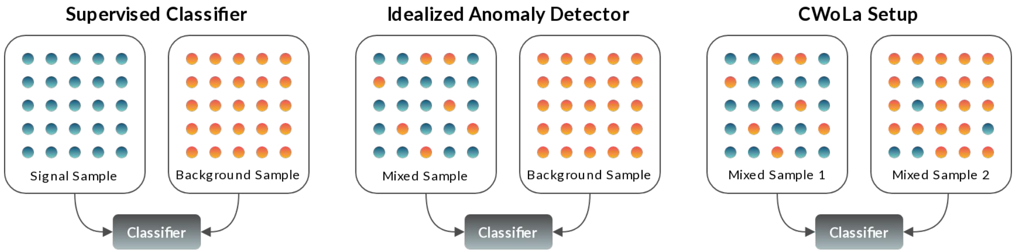
Advantages of BDT approach:

- BDT works well on small data sets with strong class imbalances
- Trains very fast and can outperform DNN on 'simple' classification problems with tabular data

Searching for Anomalous γ-Ray Sources with **Weakly Supervised Classification**

TTK Institute for Theoretical Particle Physics and Cosmology | RWTH AACHEN UNIVERSITY

01.05.2024 Kathrin Nippel



CWoLa:
Classification without Labels
Follow approach from Metodiev et al. (2017) [1708.02949]

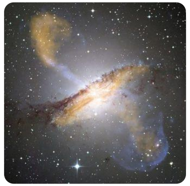
An optimal classifier trained to distinguish the mixed samples is also optimal to distinguish signal from background objects

- Optimal classifier is given by the likelihood ratio, this relates to:

$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1 p_S + (1-f_1)p_B}{f_2 p_S + (1-f_2)p_B} = \frac{f_1 L_{S/B} + (1-f_1)}{f_2 L_{S/B} + (1-f_2)}$$

- Training a classifier to maximize L_{M_1/M_2} yields the optimal classifier also to discriminate signal and background if $f_1 > f_2$
- Pure background sample ↔ idealized anomaly detector

This work: Setup exemplary approaches where *f* can be controlled and supervised classification can be used as a benchmark



Case I - Astrophysical Sources

Astrophysical data only:

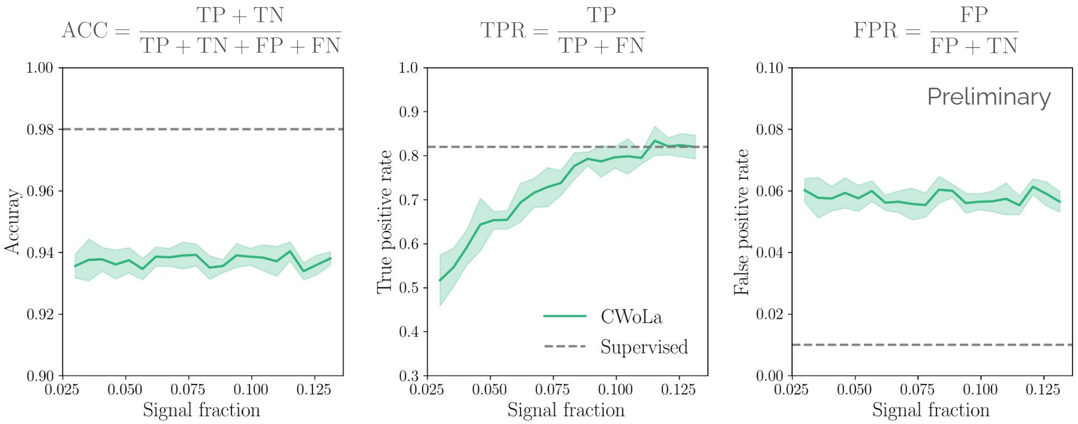
- Background: AGN
- Signal: Pulsars

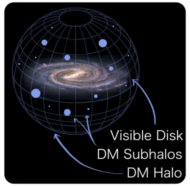
- Small signal fractions ($f_s \lesssim 0.1$)
- Background & mixed sample in $\mathcal{O}$(4FGL / UNID)

Data Augmentation:

- Use normalizing flow (rational quadratic spline) to generate additional AGN data
  - Realistic generated spectra: ACC ≤ 52.6%

- Used to probe small signal fractions without too sparse signal





Case II - Dark Matter Subhalos

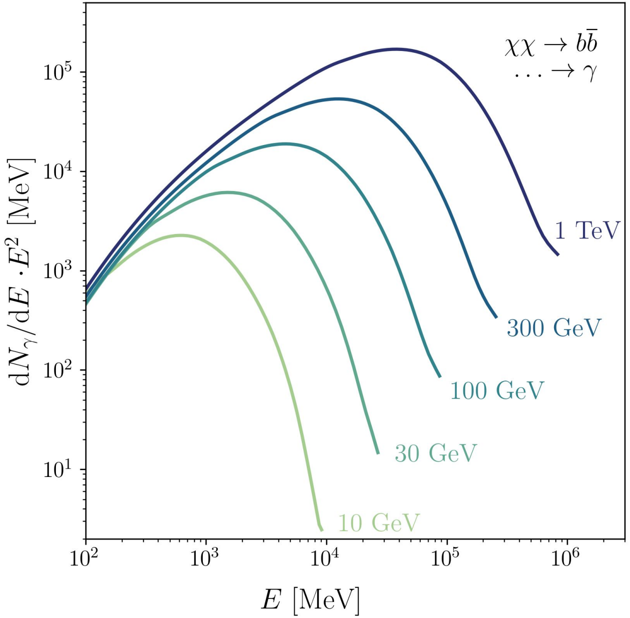
- Dark matter is clustered hierarchically in galactic halos
  - According to cosmological simulations
  - Substructures ("Subhalos") form

- Signal in γ-rays expected, following

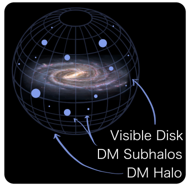
$$\phi = \frac{\langle \sigma v \rangle}{8 \cdot \pi \cdot m_{\mathrm{DM}}} \cdot \mathcal{J} \cdot \frac{\mathrm{d}N}{\mathrm{d}E}, \qquad \mathcal{J}(\Delta\Omega) = \int_{\Delta\Omega} \mathrm{d}\Omega \int_{\mathrm{l.o.s.}} \mathrm{d}s\ \rho^2(x)$$

- We simulate subhalo population with CLUMPY

- Realistic simulation of observable signal using fermipy
  - Use detectable subset ($\sigma_d > 5$) of simulated subhalo population
  - For details on full simulation procedure see [2303.07362]

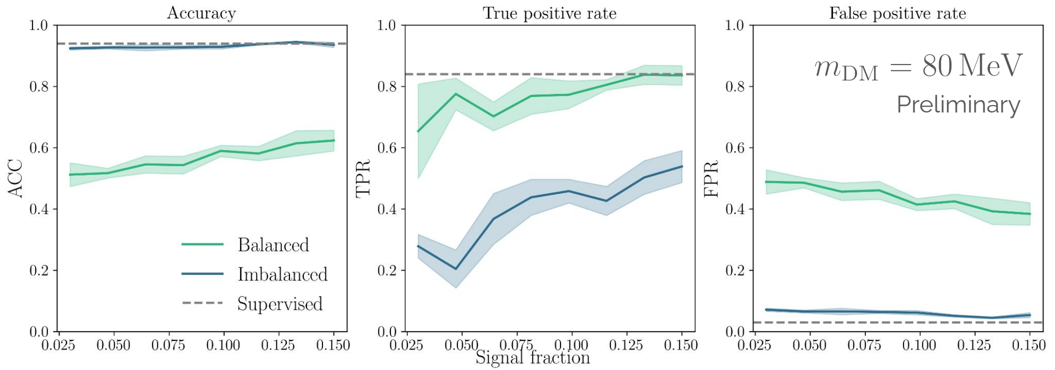


CLUMPY v3: Hütten *et al. 2019* CPC 235 336–345

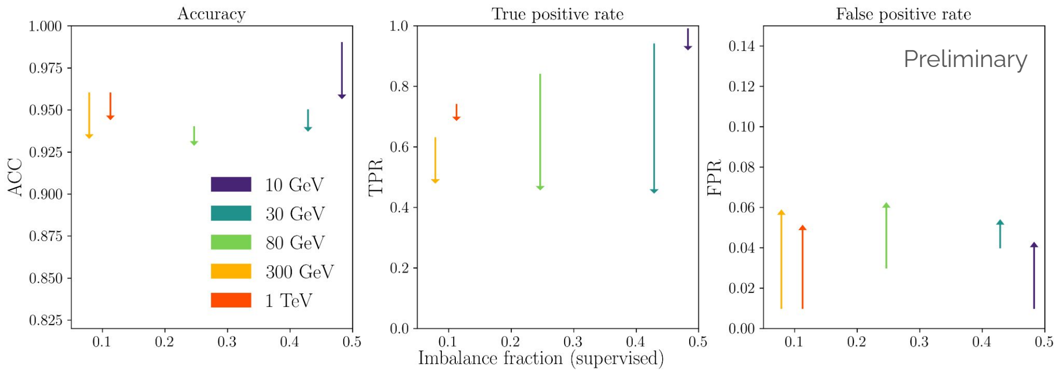


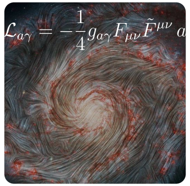
Searching for Subhalos with
Weakly Supervised Classification

- Balanced: $N_{\text{Mixed}} = N_{\text{Background}}$
- Imbalanced: $N_{\text{Mixed}} = \frac{1}{2} N_{\text{Background}}$
- CWoLa performance dependent on training class imbalance



- Comparison for different dark matter masses
- Imbalance in supervised case relevant for TPR
- More 4FGL-similar halos (intermediate masses) lose most predictive power with CWoLa



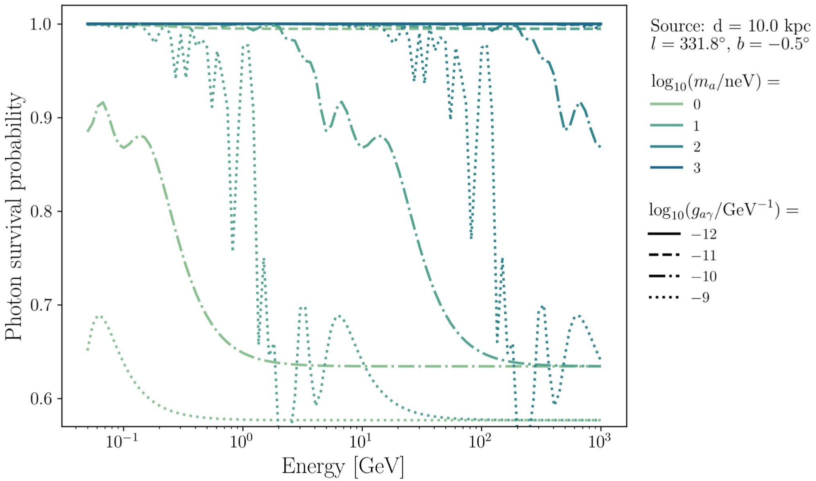


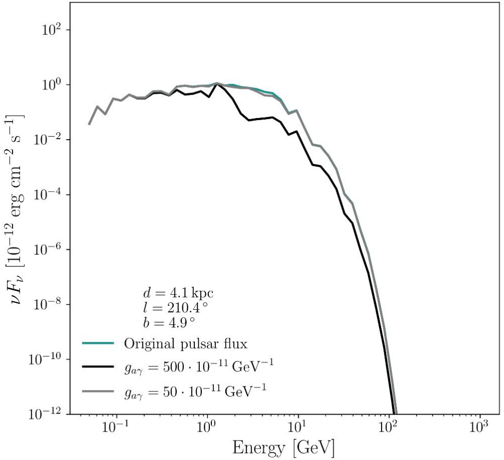
Case III - Axion-Photon Conversion

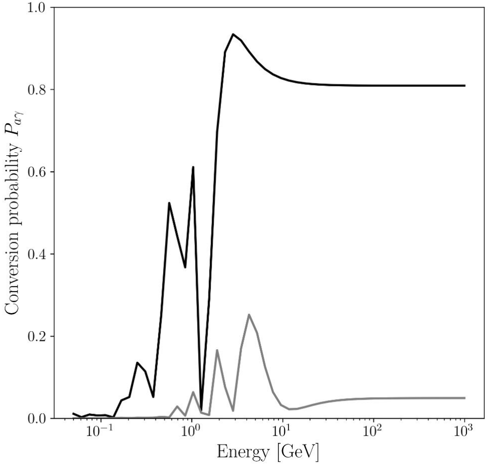
Axion-photon coupling leads to conversion in magnetic field $\mathcal{L}_{a\gamma} = -\frac{1}{4} g_{a\gamma} F_{\mu\nu} \tilde{F}^{\mu\nu}\, a$

➜ Energy dependent oscillatory suppression of γ-ray flux

- Signal: Pulsars with oscillatory pattern from ALP-photon conversion in the GMF
  - Assume fixed axion mass and axion-photon coupling strength
- Background: Unmodulated pulsar spectra







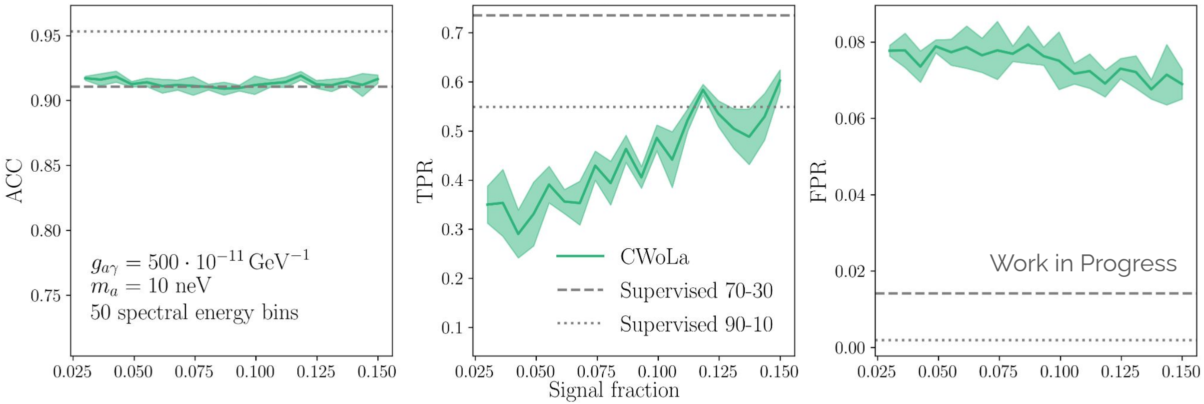
Case III - Axion-Photon Conversion

Pick parameters leading to sufficiently strong modulation:

- Mass: $m_a = 10\,\mathrm{neV}$
- Coupling: $g_{a\gamma} = 5 \cdot 10^{-9}\,\mathrm{GeV}^{-1}$

Supervised benchmark: Different initial class imbalances

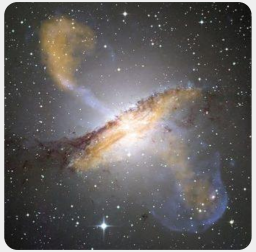
- CWoLa Training class imbalance: 70/30%
- ‘Low signal fraction’ setup: 90/10%

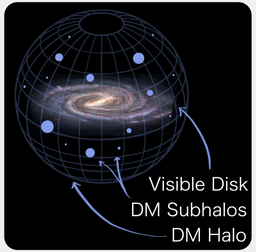


Complex setup also for supervised classification (depends on ALP parameters)

➜ In such cases, CWoLa leads to larger FPR // other features in data might overshadow signal

Conclusion and Outlook





Application of CWoLa in search for anomalous γ-ray sources

- Demonstrated potential on three exemplary cases
- Supervised limit can be reached at small signal fractions

Challenges:

- Small data sets - few signal samples overall
- Results depend on how 'well-defined' classification setup is also in supervised case

Take-aways:

- Method useful to point out interesting sources for further investigation
- Findings of test setups can help guide CWoLa setup on unlabeled data

Thank you for your attention!