Machine Learning for Dark Matter Detection via Photon Signatures at the LHC

Rosa María Sandá Seoane

Departamento de Física Teórica Universidad Autónoma de Madrid &
Instituto de Física Teórica UAM-CSIC

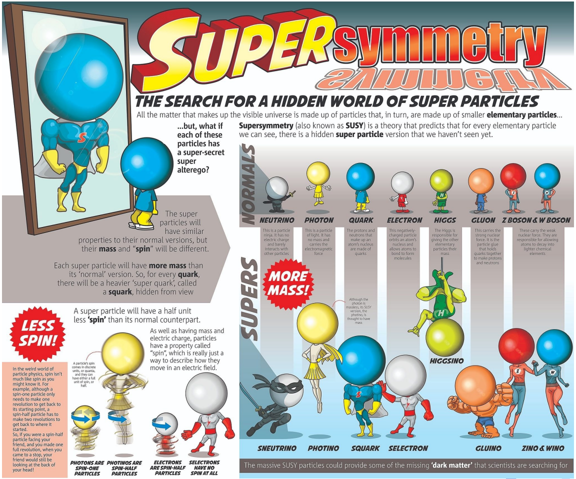
EPS-HEP 2025
July 11, 2025

References

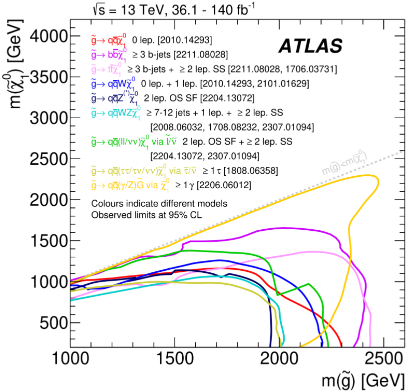
- E. Arganda, M. de los Ríos, A. D. Perez, S. Roy, RMSS, C. Wagner, **“Machine Learning Analysis for Dark Matter Detection via Photon Signatures at the LHC,”** [2508.XXXXX].
- S. Roy, and C. E. M. Wagner, **“Dark Matter searches with photons at the LHC,”** JHEP 04 (2024), 106 [2401.08917].
- E. Arganda, M. Carena, M. de los Ríos, A. D. Perez, D. Rocha, RMSS, C. Wagner,**“Machine-Learning Collider Analysis of Radiative Neutralino Decays at the LHC,”** JHEP 07 (2025), 014 [2410.13799].

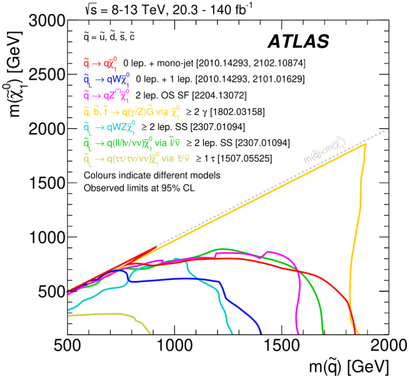
Why searching for radiative neutralino decays at the LHC?

- SUSY provides an explanation for the scale of EWSB (soft SUSY breaking scale) and for the DM (with the LSP $\tilde{\chi}_1^0$ if R-parity is conserved).

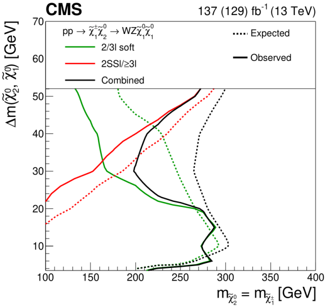


- Strong constraints at LHC for colored supersymmetric partners.



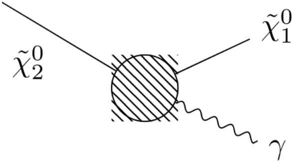


- Weakly interacting particles, instead, may be light and can be probed at the HL-LHC (e.g. MSSM, NMSSM).



- In this scenario, the proper cosmological relic density can be achieved in the co-annihilation/compressed spectra, where the mass of the LSP is close to other weakly interacting particles, like the second lightest neutralino $\tilde{\chi}_2^0$ and the charginos $\tilde{\chi}_1^{\pm}(m_{\tilde{\chi}_1^0} \sim m_{\tilde{\chi}_2^0} \sim m_{\tilde{\chi}_1^{\pm}})$.

- If the direct detection cross-section of DM (LSP) is suppressed within the compressed region, the second lightest neutralino tends to decay into the LSP and a photon.



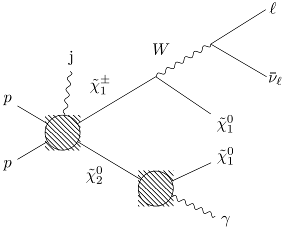
- Radiative decaying neutralinos at the LHC are highly suppressed by backgrounds (yet unexplored experimentally).
- Final state already phenomenologically explored within the MSSM, for wino-likes $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$, and bino-like $\tilde{\chi}_1^0$, with $\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 \ell \nu_\ell$ and $\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 + \gamma$,

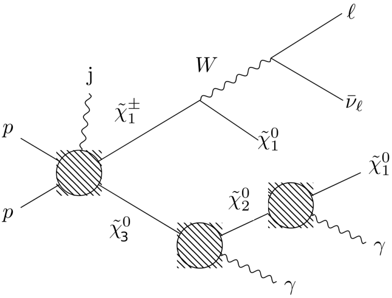
$$pp \to \tilde{\chi}_1^\pm \, \tilde{\chi}_2^0 j \to \tilde{\chi}_1^0 \, \ell \, \nu_\ell + \tilde{\chi}_1^0 \gamma + j \, .$$

Arganda, Carena, de los Rios, Perez, Rocha, RMSS, Wagner. [2410.13799]

Proposal

Search for radiative decaying neutralinos within the NMSSM, at $\sqrt{s} = 14$ TeV and a total integrated luminosity of $\mathcal{L} = 100$ fb^{-1}. We require a highly energetic ISR jet in association with the neutralino-chargino pair $\tilde{\chi}_1^\pm \tilde{\chi}_2^0 / \tilde{\chi}_1^\pm \tilde{\chi}_3^0$ to increase the MET signature.





Roy, Wagner, JHEP **04** (2024), 106

Use of ML techniques, as they would enable access to most of the otherwise inaccessible and unexplored parameter space.

Arganda, Carena, de los Rios, Perez, Rocha, RMSS, Wagner. [2410.13799]

Signal

$$p\,p \to \tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0 j/\tilde{\chi}_1^{\pm}\,\tilde{\chi}_3^0 j,\;\; \tilde{\chi}_1^{\pm} \to \tilde{\chi}_1^0 \ell \nu_\ell,\;\; \tilde{\chi}_3^0 \to \tilde{\chi}_2^0 + \gamma,\;\; \tilde{\chi}_2^0 \to \tilde{\chi}_1^0 + \gamma$$

BP	$m_{\tilde{\chi}_1^0,\chi_2^0,\chi_3^0}$ [GeV]	$m_{\tilde{\chi}_1^\pm}$ [GeV]	Br($\tilde{\chi}_2^0 \to \tilde{\chi}_1^0\gamma$), Br($\tilde{\chi}_3^0 \to \tilde{\chi}_2^0\gamma$), Br($\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 W^{\pm\star}$)	$\sigma(pp \to \tilde{\chi}_1^\pm\tilde{\chi}_2^0 j)$, $\sigma(pp \to \tilde{\chi}_1^\pm\tilde{\chi}_3^0 j)$ ($P_T(j) > 100GeV$, $\lvert\eta_j\rvert < 2.5$) (fb)	σ_{DD}^{SI} (cm^2)	$\Omega_{\tilde{\chi}_1^0}h^2$
1-1	133.6, 158.5, 164.8	161.8	0.28, 0.79, 0.98	105.1, 99.1	2.5×10^{-49}	7.5
1-2	141, 158.5, 164.8	161.8	0.47, 0.87, 0.98	105.1, 99.1	6.5×10^{-49}	0.9
1-3	147.5, 158.5, 164.8	161.8	0.73, 0.89 0.96	105.1, 99.1	1.3×10^{-48}	0.01
1-4	152.7, 158.5, 164.8	161.8	0.91, 0.89 0.88	105.1, 99.1	2.26×10^{-48}	0.028
2-1	143.2, 173.8, 180.2	161.8	0.21, 0.68, 0.98	81, 76.6	2.2×10^{-49}	13.8
2-2	151.7, 173.8, 180.2	177.2	0.36, 0.83, 0.98	81, 76.6	6.2×10^{-49}	2.96
2-3	156.1, 173.8, 180.3	177.2	0.48, 0.87, 0.98	81, 76.6	9.9×10^{-49}	0.55
2-4	161.9, 173.8, 180.3	177.2	0.7, 0.89, 0.96	81, 76.6	1.92×10^{-48}	0.11
2-5	168.9, 173.8, 180.6	177.2	0.94, 0.89, 0.89	81, 76.6	3.99×10^{-48}	0.0244
3-1	153.7, 184.0, 190.4	187.5	0.22, 0.72, 0.98	67.9, 65	3.95×10^{-49}	11.6
3-2	160.5, 184.0, 190.4	187.5	0.33, 0.83, 0.98	67.9, 65	7.3×10^{-49}	3.45
3-3	174.5, 184.0, 190.6	187.5	0.78, 0.9, 0.96	67.9, 65	2.76×10^{-48}	0.07
3-4	180.6, 184.0, 190.9	187.5	0.97, 0.89, 0.70	67.9, 65	5.1×10^{-48}	0.018

BP	$m_{\tilde{\chi}_1^0,\chi_2^0,\chi_3^0}$ [GeV]	$m_{\tilde{\chi}_1^\pm}$ [GeV]	Br($\tilde{\chi}_2^0 \to \tilde{\chi}_1^0\gamma$), Br($\tilde{\chi}_3^0 \to \tilde{\chi}_2^0\gamma$), Br($\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 W^{\pm\star}$)	$\sigma(pp \to \tilde{\chi}_1^\pm\tilde{\chi}_2^0 j)$, $\sigma(pp \to \tilde{\chi}_1^\pm\tilde{\chi}_3^0 j)$ ($P_T(j) > 100GeV$, $\lvert\eta_j\rvert < 2.5$) (fb)	σ_{DD}^{SI} (cm^2)	$\Omega_{\tilde{\chi}_1^0}h^2$
9-1	222.5, 245.0, 251.7	248.9	0.41, 0.90, 0.98	28.4, 27.1	1.6×10^{-48}	0.51
9-2	227.9, 245.0, 251.7	248.9	0.57, 0.91, 0.96	28.4, 27.1	2.3×10^{-48}	0.21
9-3	235.1, 245.0, 251.7	248.9	0.82, 0.91, 0.92	28.4, 27.1	4.2×10^{-48}	0.07
9-4	240.3, 245.0, 252.1	248.9	0.96, 0.91, 0.76	28.4, 27.1	6.8×10^{-48}	0.03
10-1	235.4, 255.1, 261.9	259.1	0.5, 0.90, 0.98	25, 23.9	2.02×10^{-48}	0.3
10-2	239.2, 255.1, 262.0	259.1	0.62, 0.91, 0.96	25, 23.9	2.64×10^{-48}	0.17
10-3	244.8, 255.1, 262.0	259.1	0.81, 0.91, 0.93	25, 23.9	4.2×10^{-48}	0.07
10-4	250.3, 255.1, 262.3	259.1	0.96, 0.91, 0.75	25, 23.9	6.7×10^{-48}	0.03
11-1	244.8, 265.2, 272.1	269.3	0.48, 0.9, 0.97	22, 21.1	2.0×10^{-48}	0.32
11-2	250.6, 265.2, 272.1	269.3	0.67, 0.91, 0.95	22, 21.1	3.0×10^{-48}	0.13
11-3	254.5, 265.2, 272.2	269.3	0.8, 0.91, 0.93	22, 21.1	4.1×10^{-48}	0.077
11-4	258.3, 265.2, 272.4	269.3	0.91, 0.91, 0.87	22, 21.1	5.8×10^{-48}	0.045
11-5	262.1, 265.2, 272.6	269.3	0.98, 0.90, 0.62	22, 21.1	8.1×10^{-48}	0.026

- NMSSM, for $\tilde{\chi}_1^0$ singlino-like, and $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$ and $\tilde{\chi}_1^{\pm}$ higgsino-like.
- In all cases $m_{\tilde{\chi}_1^0} \sim m_{\tilde{\chi}_2^0} \sim m_{\tilde{\chi}_3^0} \sim m_{\tilde{\chi}_1^{\pm}}$.
- BPs representative of the compressed region. Only a few produce the cosmological relic density. A few BPs are naively excluded by CMS multilepton searches (the ones with the lowest $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0}$).

Backgrounds

- Event generation with MadGraph5, Pythia8 and Delphes.
- Baseline Event selection criteria: at least one charged light lepton ($\ell = e, \mu$), at least one photon, and at least one jet. Leading jet with $p_T > 100$ GeV and $E_T^{\text{miss}} > 100$ GeV.

Process	Yield
$W +$ jets	60058
$W\gamma$	58462
$t\bar{t} +$ jets	18051
$Z +$ jets	3360
$t\bar{t}\gamma$	2498
Diboson	2340
Total background	147983

BP #	Yield $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0 + \tilde{\chi}_1^{\pm}\,\tilde{\chi}_3^0$	$S/\sqrt{B}$
1-3	$263 + 223$	1.26
3-3	$200 + 169$	0.96
4-4	$185 + 162$	0.90
2-4	$161 + 142$	0.79

Kinematic variables

Simple set of variables to characterize the kinematics of the studied final state including low-level detector variables (p_T, η and ϕ of the leading objects $\{j_1, \ell_1, \gamma_1\}$, E_T^{miss}, and object multiplicities), and several high-level observables:

$$H_T^{\text{jets}} = \sum p_T^{\text{jets}},$$

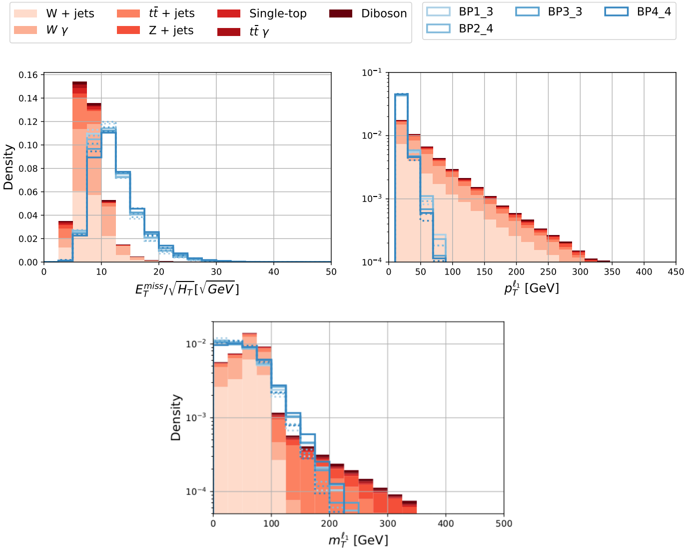
$$H_T = \sum p_T^{\text{jets}} + \sum p_T^{\tau} + \sum p_T^{e} + \sum p_T^{\mu} + \sum p_T^{\gamma},$$

$$m_T^A \equiv m_T\left(\mathbf{p}_T(A), \mathbf{E}_T^{\text{miss}}\right) = \sqrt{2p_T(A)E_T^{\text{miss}}\left(1 - \cos\Delta\phi\left(\mathbf{p}_T(A), \mathbf{E}_T^{\text{miss}}\right)\right)},$$

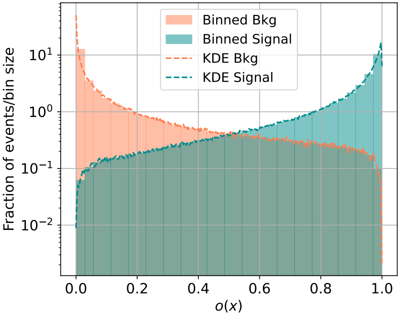
$$s_T^1 = p_T^{\ell_1} + p_T^{j_1} + p_T^{\gamma_1},$$

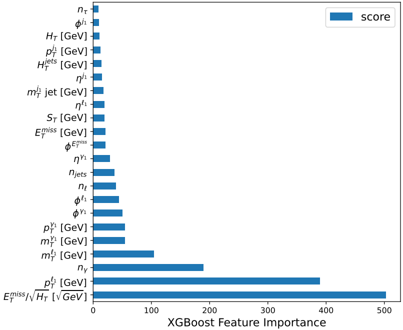
$$E_T^{\text{miss}}/\sqrt{H_T}.$$

Most important kinematic variables for discrimination:



- Training of supervised `XGBoost` classifier, with balanced dataset, and all low and high-level variables as input features.

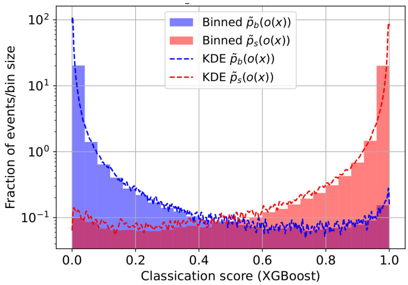




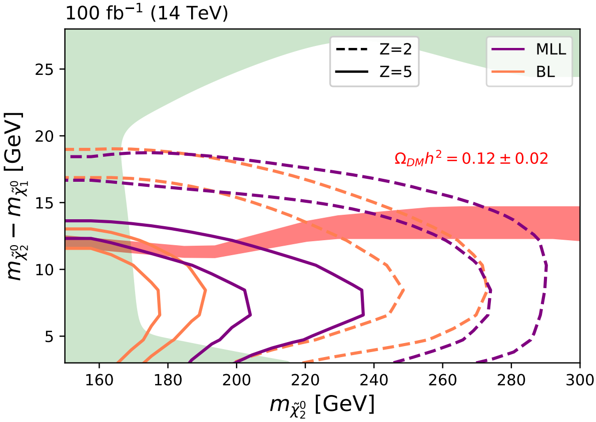
- All BPs included in the training dataset (60 in total), with $m_{\tilde{\chi}_2^0} \in [150, 290]$ GeV and $m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0} \in [3, 30]$ GeV.

- Main analysis adding a ML binary classifier as they better capture the underlying physics and correlations, and increase the sensitivity of signal over background. Discovery significance reported for two different approaches:
  - Binned Likelihood (BL) method: histogram-based analysis of the ML output.
  - Machine-Learned Likelihoods (MLL) method: unbinned fit of the ML output with Kernel Density Estimators -KDE.

Arganda, de los Rios, Perez, RMSS *Eur. Phys. J. C* 83, no.12, 1158 (2023)



Projected discovery significance in the $[m_{\tilde{\chi}_2^0}, m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0}]$ plane:

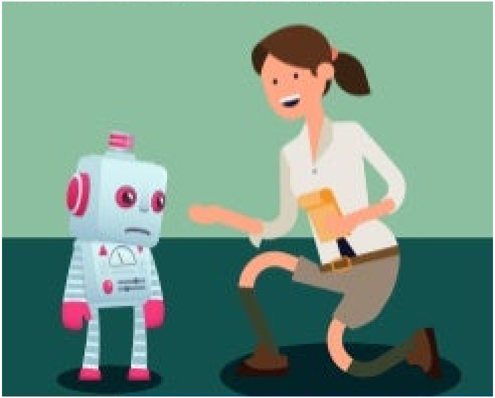


- Both the binned and unbinned ML strategies yields promising results.
- Proof-of-concept results, not systematic uncertainties included or CMS multilepton exclusion limits.

Conclusions

- We explore an alternative final state for searching neutralinos within the NMSSM, including photons and a hard ISR jet, never explored at the LHC.
- This channel dominates where the direct DM detection cross-section is suppressed. We assume $m_{\tilde{\chi}_1^0} \sim m_{\tilde{\chi}_2^0} \sim m_{\tilde{\chi}_3^0} \sim m_{\tilde{\chi}_1^\pm}$, with $\tilde{\chi}_1^0$ being singlino-like, and $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$ and $\tilde{\chi}_1^\pm$ higgsino-like.
- The significance of these type of searches for the HL-LHC is greatly improved using machine learning methods (otherwise inaccessible).
- Projected significances for the HL-LHC are promising, in line with previous study for the MSSM.
- Future outlook: inclusion of CMS multilepton searches and systematic uncertainties.

Thank you!



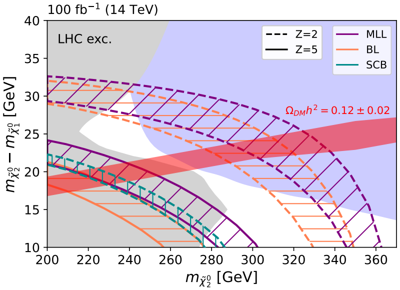
More about the NMSSM model:

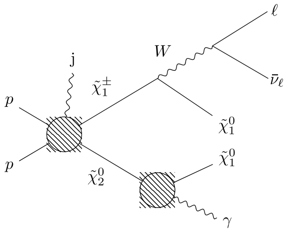
- M_1 (Bino mass parameter) fixed to $\sim$500 GeV. M_2 (Wino mass parameter) fixed to $\sim$2500 GeV.
- Mass of all gluinos and squarks set to 3 TeV.
- $\tan(\beta) = 6$.
- $m_h \sim 125$ GeV.
- Blind spot condition:

$$\left(m_{\tilde{\chi}_1^0} + \frac{g_1^2 v^2}{M_1 - m_{\tilde{\chi}_1^0}} \right) \frac{1}{\mu_{\text{eff}} \sin 2\beta} \simeq 1 \,.$$

Previous results for the MSSM

Projected discovery significance in the $[m_{\tilde{\chi}_2^0}, m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0}]$ within the MSSM, for wino-likes $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$, and bino-like $\tilde{\chi}_1^0$, with $\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 \ell \nu_\ell$ and $\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 + \gamma$:





Arganda, Carena, de los Rios, Perez, Rocha, RMSS, Wagner. [2410.13799]

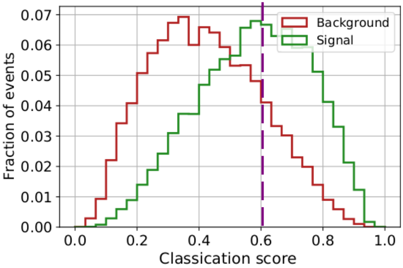
About the MSSM model:

- M_1 (Bino mass parameter), M_2 (Wino mass parameter) fixed to 500 GeV. and μ (Higgsino mass parameter) all takes values of a hundred GeV with $|M_1| \leqslant |M_2| \leqslant |\mu|$ (compressed spectra).
- Also $M_2 \times \mu > 0$ (prefered by $(g-2)$), and $M_1 \times \mu < 0$.
- Mass of all gluinos and squarks set to 2.5 TeV (generation universal).
- $\tan(\beta) = 50$.
- $m_h \sim 125$ GeV and $m_A = 2.5$ TeV.

Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... $\longrightarrow$ ML classifiers ✓
- For experimental significances, selection cuts $\longrightarrow$ Working points ✗



Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... $\longrightarrow$ ML classifiers ✓
- For experimental significances, selection cuts $\longrightarrow$ Working points ✗

Is it possible to connect the ML classifier output with the standard statistical tests without defining working points?

$\longrightarrow$ Machine-Learned Likelihood (MLL) Method

Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... $\longrightarrow$ ML classifiers ✓
- For experimental significances, selection cuts $\longrightarrow$ Working points ✗

Is it possible to connect the ML classifier output with the standard statistical tests without defining working points?

$\longrightarrow$ Machine-Learned Likelihood (MLL) Method

Can we avoid the information loss from binning the output?

$\longrightarrow$ +Kernel Density Estimators (KDE)

Method: Machine-Learned Likelihood

E. Arganda, X. Marcano, V. Martín Lozano, A. D. Medina, A. D. Perez, M. Szewc, A. Szynkman

Eur. Phys. J. C **82**, no.11, 993 (2022)

A method for approximating optimal statistical significances with machine-learned likelihoods

Ernesto Arganda,[a,b*] Xabier Marcano,[a,c†] Víctor Martín Lozano,[d,e‡] Anibal D. Medina,[b§] Andres D. Perez,[b¶] Manuel Szewc[f,g‖] and Alejandro Szynkman[b**]

[a]Instituto de Física Teórica UAM-CSIC, C/ Nicolás Cabrera 13-15, Campus de Cantoblanco, 28049, Madrid, Spain
[b]IFLP, CONICET - Dpto. de Física, Universidad Nacional de La Plata, C.C. 67, 1900 La Plata, Argentina
[c]Departamento de Física Teórica, Universidad Autónoma de Madrid , E-28049 Cantoblanco, Madrid, Spain
[d]Departament de Física Teòrica and IFIC, Universitat de València-CSIC,

E. Arganda, M. de los Rios, A. D. Perez, RMSS

PoS **ICHEP2022** (2022) 1226

PoS PROCEEDINGS OF SCIENCE

Imposing exclusion limits on new physics with machine-learned likelihoods

Ernesto Arganda,[a,b] Martín de los Rios,[a,c] Andres D. Perez[b] and Rosa María Sandá Seoane[a,*]
[a]Instituto de Física Teórica UAM-CSIC, C/ Nicolás Cabrera 13-15, Campus de Cantoblanco, 28049, Madrid, Spain
[b]IFLP, CONICET - Dpto. de Física, Universidad Nacional de La Plata,

E. Arganda, M. de los Rios, A. D. Perez, RMSS

Eur. Phys. J. C **83**, no.12, 1158 (2023)

Machine-Learned Exclusion Limits without Binning

Ernesto Arganda [a,b*] Andres D. Perez,[b†] Martin de los Rios[a,c‡] and Rosa María Sandá Seoane[a§]

[a]Instituto de Física Teórica UAM-CSIC, C/ Nicolás Cabrera 13-15, Campus de Cantoblanco, 28049, Madrid, Spain
[b]IFLP, CONICET - Dpto. de Física, Universidad Nacional de La Plata, C.C. 67, 1900 La Plata, Argentina
[c]Departamento de Física Teórica, Universidad Autónoma de Madrid, E-28049 Cantoblanco, Madrid, Spain

The MLL method

Statistical model for N independent measurements, with a high-dimensional set of observables x

$$\mathcal{L}(\mu, s, b) = p(N, \{x_i, i = 1, ..., N\} | \mu, s, b) \equiv \text{Poiss}\big(N | \mu S + B\big) \prod_{i=1}^{N} p(x_i | \mu, s, b)$$

where S (B) is the expected total signal (background) yield, and

$$p(x|\mu, s, b) = \frac{B}{\mu S + B}\, p_b(x) + \frac{\mu S}{\mu S + B}\, p_s(x)$$

The relevant test statistic to derive discovery significances corresponds to $\mu = 0$

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2 \text{ Ln } \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N} \text{ Ln } \left(1 + \frac{\hat{\mu} S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu} S\, p_s(x_i) + B\, p_b(x_i)} = 1$$

The MLL method

Statistical model for N independent measurements, with a high-dimensional set of observables x

$$\mathcal{L}(\mu, s, b) = p(N, \{x_i, i = 1, ..., N\}|\mu, s, b) \equiv \text{Poiss}\big(N|\mu S + B\big) \prod_{i=1}^{N} p(x_i|\mu, s, b)$$

where S (B) is the expected total signal (background) yield, and

$$p(x|\mu, s, b) = \frac{B}{\mu S + B}\, p_b(x) + \frac{\mu S}{\mu S + B}\, p_s(x)$$

The relevant test statistic to derive discovery significances corresponds to $\mu = 0$

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2 \text{ Ln } \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N} \text{ Ln } \left(1 + \frac{\hat{\mu}S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu}S\, p_s(x_i) + B\, p_b(x_i)} = 1$$

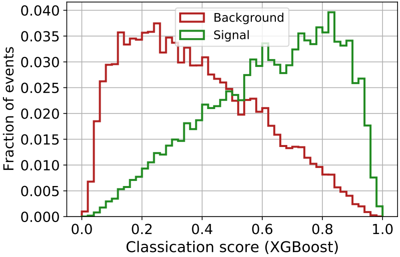
Solution: train classifier to distinguish signal from bckg with a balanced dataset. The classification score maximizes the binary cross-entropy and thus approaches

$$o(x) = \frac{p_s(x)}{p_s(x) + p_b(x)}$$

Dimensional reduction by dealing with $o(x)$ instead of x

$$p_s(x) \to \tilde{p}_s(o(x)), \qquad \text{and} \qquad p_b(x) \to \tilde{p}_b(o(x))$$

Cranmer et al, arXiv: 1506.02169



where $\tilde{p}_{s,b}(o(x))$ are the distributions of $o(x)$ for signal and background, obtained by evaluating the classifier on a set of pure signal or background events, respectively.

The relevant test statistic for exclusion limits

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2\ \mathrm{Ln}\ \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N}\ \mathrm{Ln}\ \left(1 + \frac{\hat{\mu}S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

with $\hat{\mu}$ such us

$$\sum_{i=1}^{N} \frac{\tilde{p}_s(o(x_i))}{\hat{\mu}S\,\tilde{p}_s(o(x_i)) + B\,\tilde{p}_b(o(x_i))} = 1$$

The median expected discovery significance when the true hypothesis is assumed to be the signal-plus-background ($\mu' = 1$) is

$$\mathrm{med}\ [Z_0|1] = \sqrt{\mathrm{med}\ [\tilde{q}_0|1]}$$

Summary of MLL

- MLL method allows to obtaining exclusion (and discovery) significances for additive new physics scenarios.
- Uses a single `XGBoost` classifier and its full 1D output (no working points), which allows the estimation of the S and B pdfs needed for statistical inference. Not strictly necessary to bin the output to extract the PDFs.
- Inclusion of KDE as an extension of the MLL method to avoid the binning of the ML classifier output.
- Improves results obtained by traditional techniques in toy models and realistic analysis, approaching (when possible) the ones computed with true generative functions.
- Possible improvements: unsupervised analysis, systematic uncertainties...

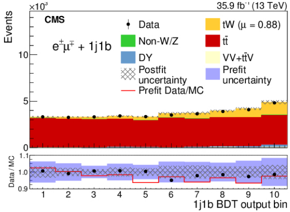
Traditional Binned-Likelihood (BL) method

$p_{s,b}(x)/\tilde{p}_{s,b}(o(x_i))$ are not known and are approximated by discrete binned distributions

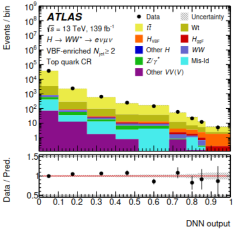
$$\mathcal{L}(\mu, s, b) = \prod_{d=1}^{D} \text{Poiss}\left(N_d | \mu S_d + B_d\right)$$

The median exclusion significance using Asimov datasets is given by

$$\text{med}[Z_\mu|0] = \sqrt{\tilde{q}_\mu|0} = \left[2\sum_{d=1}^{D}\left(B_d \ln\left(\frac{B_d}{S_d+B_d}\right) + S_d\right)\right]^{1/2} \xrightarrow[\sqrt{B}\gg 1]{S\ll B} \frac{S}{\sqrt{B}}$$



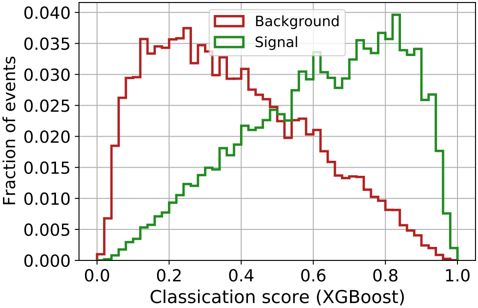
J HEP **10** (2018) 117



arXiv: 2207.00338

Density Estimation

What is the best way to extract $\tilde{p}_s(o(x))$ and $\tilde{p}_b(o(x))$?



Density Estimation

⟶Density estimation in a sense is the reverse of sampling: from given samples we want to retrieve the density function from which the samples were generated.

⟶Two types of methods for density estimation

- Parametric: model the density function as a specified functional form with a fixed number of tunable parameters.
- Non-parametric: specify a model whose complexity grows with the number of training datapoints.

Kernel Density Estimators

Kernel Density Estimators (KDE) is a non-parametric method for extracting $\tilde{p}_s(o(x_i))$ and $\tilde{p}_b(o(x_i))$

$\longrightarrow$Smoothed version of the empirical distribution $q_o(x)$ of the training data $\{x_i, i = 1, ..., N\}$

$$q_o(x) \;= \frac{1}{N}\sum_{i}^{N}\delta\left(x - x_i\right)$$

$\longrightarrow$We can smooth out the empirical distribution and turn it into a density by replacing each delta distribution with a smoothing kernel

$$\kappa_\epsilon(u) = \frac{1}{\epsilon^D}\,\kappa_1\left(\frac{u}{\epsilon}\right)$$

where $\epsilon > 0$ (bandwidth parameter) controls the width of the kernel and $\kappa_1(u)$ is a density function bounded from above (as $\epsilon \rightarrow 0$, $\kappa_\epsilon(u)$ approaches $\delta(u)$)

$$q_\epsilon(x) \;= \frac{1}{N}\sum_{i}^{N}\kappa_\epsilon\left(x - x_i\right)$$

$$\tilde{p}_{s,b}(o(x)) = \frac{1}{N}\sum_{i}^{N} \kappa_\epsilon \left[o(x) - o(x_i)\right]$$

Several options for κ_ϵ, e.g.

$$\kappa_\epsilon(u) = \begin{cases} \frac{1}{\epsilon}\frac{3}{4}\left(1-(u/\epsilon)^2\right), & \text{if } |u| \leqslant \epsilon \\ 0 & \text{otherwise} \end{cases}$$ *Epanechnikov kernel*