Estimation of Machine Learning model uncertainty in particle physics event classifiers

Miguel Cárdenas-Montes[1] miguel.cardenas@ciemat.es

José M. Hernández[1] jose.hernandez@ciemat.es **Julia Vázquez Escobar**[1] julia.vazquez@ciemat.es

[1]Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (**CIEMAT**)

AI goes MAD^2 (IFT,UAM)

October 15, 2024

CIEMAT física de partículas

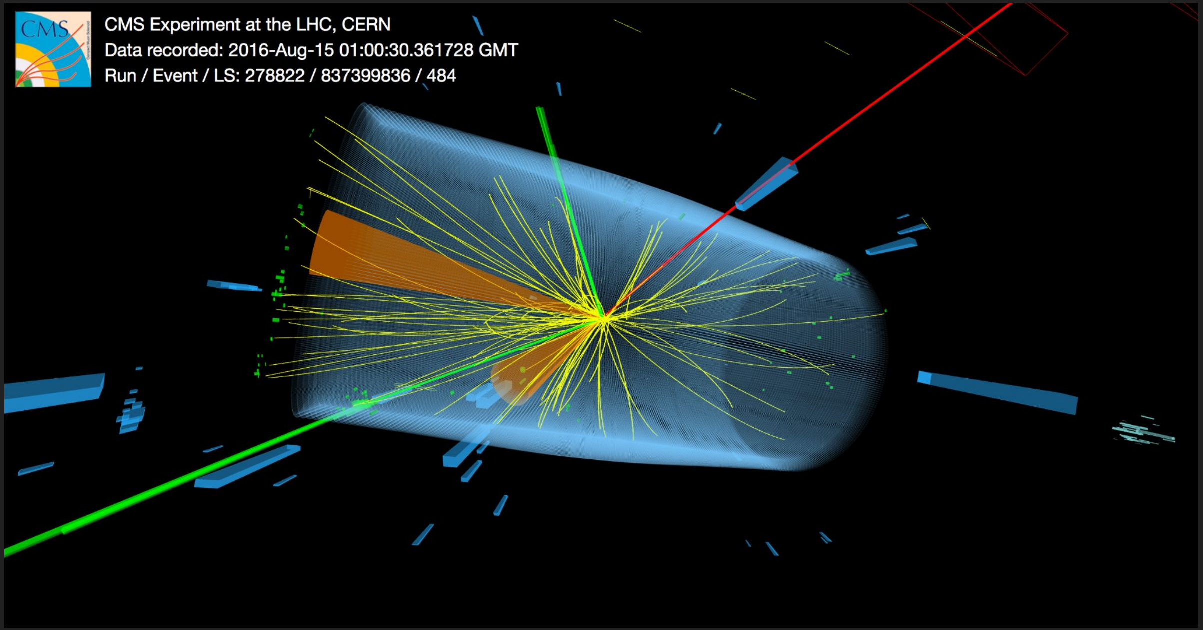
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

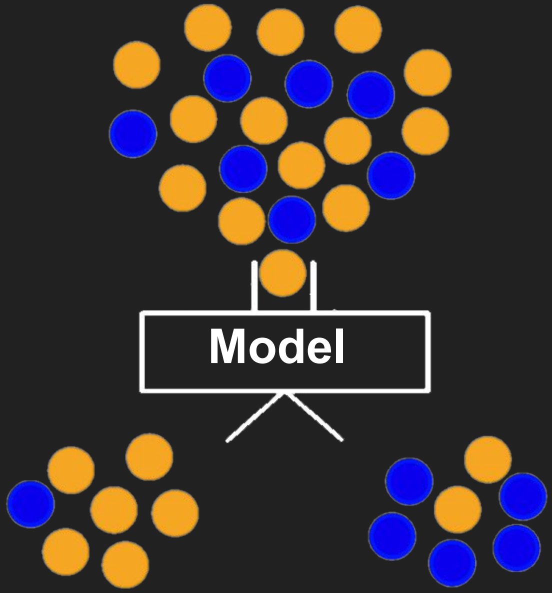
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

Paper link

Data collection in CMS experiment

Tens of petabytes of data are recorded every year at the Large Hadron Collider (LHC).

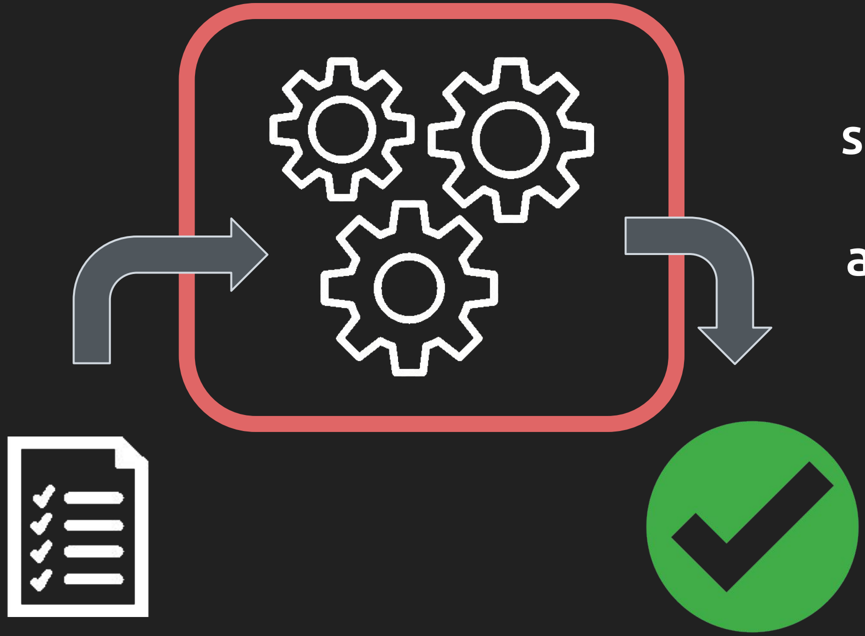




The identification of tiny amounts of signal inside a huge dataset is a complex task.

Machine Learning use for classification problem

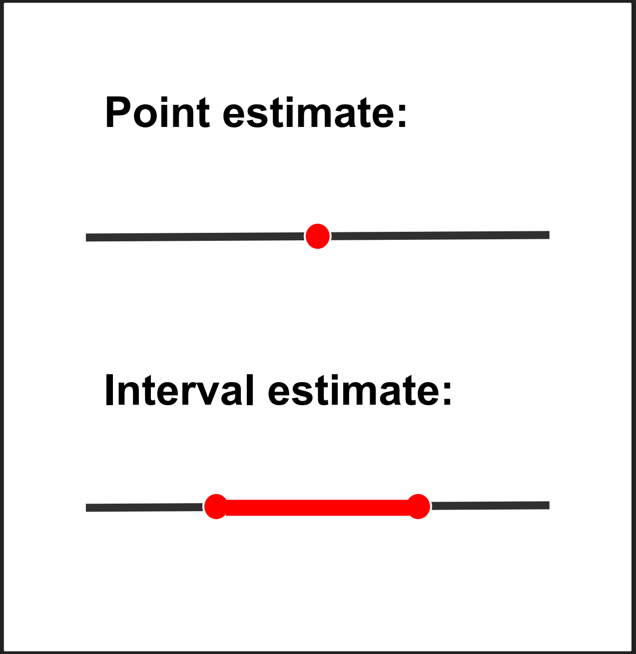
Algorithm able to improve its performance by learning from data



We measure its success with the use of specific and appropriate metrics

What about uncertainty for ML?

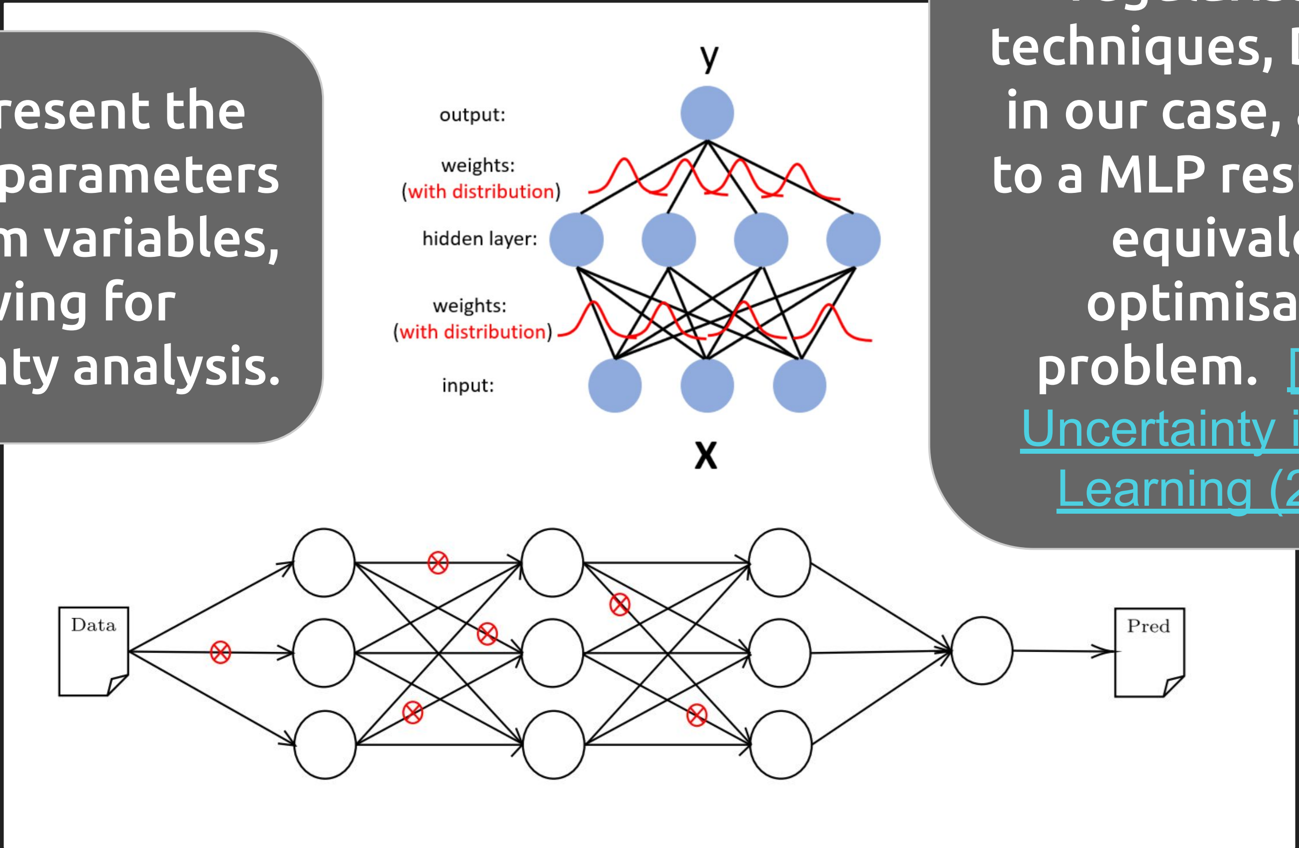
We can obtain great results with Machine learning methods, but they are usually presented as mere point estimations. Uncertainty measures such as variance offer more complete results.



Are my results reliable ?

Bayesian Approximation

BNNs present the network parameters as random variables, allowing for uncertainty analysis.



Stochastic regularisation techniques, Dropout in our case, applied to a MLP result in an equivalent optimisation problem. [Y. Gal, Uncertainty in Deep Learning (2016)]

Bayesian Approximation

Proposition *Given* $p(\mathbf{y}^*|\mathbf{f}^\omega(\mathbf{x}^*)) = \mathcal{N}(\mathbf{y}^*; \mathbf{f}^\omega(\mathbf{x}^*), \tau^{-1}\mathbf{I})$ *for some* $\tau > 0$, $\mathbb{E}_{q_\theta^*(\mathbf{y}^*|\mathbf{x}^*)}[\mathbf{y}^*]$ *can be estimated with the unbiased estimator*

$$\widetilde{\mathbb{E}}[\mathbf{y}^*] := \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}^{\widehat{\omega}_t}(\mathbf{x}^*) \xrightarrow[T\to\infty]{} \mathbb{E}_{q_\theta^*(\mathbf{y}^*|\mathbf{x}^*)}[\mathbf{y}^*]$$

with $\widehat{\boldsymbol{\omega}}_t \sim q_\theta^*(\boldsymbol{\omega})$.

We can obtain an unbiased estimator of the first and second moment!!

Proposition

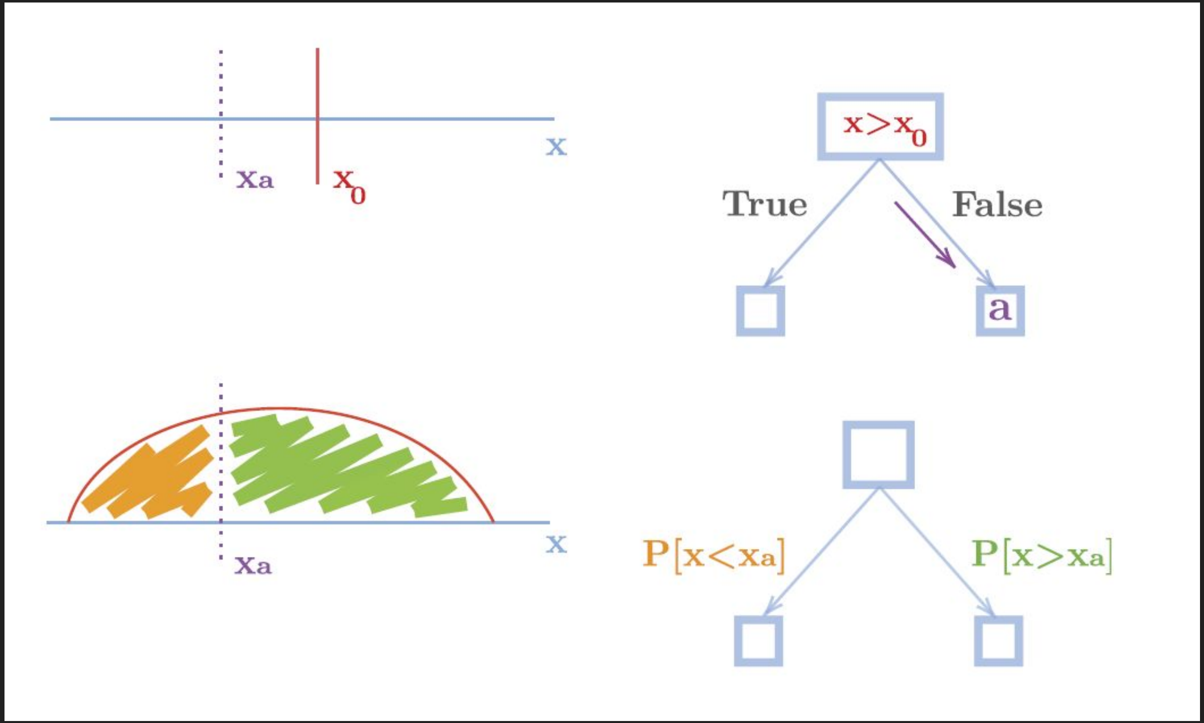
Given $p(\mathbf{y}^*|\mathbf{f}^\omega(\mathbf{x}^*)) = \mathcal{N}(\mathbf{y}^*; \mathbf{f}^\omega(\mathbf{x}^*), \tau^{-1}\mathbf{I})$ *for some* $\tau > 0$, $\mathbb{E}_{q_\theta^*(\mathbf{y}^*|\mathbf{x}^*)}\Big[(\mathbf{y}^*)^T(\mathbf{y}^*)\Big]$ *can be estimated with the unbiased estimator*

$$\widetilde{\mathbb{E}}\Big[(\mathbf{y}^*)^T(\mathbf{y}^*)\Big] := \tau^{-1}\mathbf{I} + \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}^{\widehat{\omega}_t}(\mathbf{x}^*)^T \mathbf{f}^{\widehat{\omega}_t}(\mathbf{x}^*) \xrightarrow[T\to\infty]{} \mathbb{E}_{q_\theta^*(\mathbf{y}^*|\mathbf{x}^*)}\Big[(\mathbf{y}^*)^T(\mathbf{y}^*)\Big]$$

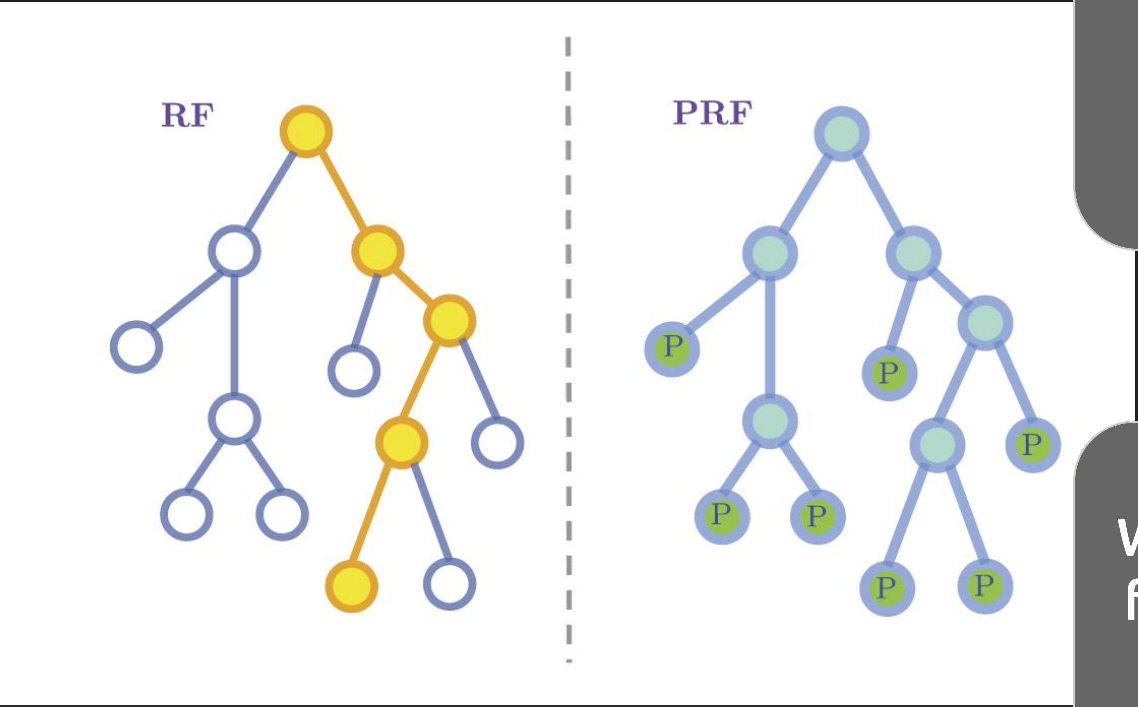
with $\widehat{\boldsymbol{\omega}}_t \sim q_\theta^*(\boldsymbol{\omega})$ *and* $\mathbf{y}^*, \mathbf{f}^{\widehat{\omega}_t}(\mathbf{x}^*)$ *row vectors (thus the sum is over the outer-products).*

Probabilistic Random Forest

The Random Forest algorithm is modified to take into account uncertainties in the input data.



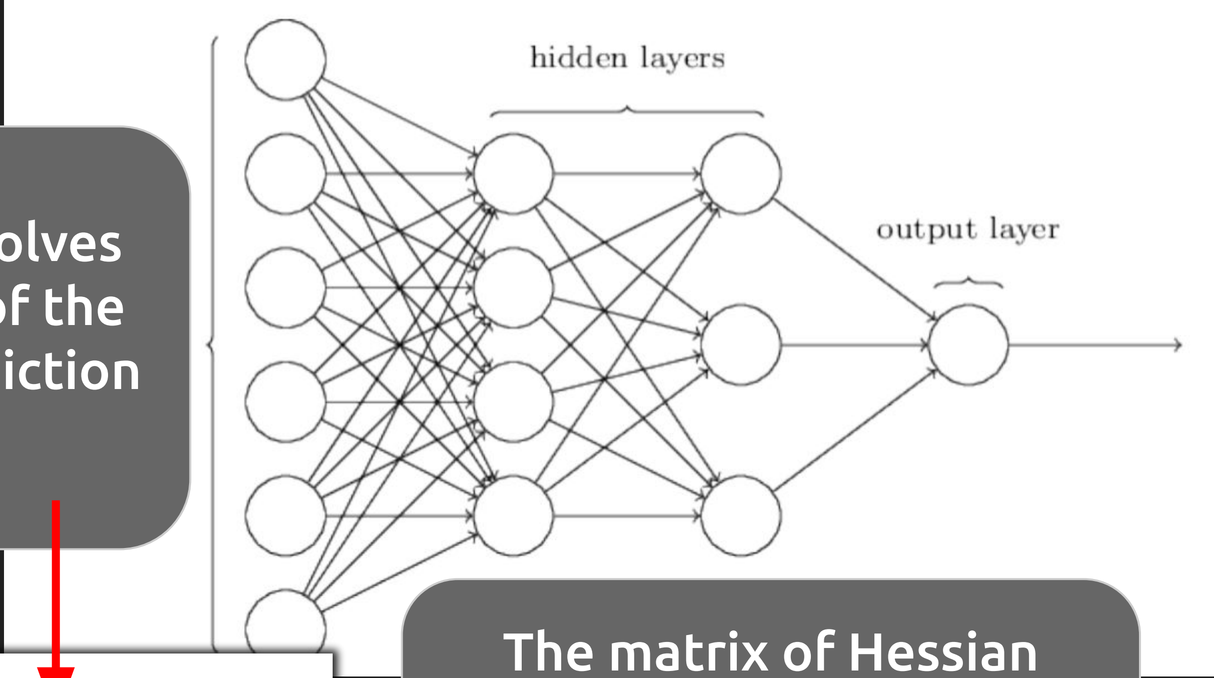
Probabilistic Random Forest



Instead of a deterministic outcome PRF calculates the probability of each possibility.

We can compute then the first and second moment of the prediction.

Local Ensembles



This method involves the calculation of the norm of the prediction gradient.

$$\mathcal{E}_m(x') = \|U_m^\top g_{\theta^\star}(x')\|_2$$

The matrix of Hessian eigenvectors spanning a subspace of low curvature is also needed.

Local Ensembles

The calculated score for a point *x'* is proportional to the standard deviation of the prediction for *x'*. [D. Madras, J. Atwood, A. D'Amour, in: ICLR 2020 International Conference on Learning Representations, (2019)]

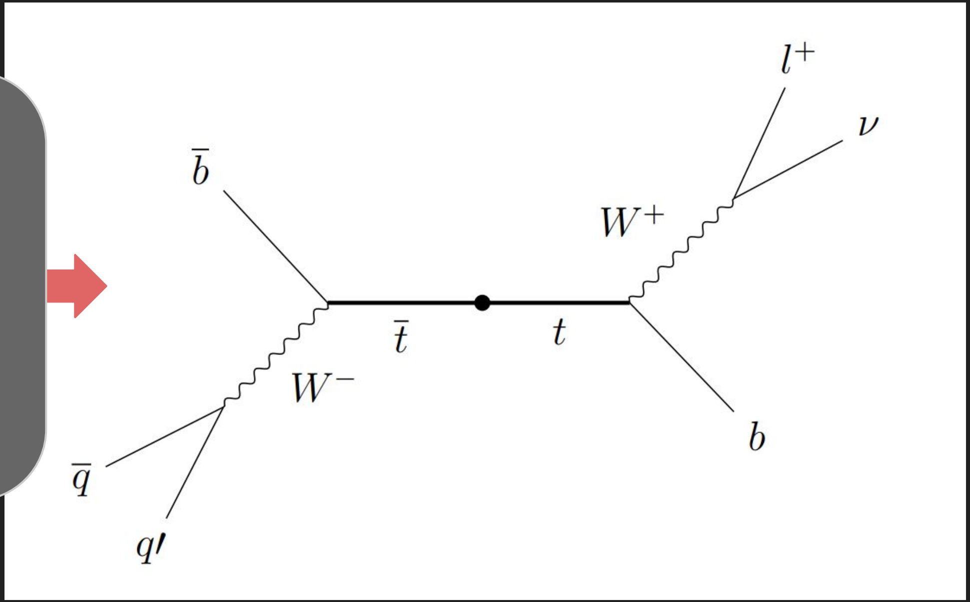
Proposition 1. *Let Δ_θ be the projection of a random perturbation with mean zero and covariance proportional to the identity $\epsilon \cdot I$ into the ensemble subspace spanned by $\{\xi_{(j)} : j > m\}$. Let P_Δ be the linearized change in prediction induced by the perturbation*

$$P_\Delta(x') := g_{\theta^\star}(x')^\top \Delta_\theta \approx \hat{y}(x', \theta^\star + \Delta_\theta) - \hat{y}(x', \theta^\star).$$

Then $\mathcal{E}_m(x') = \epsilon^{-1/2} \cdot SD(P_\Delta(x'))$.

Top antitop discrimination in CMS

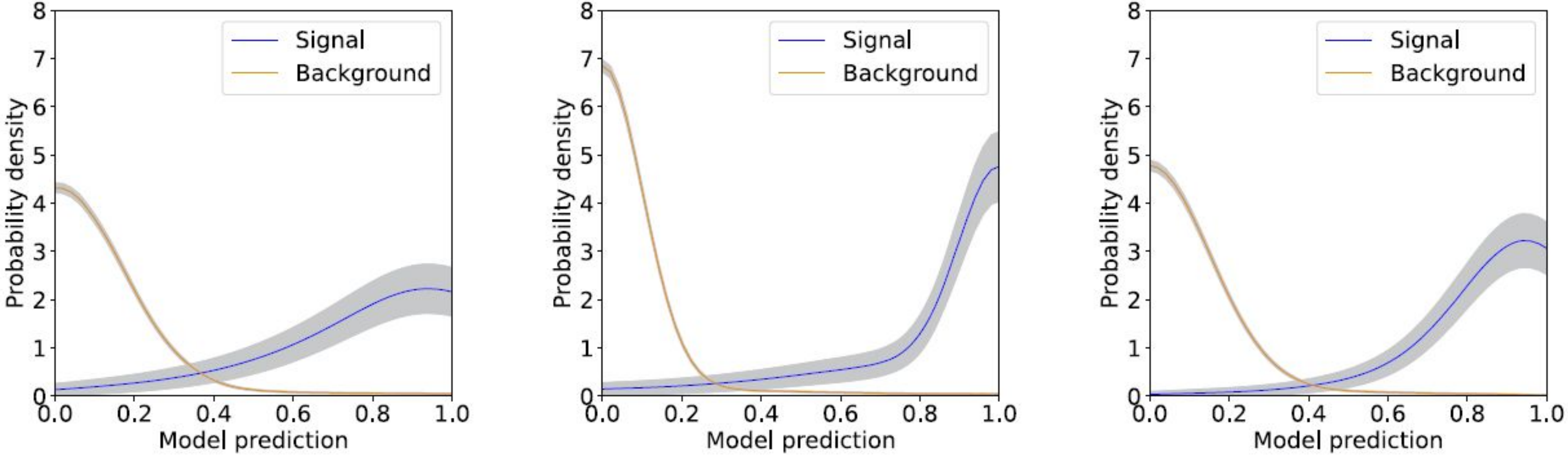
The three methods are applied to a classification problem: Discrimination of top antitop decays.



Background includes WW, WZ, ZZ or single top processes.

Results

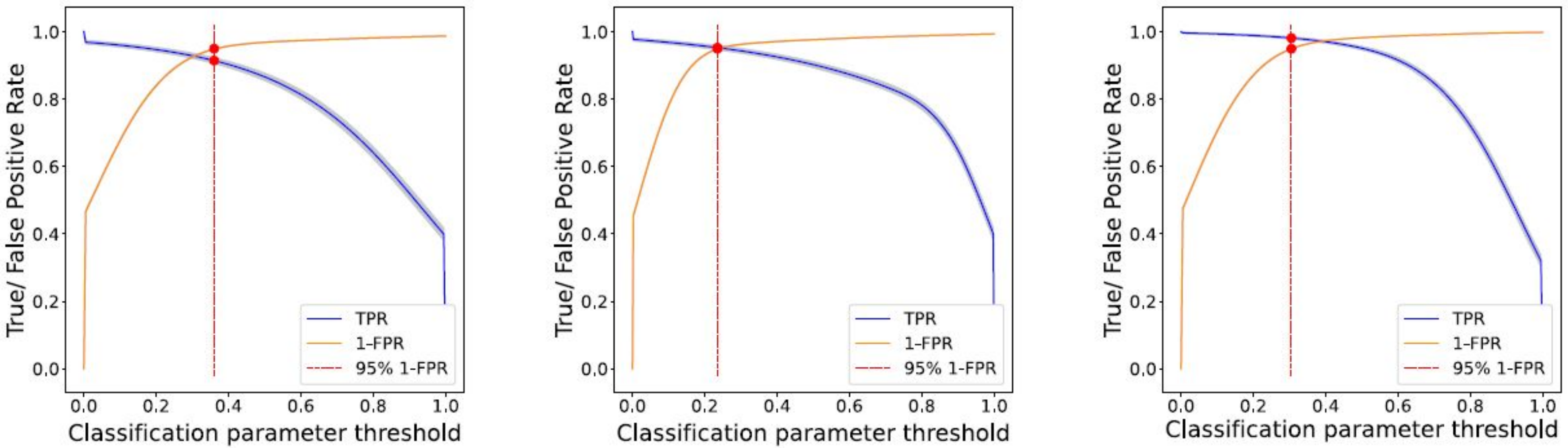
The probability density functions of the classification parameter for true signal and background events are shown.



(a) Local Ensembles (b) Probabilistic Random Forest (c) Bayesian approximation

Results

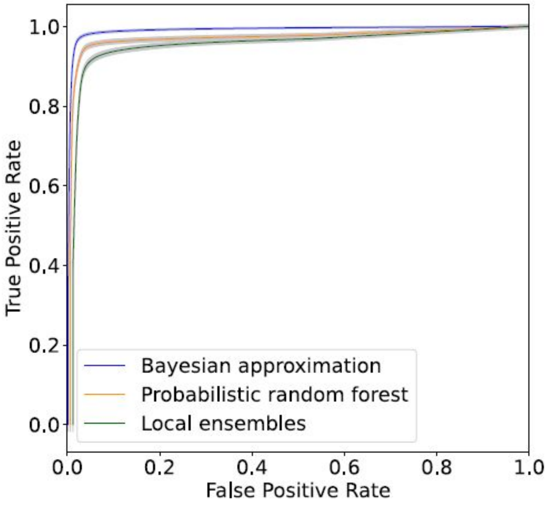
Model	TPR for $1-\text{FPR}=0.95$
Probabilistic Random Forest	0.953 ± 0.010
Local Ensembles	0.915 ± 0.008
Bayesian Approximation	0.982 ± 0.003



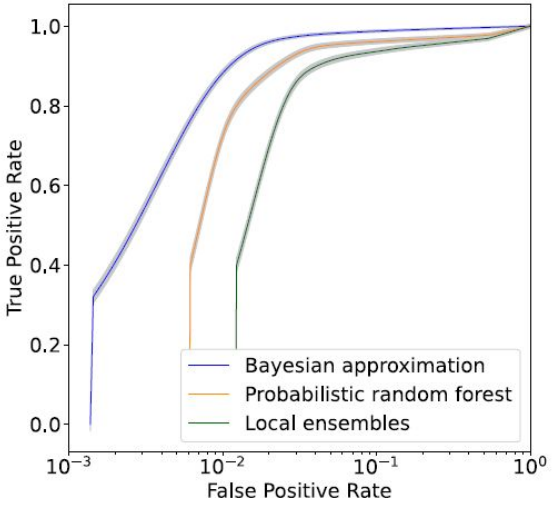
(a) Local Ensembles (b) Probabilistic Random Forest (c) Bayesian approximation

Results

Model	AUC
Probabilistic Random Forest	0.969 ± 0.005
Local Ensembles	0.951 ± 0.006
Bayesian Approximation	0.990 ± 0.001



(a) ROC curve.



(b) ROC curve with x-axis in logarithmic scale.

Conclusions

- All of them exhibit an excellent discrimination power.
- The model uncertainty measure turns out to be small, showing that the predictions are precise and robust.
- Prediction uncertainties supplied by the three methods are quite similar.
- This work foresees the prospect of generalizing the use of ML uncertainty methods in particle physics. The huge volume of data in the field makes it ideal to exploit the possibilities of these tools.

The End