Evaluate the Artificial Intelligence in Particle Physics

Iñaki Rodríguez-García

inaki.rodriguez@ciemat.es
CIEMAT Dark Matter Group
on behalf of the DEAP-3600 Collaboration

MultiDark18, Palos de la Frontera, Huelva
October 18, 2021

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

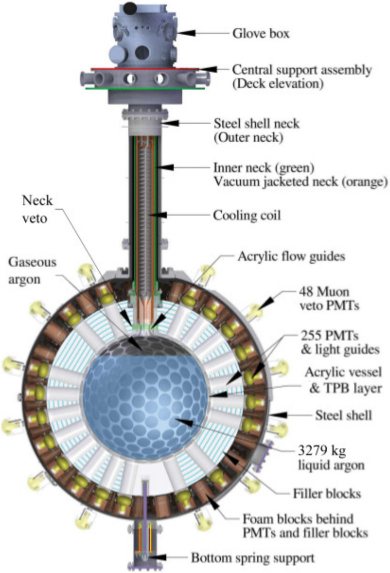
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

cfp CIEMAT física de partículas

DEAP 3600

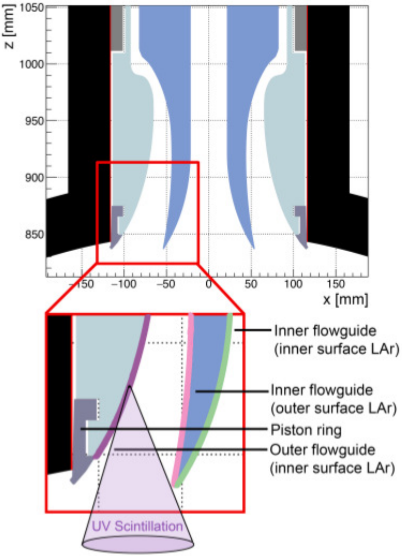
DEAP-3600 detector

- Placed in SNOLAB under 2 km of rocks ($\approx$ 6000 mwe of coverage).
- Spherical acrylic vessel with diameter of 170 cm. It is filled with 3279 kg Liquid Argon.
- Single phase detector with 255 PMTs covering the 75% of the sphere. Grouped in 35 rings of 5 or 10 PMTs.
- Liquid-Gas interface 55 cm above the *"equator"*.



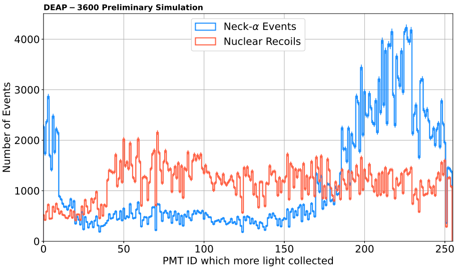
Neck alpha events background

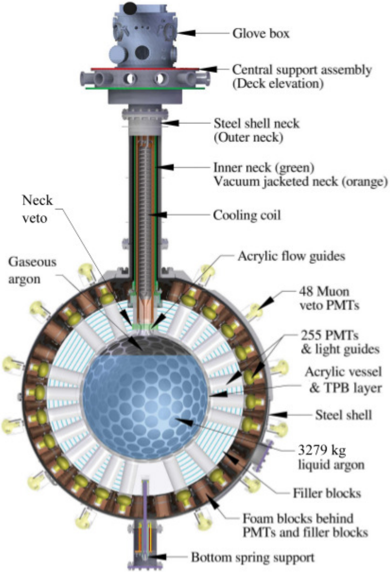
- The neck on top of the detector has two acrylic flow-guides.
- Their surfaces are covered by a thin film of LAr due to condensation, where α scintillation can happen.
- The light reaching the PMTs is a small fraction of a typical alpha event and its detection in the southern hemisphere is favoured.
- This implites that its light-pattern mimics a possible WIMP recoil on an Ar nucleus.



Neck alpha events background

- Potential paths in that direction: small fraction is reflected in the interface. Also a high proportion arrives in the PMTs of the southern hemisphere.
- Approaches based on machine learning are exploited to remove neck events.
- Goal:improving the Acceptance for Background Rejection of 99.9%.





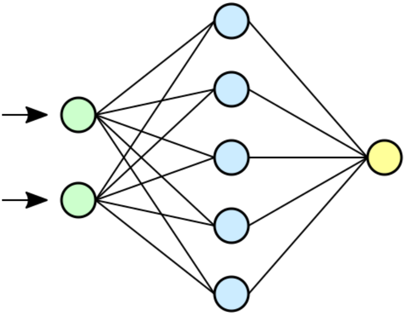
Simplified Neural Network

- Model is based on neural network with one hidden layer composed by 100 neurons.
- It isn't the best performing solution.
- However its architecture is suitable for explainable methodologies implementation.
- Input: light pattern of the 255 PMTs.

$$q = (q_0, q_1, q_2, ..., q_{254})$$

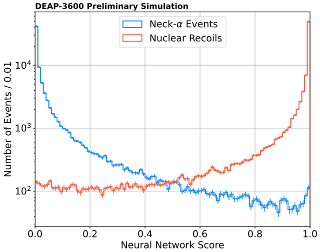
- Scaled input improves the algorithm performance.

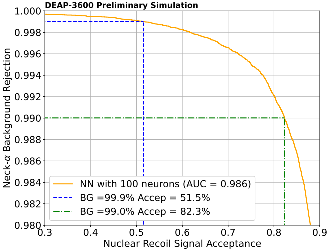
$$x_i = \frac{q_i - \min(q)}{\max(q) - \min(q)} = \frac{q_i}{\max(q)}$$



Neural Network

- Trained with 205K MC events of each class, and evaluated with 87K of each one:
  - Nuclear recoils (**signal**) labeled as 1.
  - Neck events (**background**) labeled as 0.
- Goal: maximizing the Acceptance for Background Rejection of 99.9%.
  - **Result:** 51.5%.





Neural Network II

- Input: light pattern normalized to maximum value (minimum is zero):

$$x_i = \frac{q_i - \min(q)}{\max(q) - \min(q)} = \frac{q_i}{\max(q)}$$

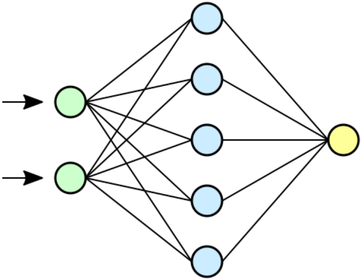
- Some necessary maths:
  - Hidden neurons expression:

$$a_n = f_h(w_n^0 + w_n^1 x_1 + w_n^2 x_2 + w_n^i x_i + \ldots + w_n^I x_I)$$

  - Activation function is $relu(z) = \max(0, z)$.
  - Output layer expression:

$$\hat{y} = \sigma(v^0 + v^1 a_1 + v^n a_n + \ldots + v^N a_N)$$

  - Activation function is *sigmoid*



Explainable Artificial Intelligence (XAI)

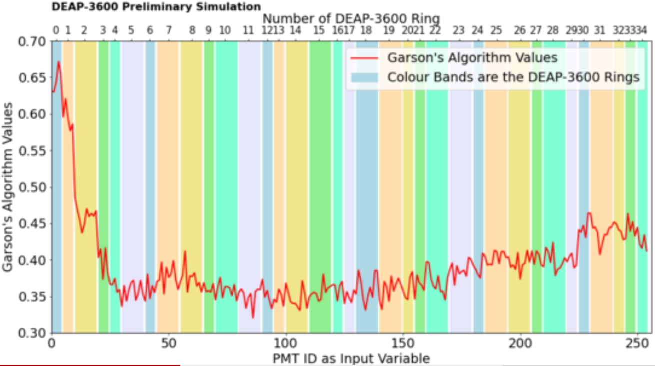
- Global interpretability: criteria based on underlying properties of the neural network.
  - Garson Algorithm, Connection Weight Approach.
- Local interpretability: criteria evaluated individually for each event. Individual explanations can be combined in one as whole of the algorithm.
  - SHAP.

Garson Algorithm

- Limited to neural networks with one single hidden layer.

$$c_i = \sum_{j=1}^{N} \frac{|w_j^i \cdot v_j|}{s_j} \ , \ s_n = \sum_{i=1}^{I} |w_n^i \cdot v_n| \text{ where } \begin{cases} w_n^i & \text{neuron n input i} \\ v_n & \text{out weight neuron n} \end{cases}$$

- Top PMTs (specially three first rings) are the most relevant ones.
- Little bump in the southern hemisphere (beyond ring 26).

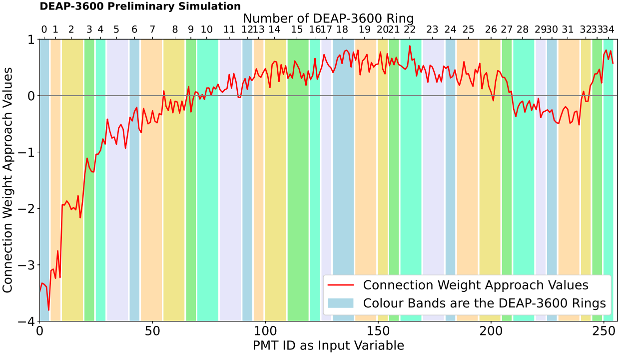


Connection Weights Approach

- Limited to neural networks with one single hidden layer as well.

$$c_i = \sum_{j=1}^{m} w_i^j \cdot v_j$$

- The sign can evidence the direction of influence, $c_i < 0 \rightarrow$ neck and $c_i > 0 \rightarrow$ argon.

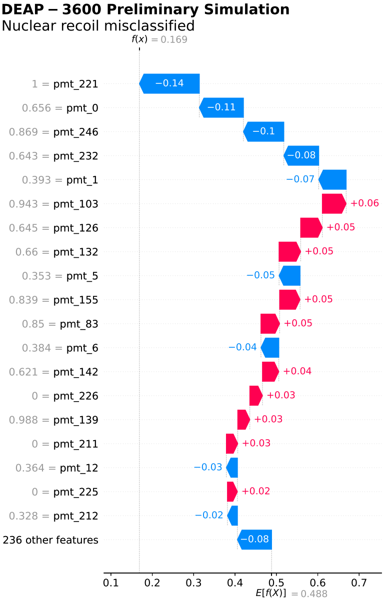


SHapley Additive exPlanations (SHAP)

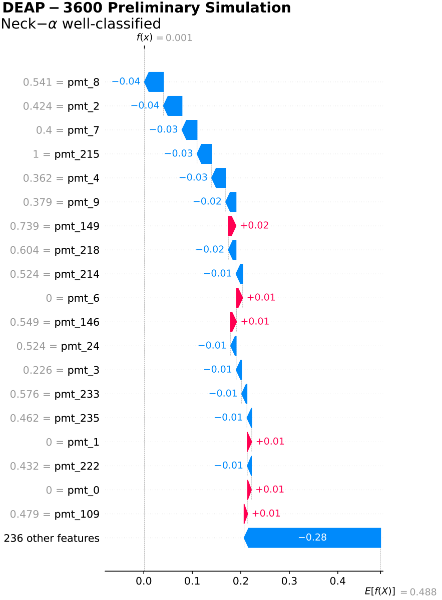
- Agnostic algorithm to retrieve interpretability for all ML models.
- SHAP calculates individual explanations for each event.
- It explains the difference between the output value and an expected value as:

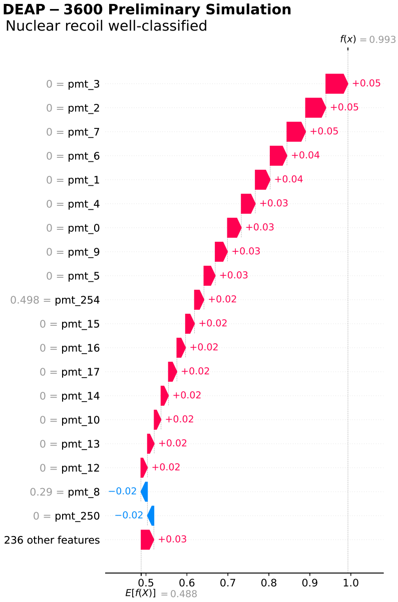
$$E[\hat{y}] + \sum_{n=0}^{N_{var}} SHAP_n(x_i) = \hat{y}_i \ \forall i$$

- The expected value is the mean of the model predictions in 5k events.



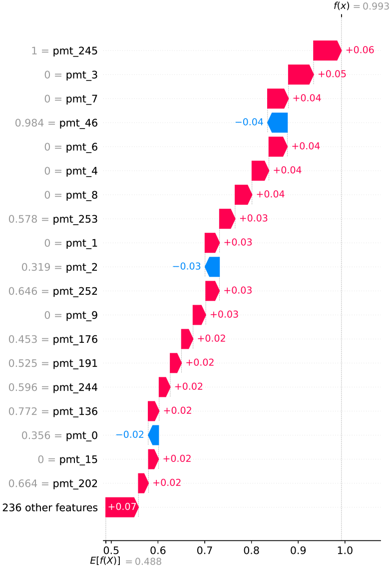
SHAP on well-classified examples



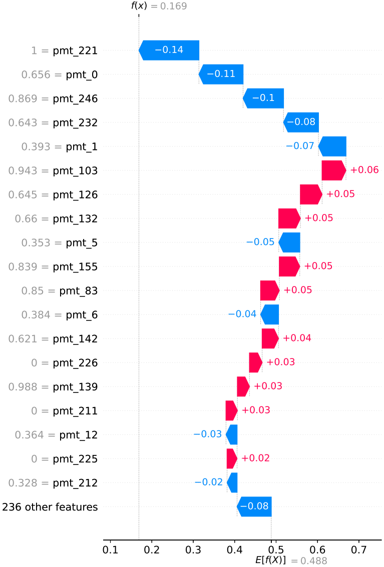


SHAP on misclassified examples

DEAP − 3600 Preliminary Simulation
Neck−α misclassified



DEAP − 3600 Preliminary Simulation
Nuclear recoil misclassified



Conclusions

- XAI methods are helpful to understanding how predictions are made, and discarding the idea of neural networks as black boxes.
- Bias detection on background classifiers.
- Understanding the prediction improves the event selection, and hence the acceptance increases. (**Under study**)
- Applying these methods to other algorithms in background problems is also under consideration.
- CIEMAT-DM group is increasing the XAI techniques in its portfolio of algorithms on DEAP-3600 and other problems.