Electron identification with Convolutional Neural Network in the ATLAS Experiment

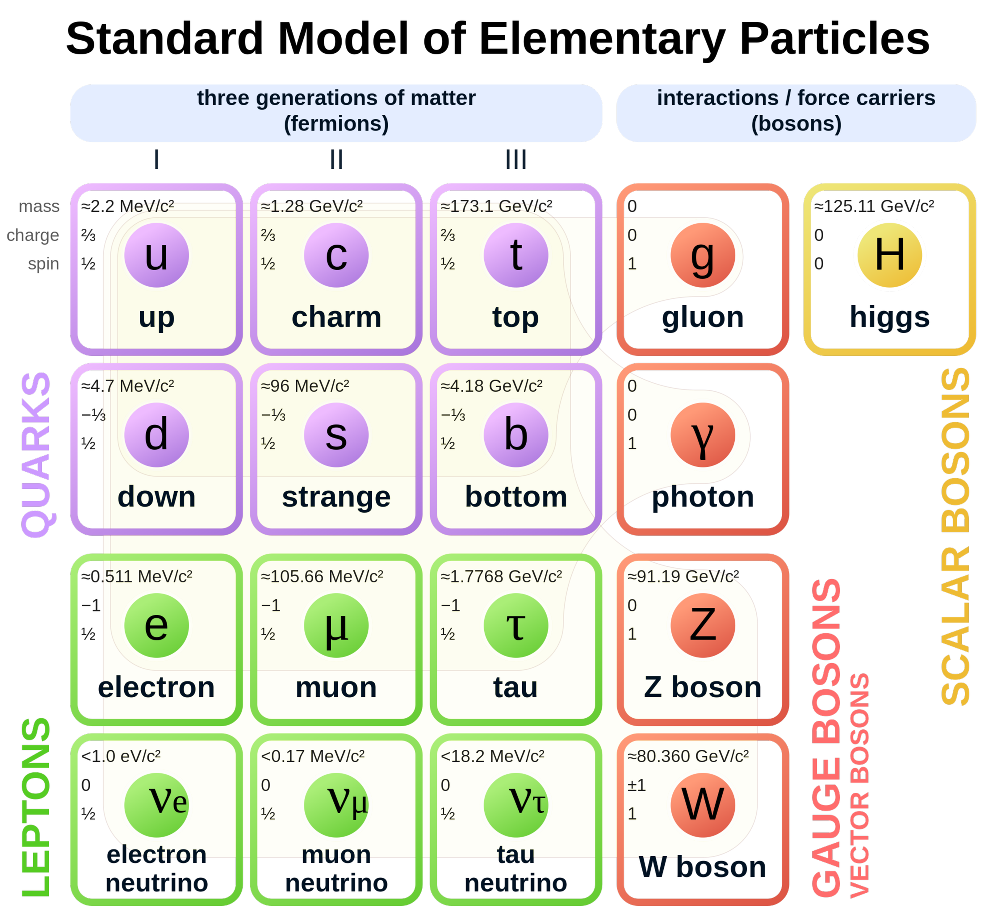
Jean-François Arguin, Olivier Denis, Bruna Pascual Dias, Dominique Godin, Hoang Dai Nghia Nguyen

6th Inter-experiment Machine Learning Workshop

ATLAS EXPERIMENT

Université de Montréal

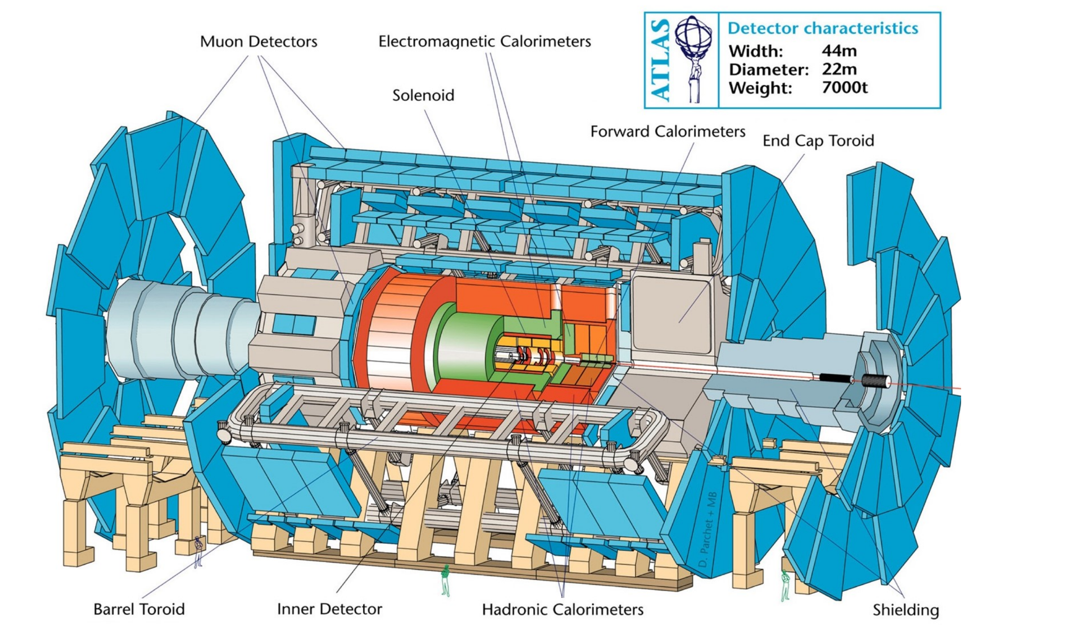
Standard Model (SM) of particle physics



- A theory that describes elementary particles (fermions) and their interactions (bosons)
  - Based on quantum mechanics + special relativity
- Fermions categorized into **quarks** and **leptons**
  - **Quarks**: sensitive to all SM interactions
  - **Leptons**: sensitive to EM and weak interactions
  - 3 generations: 3 copies with increasingly larger masses
- **Bosons** describe fundamental interactions
  - **Gluon** associates with **strong interaction**
  - **Z** and **W$^{\pm}$** bosons associate with **weak interaction**
  - **Photon** associates with **electromagnetic (EM) interaction**
  - **Higgs** boson generates mass for all massive particles
- The SM is experimentally established with high precision
- In 2012, the last piece of the particles in SM picture, the **Higgs** boson, was discovered by ATLAS + CMS at LHC
- However, SM is not yet a complete theory, unsolved questions remain
  - What is the composition of dark matter?
  - What is the origin of matter-antimatter asymmetry?
  - ...

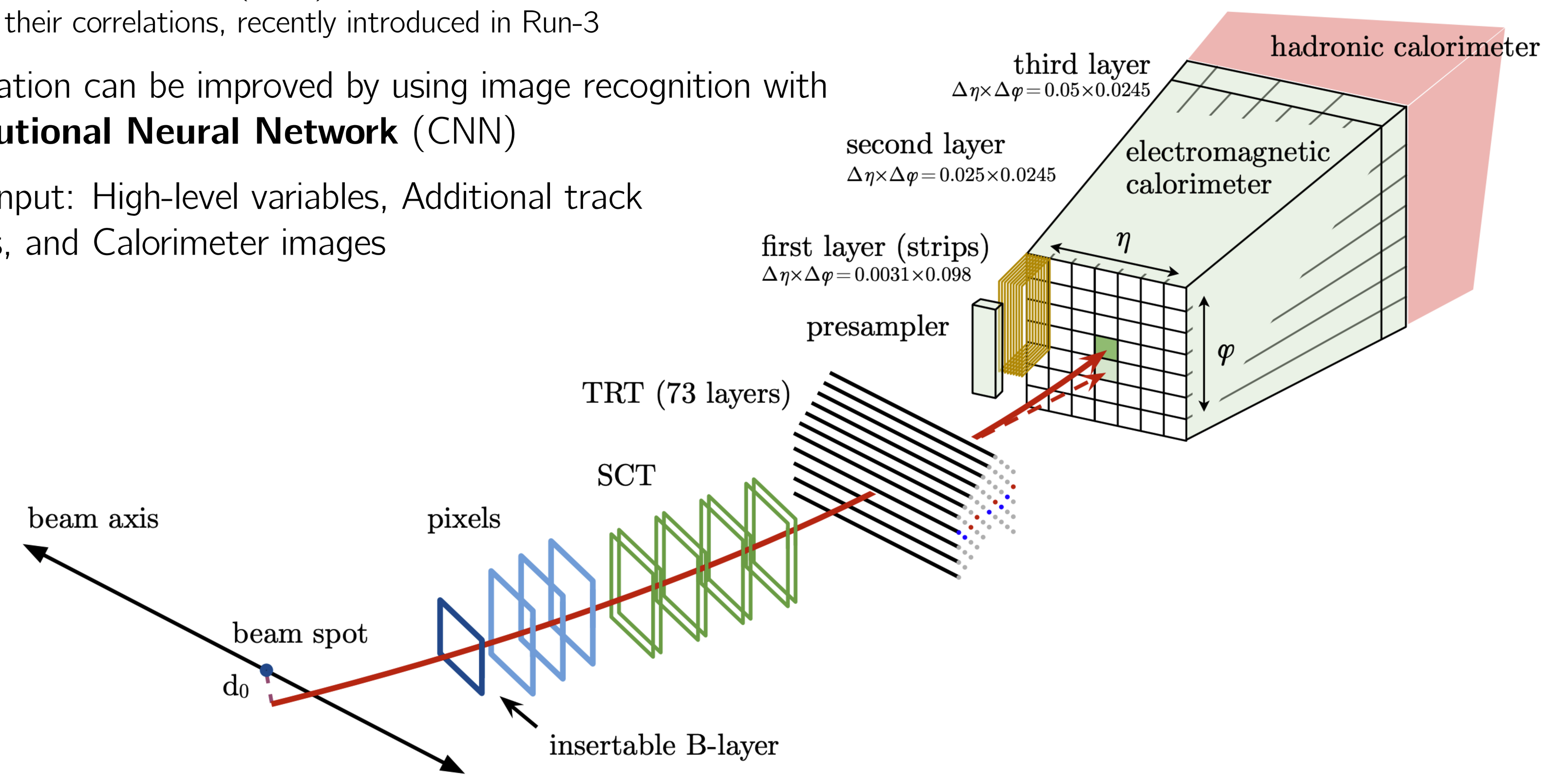
ATLAS detector

- A multi-purpose detector, designed to study Higgs boson physics, SM precision measurements and new physics searches
- Consists of the Inner Detector, Calorimeters and Muon Spectrometer
- Particles produced after collisions leave **different signatures at each layer** of the detector
- Each particle can be reconstructed and **identified** by combining these signatures from several sub-detectors



Electron identification

- Electron identification has an important role in a large fraction of ATLAS physics analyses
- ATLAS currently deployed two electron identification techniques:
  - **Likelihood** (LH): has been mainly used since 2012, takes shower shape, track and track-cluster variables as input
  - **Deep Neural Network** (DNN): uses same input variables as the LH, exploit their correlations, recently introduced in Run-3
- Identification can be improved by using image recognition with **Convolutional Neural Network** (CNN)
- CNN's input: High-level variables, Additional track variables, and Calorimeter images

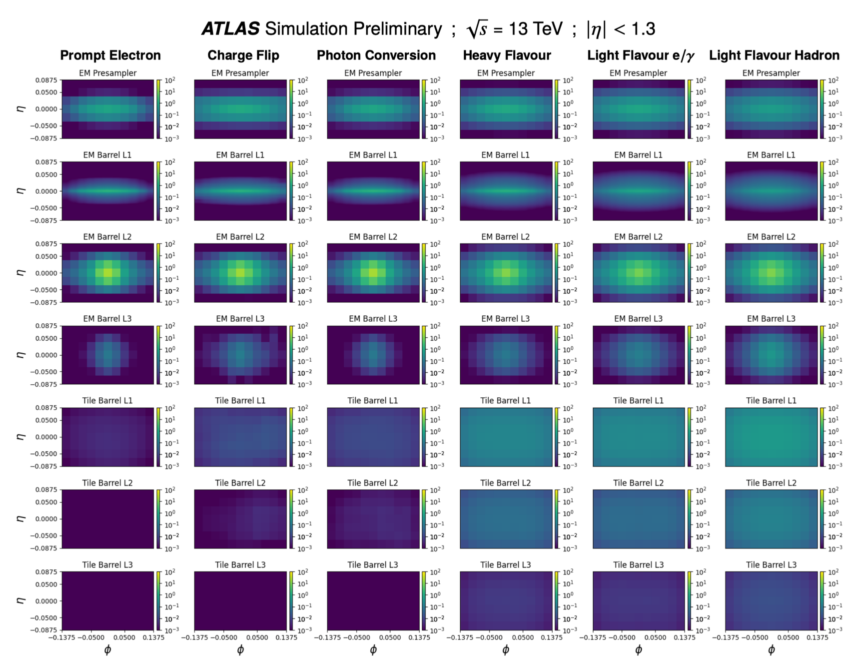


Input variables for CNN

- Electron candidates are split into six electron classes (**Prompt electrons**, Charge-flip, Photon conversion, Heavy flavor, Light flavor e/γ and Light flavor hadrons) based on their truth information
  - In most use-cases, charge-flip electrons are considered as signal, just like the prompt electron class
  - Other classes are considered as background
- Same **High-Level** inputs as LH and DNN plus two ECIDS variables developed for charge-flip identification
- **Additional tracks** contain important information that is used by the CNN algorithm for up to five tracks
- **Calorimeter images** represent the mean energy deposited in cells divided by the electron energy (in %)

CNN Input Variables for Each Track (up to five)

Type	Description	Symbol
Matching variables	Ratio of the momentum of track j to the energy of the electron candidate	p_j/E
	$\Delta\eta$ between the track j and the electron candidate position	$\Delta\eta_j$
	$\Delta\phi$ between the track j and the electron candidate position	$\Delta\phi_j$
Number of hits	Number of hits in the pixel detector	n^j_{Pixel}
	Number of hits in the SCT detector	n^j_{SCT}
	Number of hits in the TRT detector	n^j_{TRT}
Track parameters and fit quality	Transverse impact parameter relative to the beamline	d^j_0
	Uncertainty on d_0	$\sigma(d_0)_j$
	Longitudinal impact parameter relative to the beamline	z^j_0
	Charge of the track	q_j
	χ^2 of the track fit	χ^2_j
	Number of degrees of freedom of the track fit	ndof_j
	Matched vertex index number	vtx_j

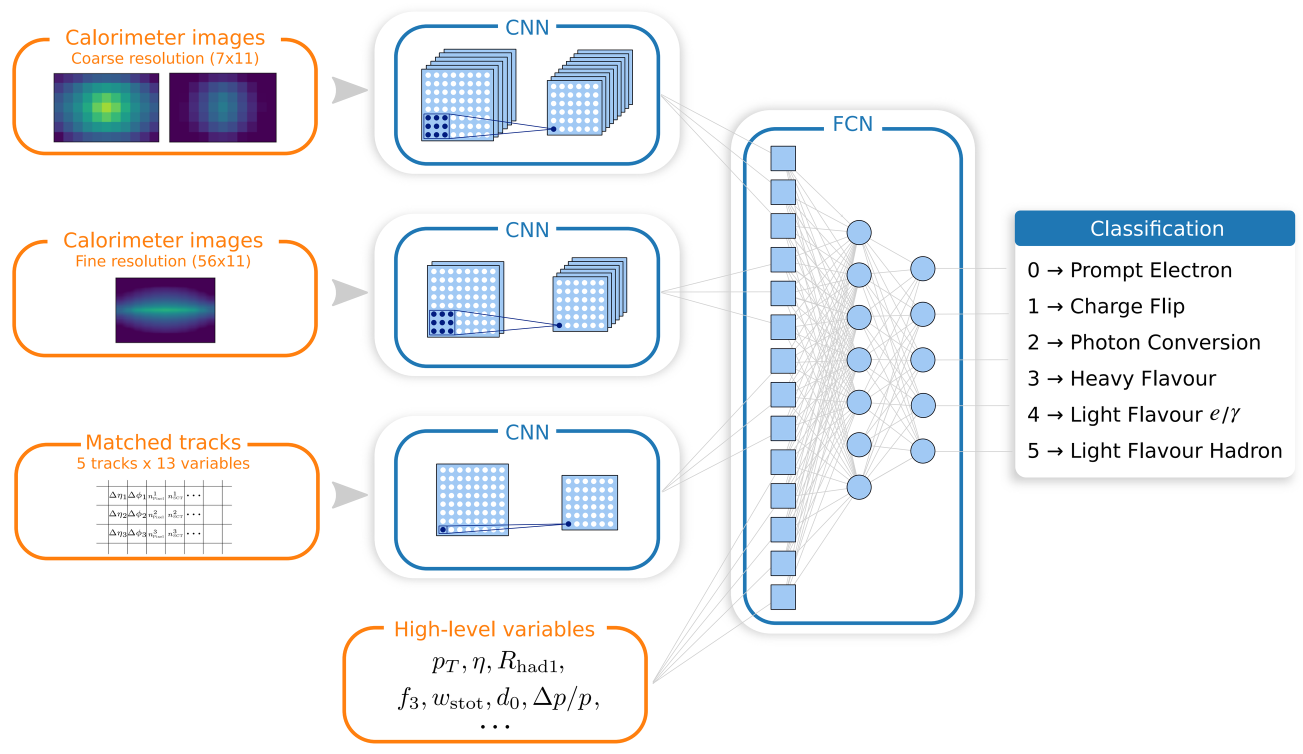


Convolutional Neural Network architecture

- Global architecture has 3 CNNs dedicated to coarse images, fine images and tracks
- First CNN processes images from EM L1 layer in resolution $\eta\times\phi = 56\times11$
- Second CNN processes all images from the calorimeter in resolution $\eta\times\phi = 7\times11$
- Third CNN uses a 1×1 kernel to process information of additional tracks
- CNNs outputs are concatenated with the high-level variables and linked to a FCN which outputs probability vectors
- Network trained with TensorFlow and Adam optimizer with an adaptative learning rate is used for gradient descent
- More powerful discriminant obtained from each electron class probabilities and their corresponding adjustable weights:

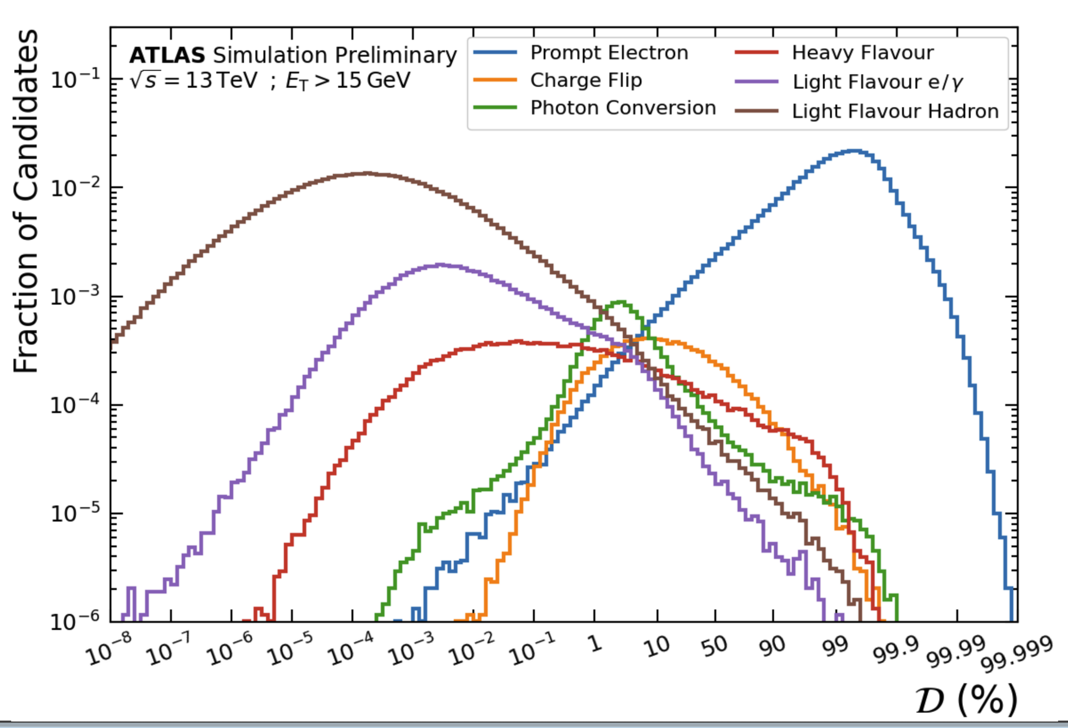
$$D = \frac{\sum(w_{\text{sig}}p_{\text{sig}})}{\sum(w_{\text{bkg}}p_{\text{bkg}})} \quad \rightarrow \text{Can be transformed and bounded between 0 and 1 with } D\rightarrow D/(D+1)$$

- Values of w_{sig} and w_{bkg} can be customized to target specific background rejection
  - Perform the best when w_{sig} and w_{bkg} correspond to the actual fractions of the validation sample

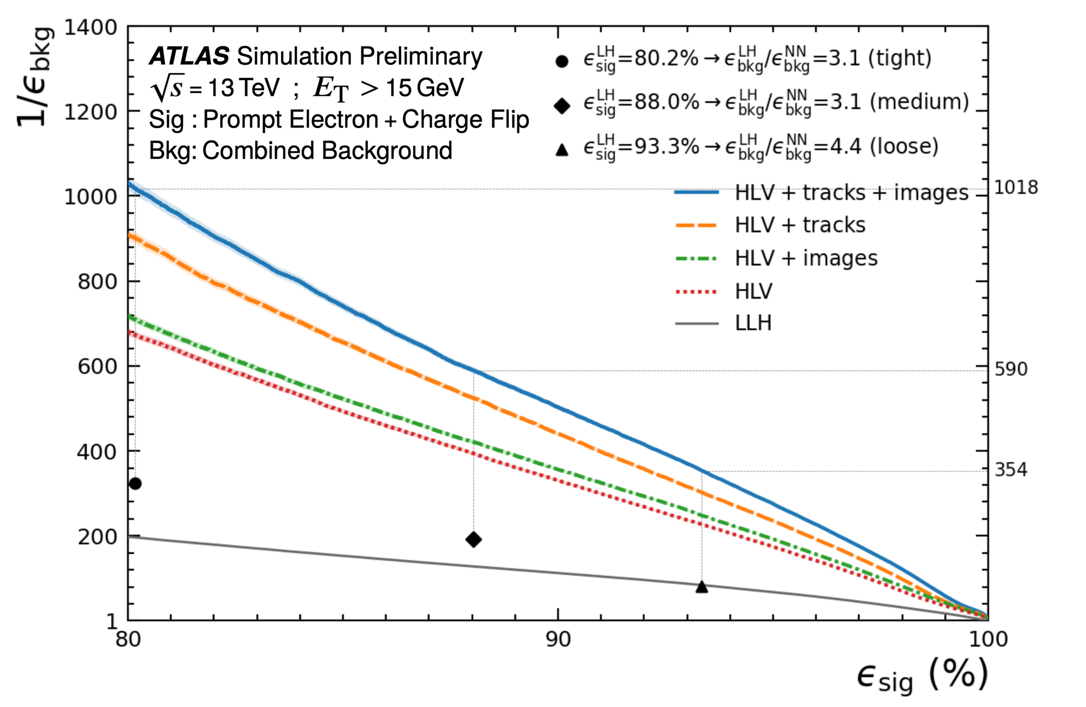


Neural Networks Performance

- Excellent separation observed between signal and all background classes, especially in the case of light-flavour hadron
- Other classes like heavy flavor or photon conversion are more challenging



- Signal = prompt + charge-flip
- Four CNN models trained to compare different components of the global architecture
- When all components are used together, best performance achieved



Towards training CNN in data

- CNN has been trained only on electron candidates generated by Monte Carlo simulation
- Simulation of fake electron object is imperfect, particularly the charged hadrons faking electrons since they involve hadronic showers
- Additional information (tracks, calorimeter images) used by the CNN might add sensitivity to data/MC differences
- A **sample enriched in light-flavour background** was designed and driven from data to provide training example for the CNN
- Multiple selections are applied on the sample to ensure purity in light flavor background
- Light flavor is the most common background in electron identification
- Hence, this class is where we expect the most significant improvement.

Acknowledgments

- *This work is supported in part by IVADO (Institut de valorisation des données)*
- *Special thanks to the ATLAS group from Université de Genève for providing Monte Carlo training samples*

IVADO

UNIVERSITÉ DE GENÈVE