INFIERI 2021 @ UAM

A machine learning algorithm for tau leptons identification at L2 Trigger in the CMS experiment

Valeria D'Amante[1,2,3]

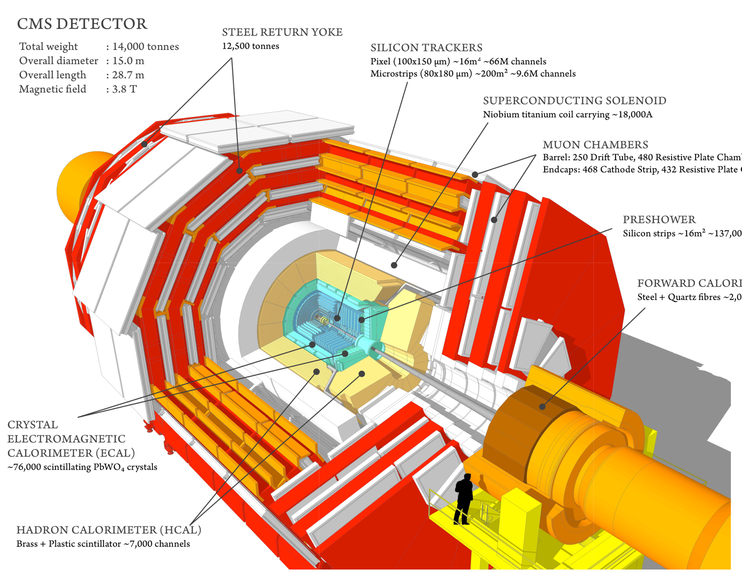
[1]Università di Siena, [2]INFN Sezione di Pisa, [3]CMS Collaboration

Introduction

The **τ lepton** is the only one kinematically allowed to decay into hadrons. It decays into its associated neutrino and into hadrons (τ_h) almost the 65% of times or, in the remaining cases, into μ/e and their associated neutrinos. Belonging to the **third generation**, it plays a key role in a wide variety of SM measurements and searches for new physics at the LHC. Improvements in τ reconstruction and identification for the future LHC data taking are crucial to increase the sensitivity of analyses involving tau leptons.

The CMS Experiment @ LHC

The **Compact Muon Solenoid** (CMS) experiment is a general purpose detector hosted at the Large Hadron Collider (LHC). It is a 21.6 m long cylindrical detector, with a total diameter of 14.6 m, centered to the LHC beam line.
At the heart of CMS sits a 13 m long, 5.9 m inner diameter, 4 T superconducting solenoid.



CMS Trigger

LHC bunch crossing every 25ns, corresponding to 40MHz event rate: too much data for us to record (~40TB/s), and most of events are not interesting.
Two-level triggering system:
- L1: hardware based trigger with a decisional time of 4μs, reduces the rate to 100kHz. Hardware trigger.
- HLT: software based trigger with full event information available, running on CPU + GPU based farm, reducing the rate to ~1kHz.

Tau Trigger @ CMS

Run 2 **Tau reconstruction at HLT** :
- **L2:** Calorimeter jets build around L1 seeds.
- **L2p5**: Pixel track based isolation around L2 hadronic tau leptons (only di-τ_h triggers)
- **PF event** reconstruction: regional PF reconstruction for $\tau_h\tau_h$ triggers
- **L3 tau** reconstruction:
  - From PF candidates HPS algo for ττ, μτ , eτ and cone based algo: single τ, τ+MET
  - Isolation criterion based on tracks from charged particles around τ lepton

There are ongoing developments for all parts of the reconstruction chain to update the tau reconstruction for Run 3. The goal of my project is to implement a **Machine Learning based L2 hadronic tau (τ_h) identification** exploiting a Convolutional Neural Network (CNN).

¿Why a ML Based approach for L2?

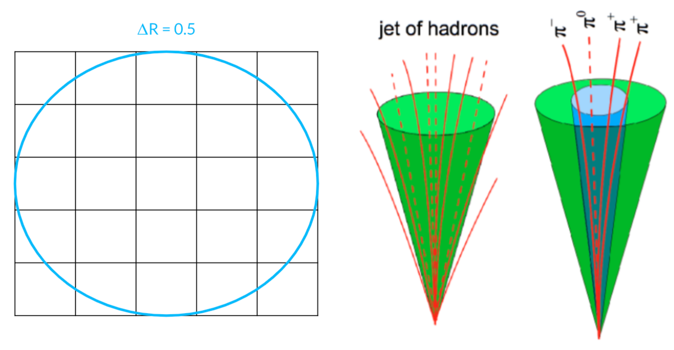
Current L2/2.5 tau selection is one of the most time-consuming steps and consists in a cut based algorithm. For future data-taking at LHC, we want to redesign L2/2.5 in order to maintain high efficiency and small timing with the increased instantaneous luminosity, exploiting availability of the GPU-based pixel tracks.
Machine learning (ML) algorithms make possible to perform a multi-variate analysis capable to capture all of the available classification power in order to discriminate between signal and background classes, tending to reach the optimal performances.
Hence, ML techniques are well suited for the L2 τ_h reconstruction step.

Data Samples

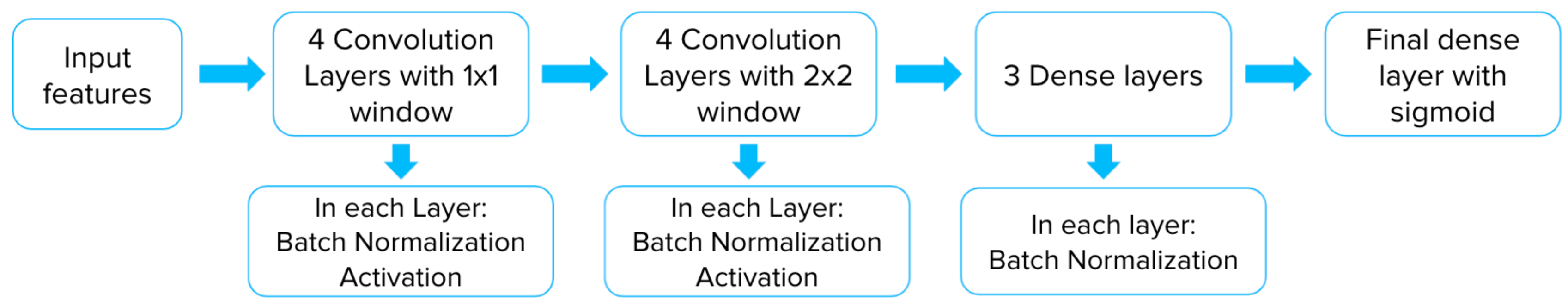
ML Training independent on τ lepton production mechanism:
Signals: Drell-Yan (DY), $t\bar{t}$ and W+jets MC samples with genTau p_T distribution weighted to obtain an uniform yield in each p_T bin. **Background:** QCD MC samples, with L1 Tau p_T distribution has been reweighted to L1 Tau p_T profile in data used for rate evaluation.
Signals and background have been divided in three subsamples for train, test and validation.
The performances of the ML algorithm after the training with respect to the cut-based procedure are evaluated in terms of **efficiency** on Vector Boson Fusion H→ττ and Z'→ ττ MC samples and **rate** on EphemeralHLTPhysicsX (X = 1,…,8) samples

Input features and CNN architecture

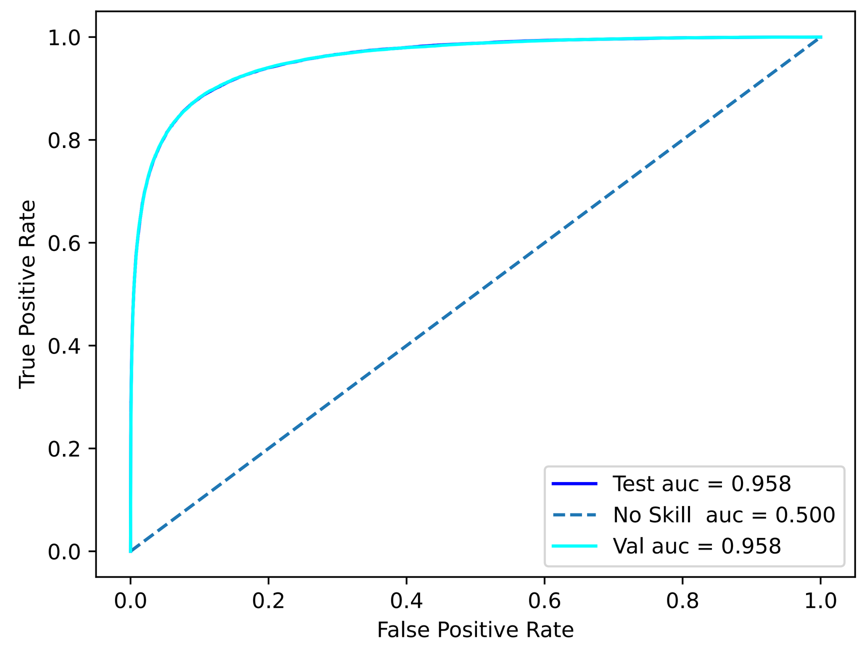
Objects have been associated with taus by requiring them to fall into a ΔR cone of 0.5 around the L1 τ candidate direction



Global observables: Number of patatrack vertices
L1 Tau observables: p_T, η, hwIso
CaloRecHit (electromagnetic and hadronic) and **Patatrack** observables related to each L1 τ.
To feed the CNN, the input features have been divided into **5x5 cells** in ηxϕ, resulting in 25 cells
Input features are **standardized** and truncated (when it's required)

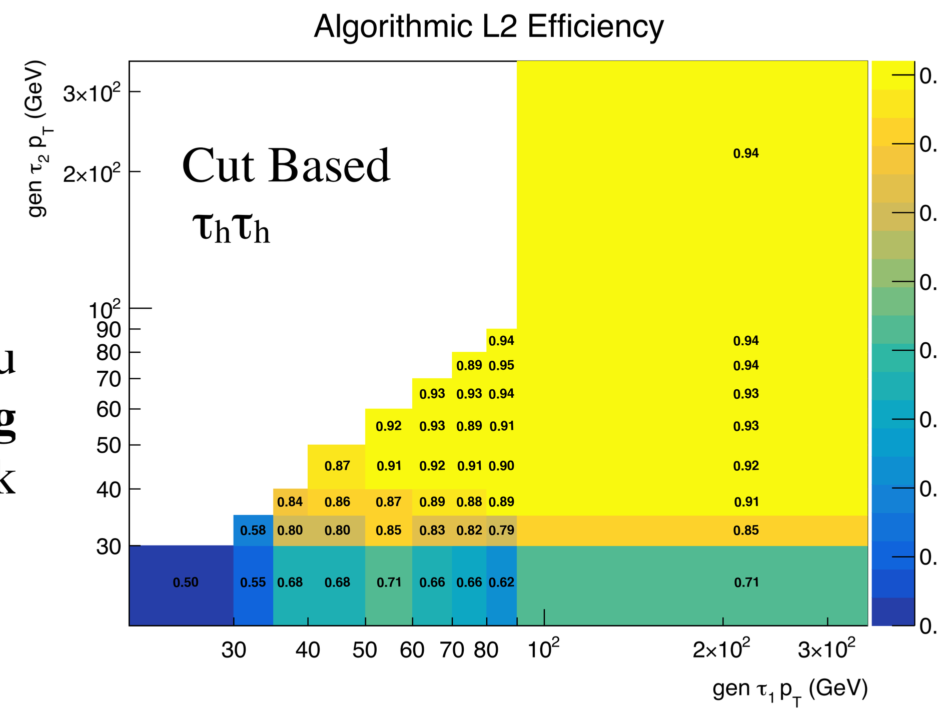


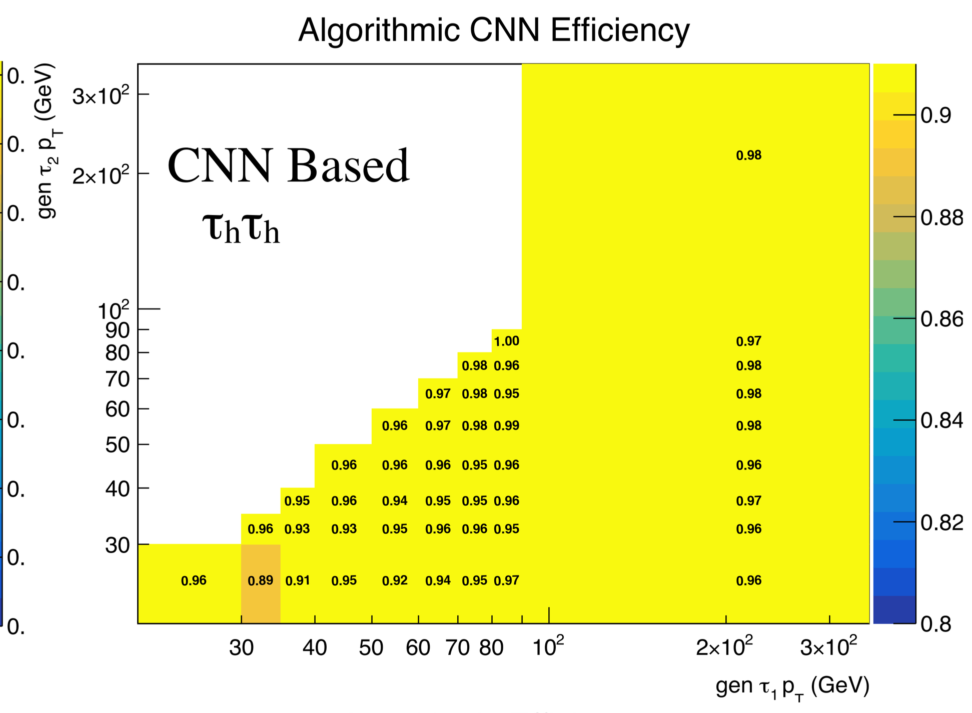
Number of trainable parameters = 23701
Activation function: **relu**
Optimizer: **Adam** with learning rate = 0.001
Loss function: **BinaryCrossEntropy**
Metrics: **Accuracy** (percentage of correctly identified events in all classes)
Saving condition: the model that reaches the **smallest loss for validation** sample is saved
Early stopping condition: the **training stops when the validation loss has no more improved in 10 epochs**



Results and Comparisons

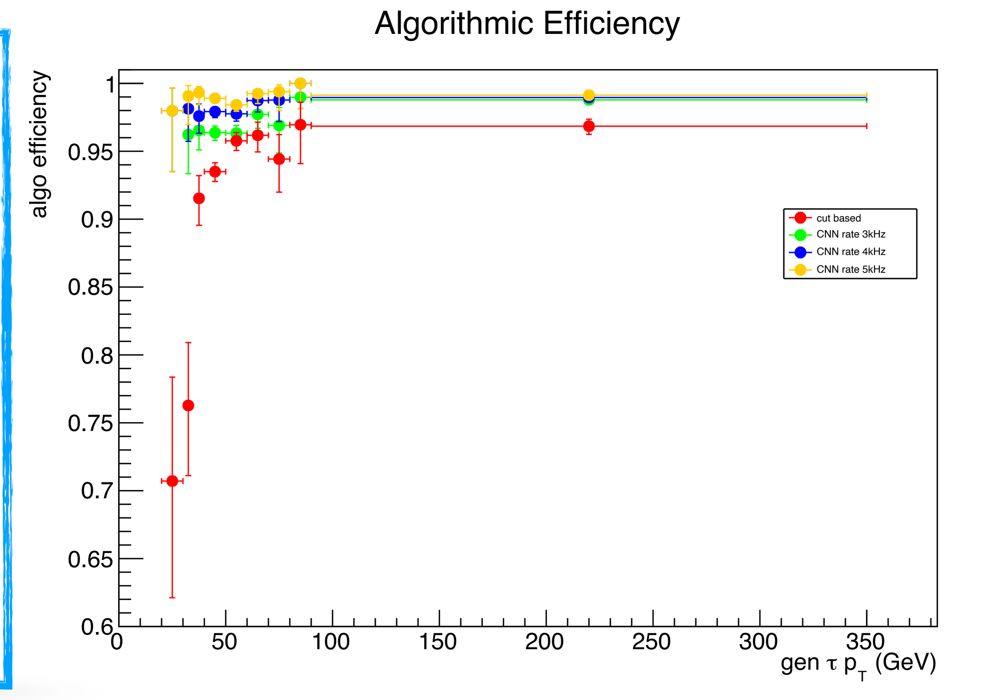
3 working point defined corresponding to different rate values to reach: 3, 4 and 5kHz. For $\tau_h\tau_h$ comparisons here it is reported the distribution for efficiency given the 4kHz working point .





To compare CNN with L2 cut-based performance in the single tau identification, the square root of 2D distribution has been considered.

The algorithmic efficiency for all 3 CNN selection scenarios is always better w.r.t. previous approach and above 90% in the whole gen τ p_T range.



Conclusion and Future plans

- **CNN-based approach for L2 Tau selection shows remarkable improvement** allowing considerable rate reduction, while keeping the algorithmic signal efficiency above 90% in the whole tau p_T range
- Inclusion in CMS Software (CMSSW) - ongoing
- Confirm physics performances and timing measurements - ongoing
- CNN hyperparameters optimisation