

Non-prompt background rejection with machine learning for soft lepton searches

CERN summer student project

Antti Pirttikoski,

in the supervision of Juliette Alimena and Marco Peruzzi

9.9.2021

Abstract

Many different physics theories beyond the Standard Model are studied at the LHC. One branch of these theories is concerning supersymmetry and especially models with compressed mass spectra, which could be detected via production of two or three soft leptons and missing transverse momentum. An important background for the soft leptons comes from $t\bar{t}$ events. Lepton candidates in such events can arise either from prompt production, or from non-prompt production and misidentification. The non-prompt background is also referred to as "fakes".

The main task of this project was to classify the prompt leptons from the fakes in the $t\bar{t}$ background in the context of new physics searches with the CMS experiment. This was done by training a neural network, which predicts the probability that the lepton is prompt. We found that the neural network succeeded very well in the task.

1 Introduction

The Higgs boson, which was discovered at the LHC in 2012, was the last missing building block of the Standard Model (SM) [1, 2] to be experimentally observed. After that, searches have focused more and more on physics beyond the Standard Model (BSM), to explain multiple phenomena not explainable by the SM, such as the existence of dark matter.

Supersymmetry (SUSY) is one branch of BSM theories and can be realized in many different forms. The minimal supersymmetric extension to the Standard Model (MSSM) predicts the existence of a supersymmetric partner for every fermion and boson in SM. These supersymmetric partners have otherwise the same properties as the SM counterparts but their spin differs by a half a unit [3].

However, a drawback with the MSSM is that it introduces over 100 new free parameters. One solution is to lower the number of parameters by making additional constraints to the model using experimental data. These constrained models are referred to as phenomenological-MSSM (pMSSM), and they typically contain 19 or less free parameters [4].

One of the potential candidates for dark matter is weakly interacting massive particles (WIMPs). In theories where the WIMP is the lightest among the new particles and the spectrum is nearly degenerate, meaning that another slightly more massive particle exists, the production of WIMPs at the LHC could be detected in final states with soft particles and low missing transverse momentum (MET). Such a particle spectrum is typically called compressed.

In such models, most of the energy of the higher-mass states is carried away by the lightest supersymmetric particle (LSP) [5]. Consequently, the detectable SM particles from the decay have low energy and the SM backgrounds are high. For this reason, hard initial state radiation (ISR) is required which will boost the SUSY particles and also produce large MET. However, background contributions from QCD multijet production and Z+jets production would still be dominant. In order to decrease these backgrounds, a requirement of two or three soft light leptons is also used.

The search focuses on leptonic decays of neutralinos ($\tilde{\chi}_2^0$) and charginos ($\tilde{\chi}_1^\pm$) into LSP neutralinos ($\tilde{\chi}_1^0$). The neutralinos and charginos are collectively referred to as electroweakinos. In the TCHIWZ model, $\tilde{\chi}_2^0 - \tilde{\chi}_1^\pm$ pair production is studied, where $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ are taken to be mass degenerate and decay into LSP via virtual W and Z bosons. Figure 1 shows the Feynman diagram of the studied TCHIWZ model. Another considered model is a simplified Higgsino model, where the production happens via the $\tilde{\chi}_2^0\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^\pm\tilde{\chi}_1^0$ channels. The search also considers top squark pair production.

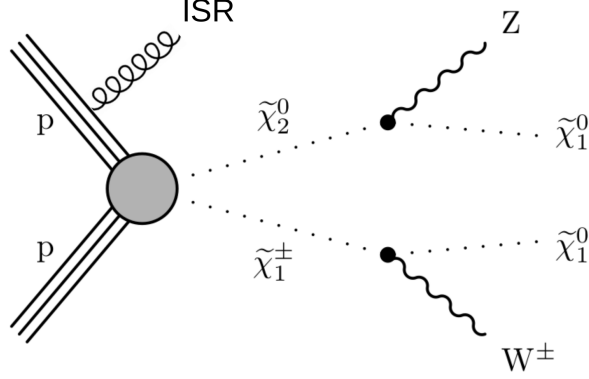


Figure 1: Studied TCHIWZ event, where neutralino and chargino decay into LSP neutralino while emitting Z and W bosons. Also hard ISR is emitted from one of the partons from the hard process. Figure modified from Ref. [5].

2 Methods

2.1 Event yields

Different backgrounds were earlier studied in more detail in the context of the search done in Ref. [5]. The event yields for the MC background, experimental data and predicted signal from TCHIWZ model were plotted with respect to multiple different variables, *e.g.* lepton p_T , η and impact parameter (IP_{3D}). These event yields were studied for combined data from 2016 to 2018 for different MET cuts.

Figure 2 shows the event yields as a function of subleading lepton p_T in the signal region at the high MET-bin, which is defined to be in range $240 < p_T^{\text{miss,corr}} < 290$ GeV, where $p_T^{\text{miss,corr}}$ is the corrected missing transverse momentum with the contribution of muons removed. As can be seen from Fig. 2, the dominant backgrounds are coming from Drell-Yan (DY), $t\bar{t}$, diboson and WZ events. Figure 2 also shows that the data agrees with MC background well and the largest predicted signal is coming from the TChi150/20 mass point, which correspond to a next-to-lightest SUSY particle (NLSP) with a mass of 150 GeV and a mass difference of 20 GeV to the LSP.

2.2 $t\bar{t}$ background

The main interest of this work was the $t\bar{t}$ background, the Feynman diagram for which is shown in Figure 3. As the t quark decays almost exclusively into a b quark and a W boson, and the W boson can decay into lepton and neutrino, $t\bar{t}$ events can produce also two leptons and MET.

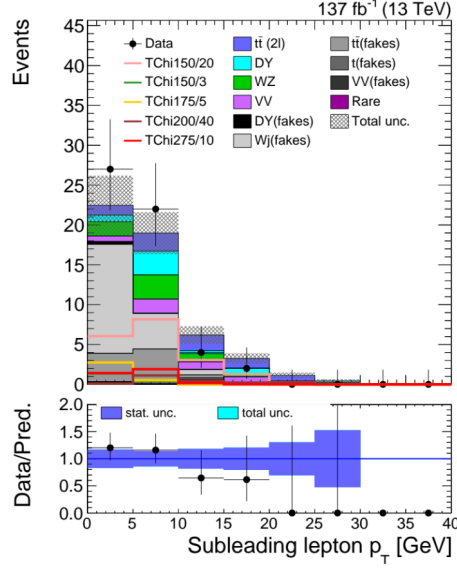


Figure 2: Event yields as function of subleading lepton p_T in the signal region at the high MET-bin with TCHIWZ model signal predictions.

This background can be reduced by requiring the absence of reconstructed b jets, although it still remains one of the major prompt backgrounds in the signal region.

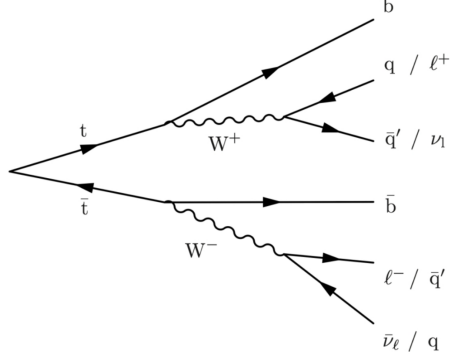


Figure 3: $t\bar{t}$ decay into b quarks and W bosons. Each W boson then decays to either a lepton and a neutrino or to a quark-antiquark pair.

The event yields were also plotted in the $t\bar{t}$ control region using 2016-2018 data, different MET bins and as a function of multiple variables. Figure 4 shows event yields as a function of subleading lepton p_T , $|\eta|$ and IP_{3D} . A considerable amount of fakes are also produced along with prompt $t\bar{t}$ events, as can be seen from Fig. 4 (also from Fig. 2). These fakes are leptons that are either produced away from primary vertex (PV) or fail the isolation requirement.

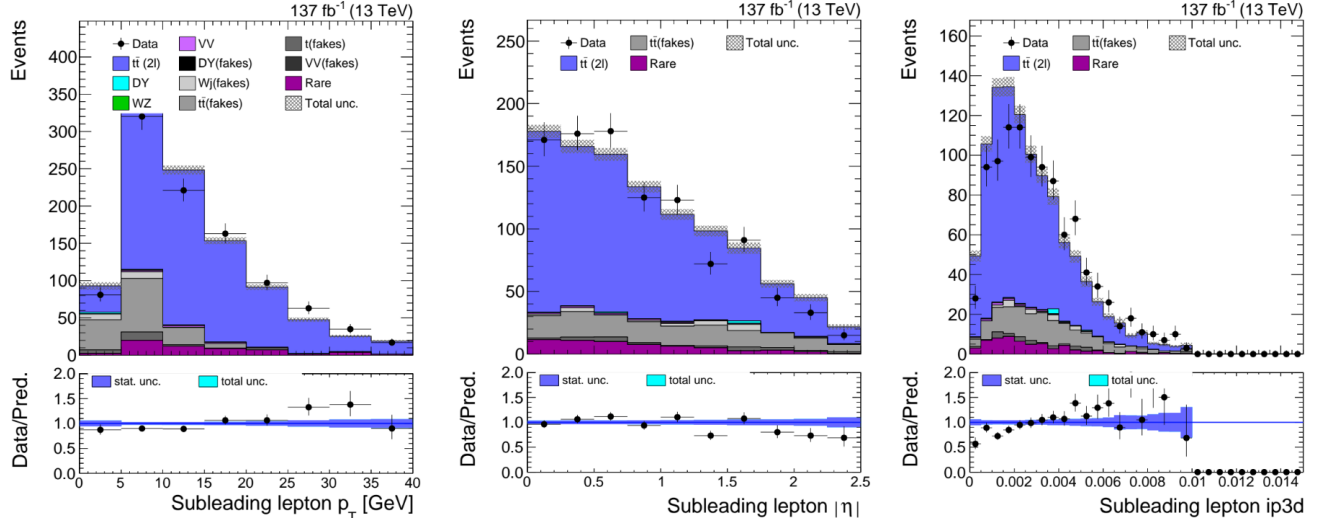


Figure 4: Event yields as a function of subleading lepton p_T (left), $|\eta|$ (middle) and IP_{3D} (right).

2.3 Classification with neural network

A neural network was trained to classify the prompt leptons and the fakes. A prompt lepton was defined to be a lepton which was matched with a generated lepton, while in the case of fake lepton we cannot find a generator lepton to match to. Five features related to the lepton were given as inputs to the network: p_T , $|\eta|$, IP_{3D} , significance of IP_{3D} (SIP_{3D}) and relative isolation (RelIso). The SIP_{3D} is defined as IP_{3D} divided by its uncertainty and RelIso as the sum of the energy deposits of other particles within $\Delta R < 0.3$ from the lepton divided by the lepton p_T .

The network consists of 5 dense layers with 128 neurons each. A rectified linear unit activation function is used in the dense layers and a sigmoid is used in the output layer. Batch normalization with momentum of 0.9 is applied before the first dense layer and after the third one. The network outputs the probability that the lepton is prompt.

Separate networks are trained for electrons and muons. For the training and testing, $t\bar{t}$ simulation from 2016 to 2018 was used, from which 80% was used for the training and 20% for testing. The networks are trained for 30 epochs. The optimal amount of epochs were studied by training the network for 200 epochs and then inspecting how the training and validation losses behave. Figure 5 shows the validation and training losses for both electron and muon networks. As can be seen from Fig. 5, the validation loss have reached a plateau after 30 epochs in both networks. The networks start to overtrain after about 75 and 125 epochs in the case of electrons and muons, respectively. This happens when the validation loss starts to shift upwards after the plateau [6].

The network was implemented using the DeepJetCore package [7], which is based on Keras [8] and TensorFlow [9] programs. The program was run on the lxplus computer cluster and the implementation and plotting scripts can be found at [10].

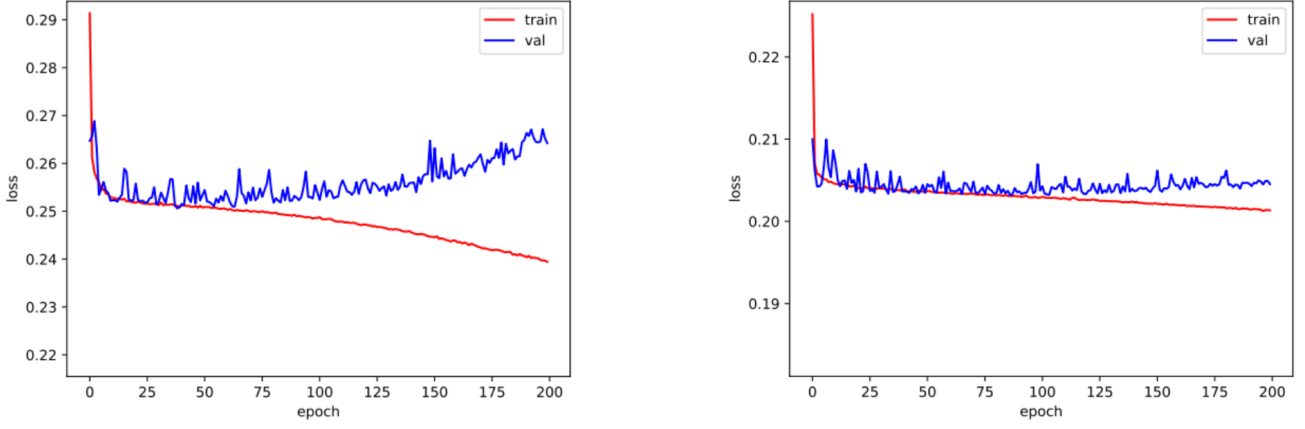


Figure 5: Training and validation losses for electron (left) and muon (right) neural networks.

3 Results

3.1 ROC curves

To visualize how well the network succeeded in the classification, ROC curves were created for both of the networks, which are shown in Fig. 6. As can be seen from Fig. 6, both networks performed very well in the classification task, as the fake efficiency starts to increase only at rather high prompt efficiency values. It can be also seen by comparing the ROC curves that the muon network performs a bit better, as its ROC curve starts increasing at higher prompt efficiency values.

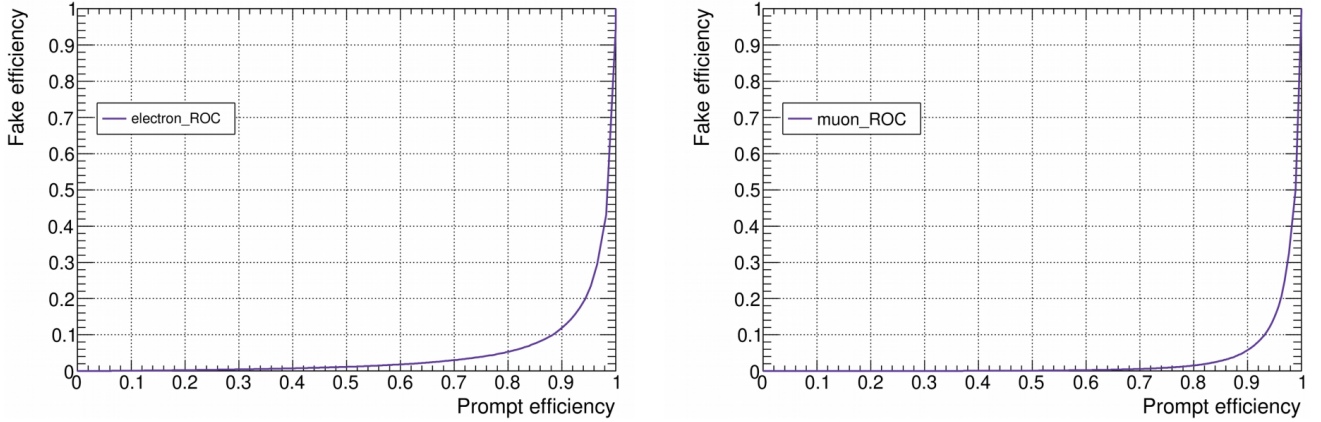


Figure 6: ROC curves for electron (left) and muon (right) neural networks.

3.2 Efficiency plots

In order to visualize the performance of networks as a function of a variable of interest for the analysis, efficiency plots were created as function of all the input variables. The efficiencies were

plotted for five different probability threshold values, after which the leptons were classified as prompt. These threshold were 0.7, 0.8, 0.9, 0.95 and 0.98. Figures 7 and 8 show the efficiencies for p_T and η , respectively. The efficiencies with respect to IP_{3D}, SIP_{3D} and RelIso are shown in Appendix A.

The efficiency as function of lepton p_T is shown in Fig. 7. As can be seen from Figure 7 the efficiency increases with the p_T . This was the expected behaviour as lower energy particles are harder to identify.

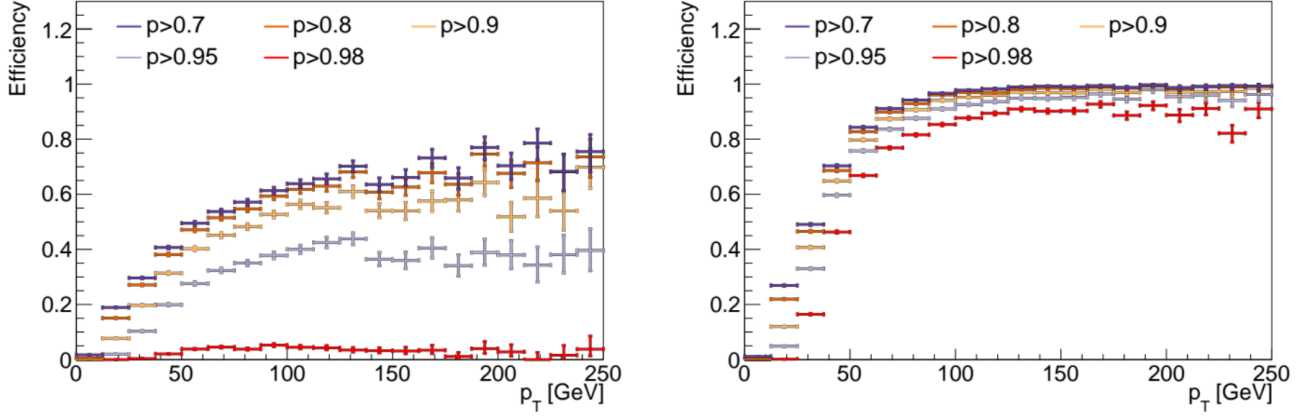


Figure 7: Efficiency as function of lepton p_T with multiple classification probability thresholds for electrons (left) and muons (right).

Figure 8 shows the efficiency as function of lepton η . It clear that best efficiencies are obtained at central η values. The faster decrease of the efficiency is also noticeable outside of the barrel region, *i.e.* $|\eta| > 1.3$, where the particle identification is less performant.

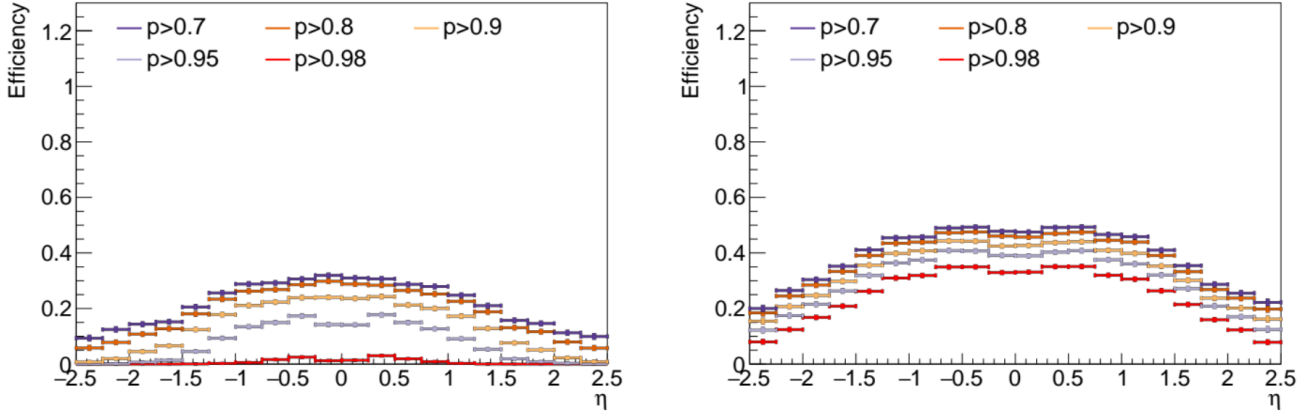


Figure 8: Efficiency as function of lepton η with multiple classification probability thresholds for electrons (left) and muons (right).

4 Conclusions

The neural network successfully classified prompt and fake leptons in a sample of $t\bar{t}$ background simulation. In the future this could be used as alternative method in studies like Ref. [5]. Since fakes are an important background in this search, regardless of the SUSY signal model being considered, the neural network classification is relevant for all of them.

In the future the neural network classification presented in this work could be improved in several different ways. For example, more variables could be used as input for the network. Also the optimal split between training and test data could be studied in more detail. It would also be important to compare the performance of the neural network to the classification methods which are currently used in the analyses.

Overall the fakes are an important background that needs to be quantified. It is advantageous to have multiple different methods to evaluate the amount of fakes in the background. The best available methods should be used in the fake estimation, which will lead to a more performant identification of soft leptons, and ultimately to more sensitive searches for compressed SUSY mass spectra.

References

- [1] CMS collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.
- [2] ATLAS collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sep 2012.
- [3] P.A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01.
- [4] I. Melzer-Pellmann and P. Pralavorio. Lessons for SUSY from the LHC after the first run. *The European Physical Journal C*, 74(5), May 2014.
- [5] CMS collaboration. Search for physics beyond the standard model in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at 13 TeV, CMS Physics Analysis Summary CMS-PAS-SUS-18-004, CERN, Geneva, 2021. <https://cds.cern.ch/record/2758359>.
- [6] Y. Bengio I. Goodfellow and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [7] Package for training and evaluation of deep neural networks for HEP. <https://github.com/DL4Jets/DeepJetCore>, Last accessed 26 September 2021.
- [8] F. Chollet et al. Keras, 2015. <https://keras.io>.
- [9] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. <https://www.tensorflow.org/>.
- [10] The repository of the program which was used for neural network studies in this project. <https://gitlab.cern.ch/jalimena/soslepid>, Last accessed 26 September 2021.

A Efficiency plots for IP_{3D} , SIP_{3D} and RelIso

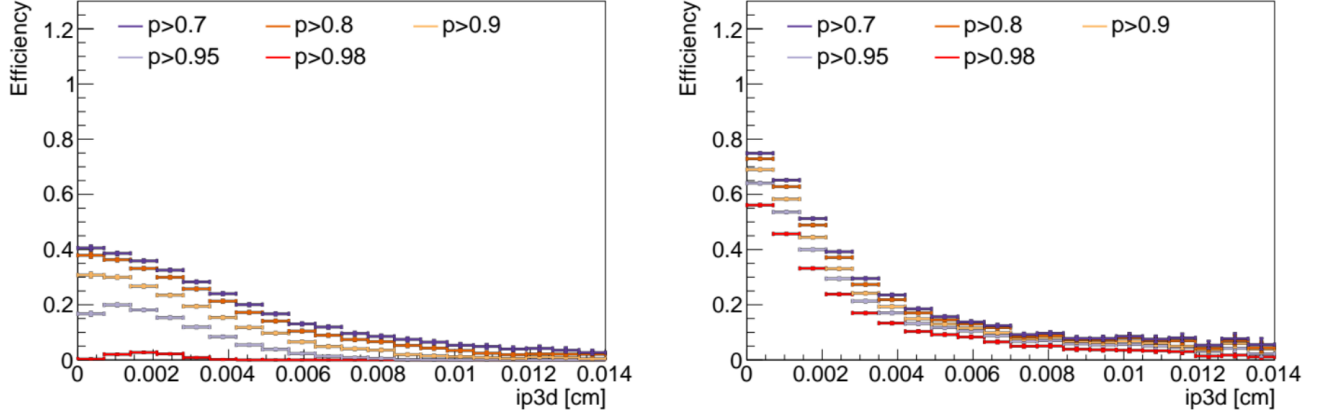


Figure 9: Efficiency as function of lepton IP_{3D} with multiple classification probability thresholds for electrons (left) and muons (right).

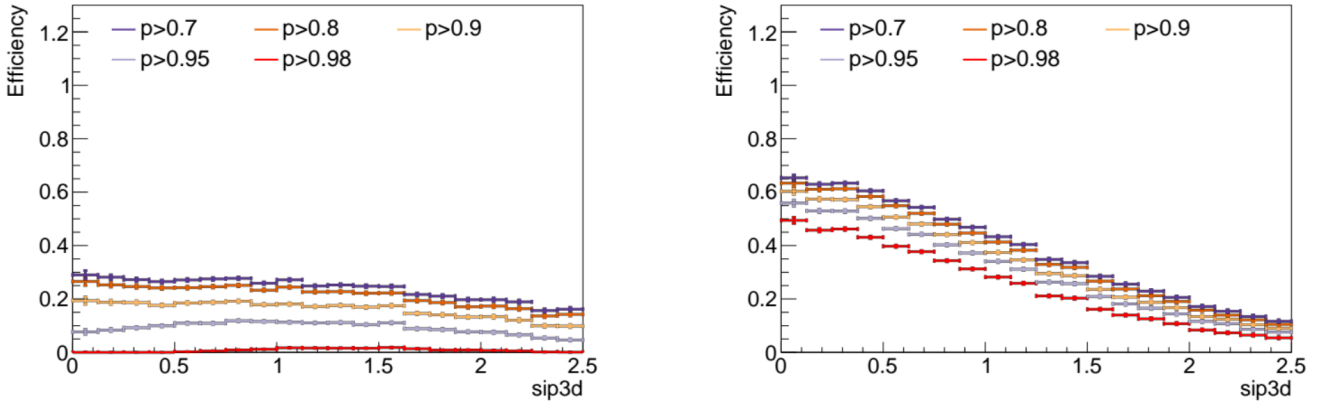


Figure 10: Efficiency as function of lepton SIP_{3D} with multiple classification probability thresholds for electrons (left) and muons (right).

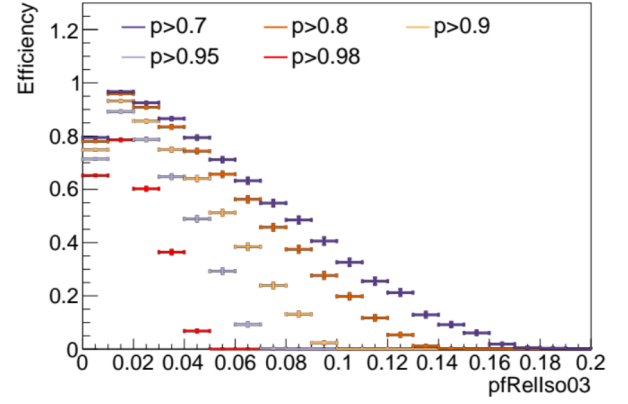
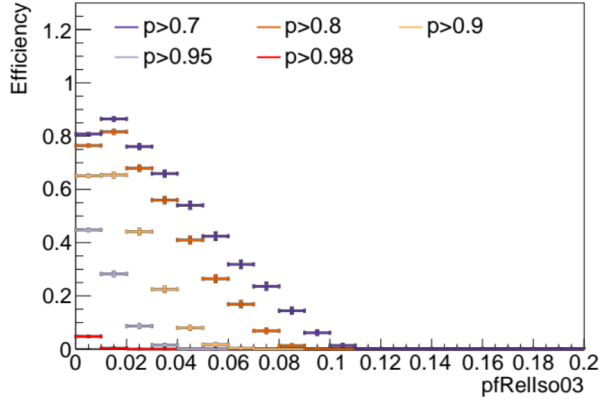


Figure 11: Efficiency as function of lepton relative isolation with multiple classification probability thresholds for electrons (left) and muons (right).