Diplomarbeit

Lepton identification for the CMS experiment with recurrent neural networks

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Technische Physik

eingereicht von

Maximilian Alexander Moser, BSc

Matrikelnummer 01429387

durchgeführt am
Institut für Hochenergiephysik
Österreichische Akademie der Wissenschaften
Apostelgasse 23
1030 Wien

betreut durch

DI Dr. Robert Schöfbeck, Priv.Doz.

Wien, 31.01.2022

Unterschrift Verfasser

Unterschrift Betreuer

Contents

Abstract

From the collision of two protons at the large hadron collider (LHC) massive vector bosons, W and Z bosons, that are sensitive probes of the electroweak interaction, can emerge and subsequently decay into lepton pairs. Among the leptons in such collisions, electrons and muons are efficiently detected by the CMS experiment at the LHC. The determination of the origin of these leptons is an important target of the event reconstruction.

This thesis describes an approach to lepton identification using Deep Learning techniques that incorporate the information from all reconstructed particles in the vicinity of the lepton candidate. This leads to a better identification efficiency, as the surrounding particles contain information about the leptons origin. Deep neural networks were trained for electron and muon identification and their performance was evaluated. The training and performance evaluation were done separately for the LHC data taking periods of the years 2016, 2017, and 2018.

For optimal performance, the fully connected deep neural networks were supplemented by long-short term memory units. The results show that this approach has a significantly better classification efficiency than traditional approaches, especially for low p_T lepton candidates. The new classifier is subsequently used to improve the background suppression in a search for supersymmetry with low-energetic leptons, improving the limit on the masses of superpartners, on average by about 5 GeV.

Finally, the relative importance of the input variables was determined and the timing and memory usage of different network configurations were compared.

Kurzfassung

Bei der Kollision von zwei Protonen am Large Hadron Collider (LHC) können massive Vektorbosonen, W- und Z-Bosonen, entstehen, die empfindliche Sonden für die elektroschwache Wechselwirkung sind und anschließend in Leptonenpaare zerfallen. Unter den Leptonen in solchen Kollisionen werden Elektronen und Myonen durch das CMS-Experiment am LHC effizient nachgewiesen. Die Bestimmung des Ursprungs dieser Leptonen ist ein wichtiges Ziel der Ereignisrekonstruktion.

In dieser Arbeit wird ein Ansatz zur Leptonenidentifikation mit Hilfe von Deep-Learning-Techniken beschrieben, die die Informationen aller rekonstruierten Teilchen in der Umgebung des Leptonenkandidaten einbeziehen. Dies führt zu einer besseren Identifikationseffizienz, da die umgebenden Teilchen Informationen über den Ursprung des Leptons enthalten. Tiefe neuronale Netze wurden für die Identifizierung von Elektronen und Myonen trainiert und ihre Leistung wurde bewertet. Das Training und die Leistungsbewertung wurden getrennt für die LHC-Datenaufnahmezeiträume der Jahre 2016, 2017 und 2018 durchgeführt.

Für eine optimale Leistung wurden die vollständig verbundenen tiefen neuronalen Netze durch Einheiten mit Langzeit-Kurzzeitgedächtnis ergänzt. Die Ergebnisse zeigen, dass dieser Ansatz eine deutlich bessere Klassifizierungseffizienz aufweist als herkömmliche Ansätze, insbesondere für Leptonenkandidaten mit niedrigem p_T. Der neue Klassifikator wird anschließend verwendet, um die Hintergrundunterdrückung bei der Suche nach Supersymmetrie mit niederenergetischen Leptonen zu verbessern, wobei die Grenze für die Massen der Superpartner im Durchschnitt um ca. 5 GeV verbessert wird.

Schließlich wurde die relative Bedeutung der Eingabevariablen bestimmt und die Zeit- und Speichernutzung verschiedener Netzwerkkonfigurationen verglichen.

Chapter 1

Introduction

Event analyses at the CMS (Compact Muon Solenoid) [1] experiment at the LHC (Large Hadron Collider) [2] rely heavily on accurate event selection, as minimizing statistical and systematic errors by reducing background events and increasing signal events improves, in general, the sensitivity. As the event selection is partly dependent on particle reconstruction and identification, the various identification and classification methods become very important.

For the purpose of this thesis, I group the reconstructed leptons' origin into three categories: Prompt, from the decay of a vector boson that was created at the primary proton-proton interaction point, non-prompt, from decays of short lived particles that emerged in the hadronisation process, and fake, particles falsely reconstructed as leptons. In most analyses, the prompt leptons are considered as signal, whereas non-prompt and fake leptons are considered background. Traditional classification methods range from simple cut-based approaches to machine and DL techniques. This thesis will cover an improved lepton classifier, leveraging the flexibility of neural networks with long-short term memory units (LSTM) [3] to use not only the lepton candidate's properties, but also the particles in close vicinity of the lepton.

Many traditional approaches in lepton identification show good performance for high momentum leptons, but recent analyses focus on searches for signals that predict leptons with low momenta. As these searches often suffer from non-prompt backgrounds with poorly isolated leptons, large improvements to the background suppression of these analyses can be made. The classifier developed in this thesis shows big improvements over traditional techniques.

On the technical side, we use the DeepJet framework [4] that was developed for classifying particle jets instead of leptons.Furthermore, we build on preliminary work on partial datasets [5].

This thesis organises as follows. In Sec. 2 I describe the LHC and the CMS experiment. In Sec. 3 I discuss DL. In Sec. 4 I explain lepton identification and in Sec. 5 I introduce the DeepLepton algorithm. In Sec. 6 I discuss the results and in Sec. 7 I give a summary and the conclusion of the thesis.

Chapter 2

Collider and Experiment

2.1 The Large Hadron Collider

The LHC [2] is located at the European Organisation for Nuclear Research (CERN) near Geneva, Switzerland. It is currently the highest energy particle accelerator and has been in operation since September 2008. The LHC was built in the 27 km long tunnel that had been built for the Large Electron Positron Collider (LEP) at a depth of 45 to 170 m underground. It collides two proton beams, separated into 2800 bunches each, travelling close to the speed of light in opposite directions in two separate beam pipes which are brought to collision at its four main experiments. These are:

- **ATLAS (A Toroidal Lhc ApparatuS)** [6] is a general purpose detector with the aim of investigating a wide range of physics, ranging from precision measurements of the standard model (SM) [7] to searches beyond the SM.
- **CMS (Compact Muon Solenoid)** [1] is a general purpose detector with the additional design requirement to reconstruct muon tracks with high efficiency. The main differences between the ATLAS and the CMS detectors are the configurations of the magnet and muon systems.
- **ALICE (A Large Ion Collider Experiment)** [8] was built with the goal of studying the properties of the quark gluon plasma in the collision of heavy nuclei.
- **LHCb (Large Hadron Collider beauty)** [9] was designed to investigate the decays that contain b quarks in an attempt to understand the matter-antimatter symmetry and CP violation.

2.1.1 Accelerator Complex

Before the injection into the LHC, the proton beams are pre-accelerated by a chain of smaller accelerators. The initial protons from a hydrogen source are accelerated by the Linear accelerator 4 (Linac4) [10] to an energy of 160 MeV to prepare them for the Proton Synchrotron Booster [11] which accelerates the protons to 2 GeV. They are then further accelerated to 26 GeV by the Proton Synchrotron (PS) [12] and to 450 GeV by the Super Proton Synchrotron (SPS) [13]. A schematic illustration of the LHC with its pre-accelerators is given in Fig. 2.1. In order to enable the high beam brightness of the LHC beams, all links of the pre-accelerator chain had to be upgraded for LHC operation.

2.1.2 Controlling the Proton Beams

The LHC is not perfectly circular. The ring consists of eight straight and eight curved sections, with the four interaction points located at the straight sections and the two high-luminosity

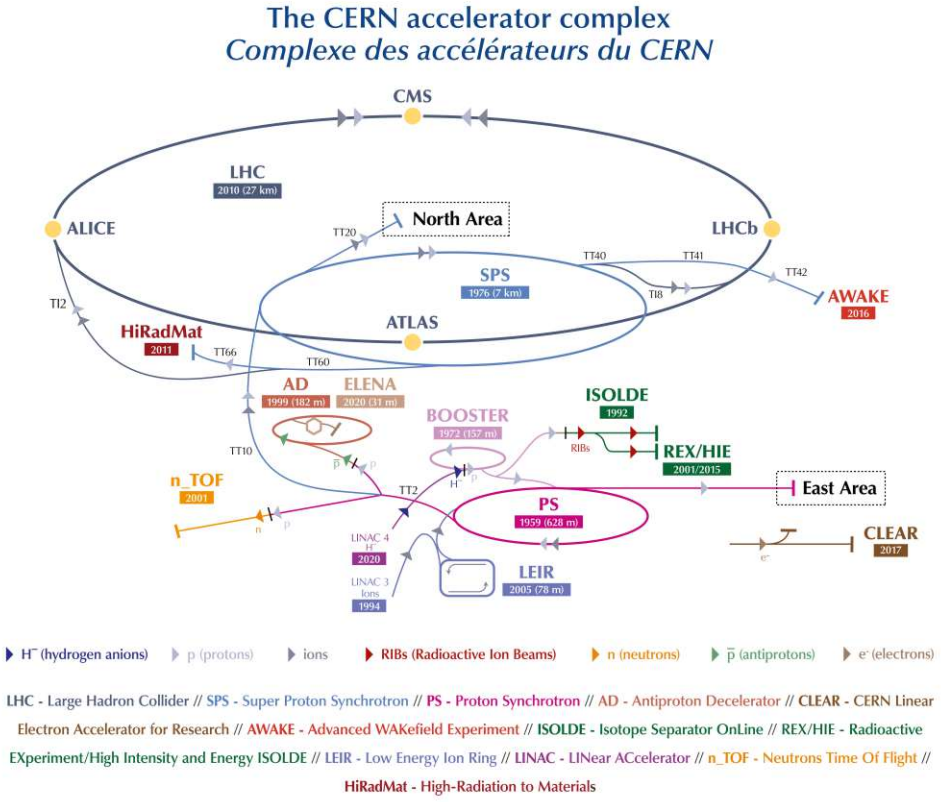


Figure 2.1: Schematic drawing of the CERN LHC accelerator complex with its main experiments and pre-accelerators. Image taken from [14].

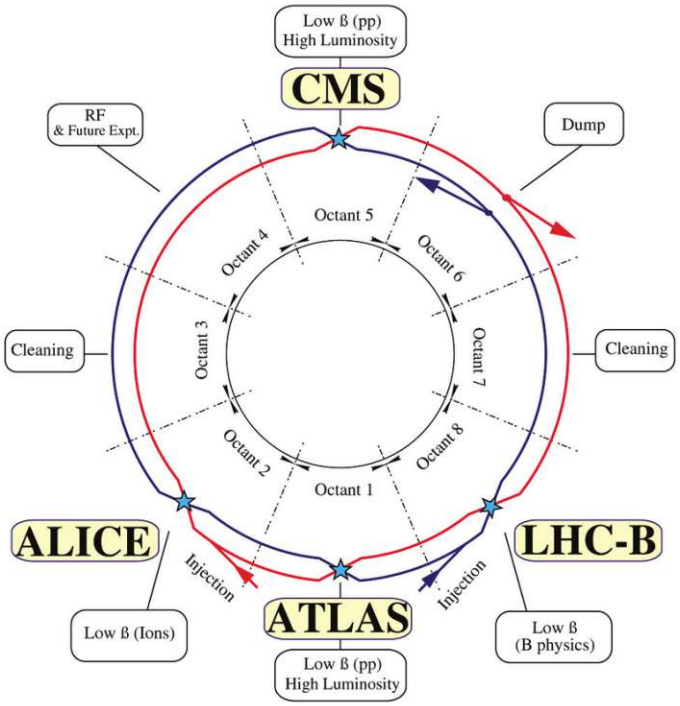


Figure 2.2: A Schematic drawing of the layout of the LHC illustrating the octants and the relative location of the interaction points and the main experiments. Image taken from [16].

experiments opposite of each other. This is shown in Fig. 2.2. At opposite points, between the ATLAS and CMS experiments, the proton beams are cleaned, meaning particles in the halo of the proton beams are removed to reduce the background of the experimental detectors and to minimize the beam loss in the cryogenic parts of the accelerator. In octant 6 in Fig. 2.2 the beam can be dumped immediately in case of an emergency such as a vacuum leak [15].
There are 1232 dipole magnets for bending the beam. They produce a magnetic field with a strength of 8 T at an operating temperature of below 2 K. The beams are focused by 392 quadrupole magnets. The acceleration of the beam to operating energy is handled by eight radio-frequency (RF) cavities [17] located in octant 4. In addition, the RF cavities also keep the proton bunches compact by accelerating and decelerating protons such that the difference to the nominal beam energy is reduced.

2.1.3 Energy and Luminosity

One very important metric of a particle accelerator is the center-of-mass energy $\sqrt{s}$ of the colliding particles. In natural units, with $c = 1$, this is given with [18]

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{2} E_i\right)^2 - \left(\sum_{i=1}^{2} \boldsymbol{p}_i\right)^2} \tag{2.1}$$

where $i = 1, 2$ labels the protons, E_i denotes the protons' energies and p_i denotes the protons' three-momenta. The center-of-mass energy of the LHC during Run II operations equals 13 TeV.
Another important metric of a particle accelerator is the instantaneous luminosity $\mathcal{L}$ [19], which defines the rate of proton-proton collisions at an interaction point.Integrated over the data-taking period, the cross-section σ of the process and the luminosity $\mathcal{L}$ are related to the total number of events by [18]

$$N = \sigma \int \mathcal{L}(t)\mathrm{d}t. \tag{2.2}$$

The instantaneous luminosity is given by

$$\mathcal{L} = \frac{f_{rev} n_b N_p^2}{4\pi\sigma_x\sigma_y} F(\theta_c, \sigma_x, \sigma_y) \tag{2.3}$$

where

$f_{rev} = 11\,245.5\,\mathrm{Hz}$	is the bunch revolution frequency,
n_b	is the number of proton bunches in the machine,
N_p	is the number of protons per bunch,
σ_i	is the transverse beam widt and
$F(\theta_c, \sigma_x, \sigma_y)$	is the luminosity reduction due to the beam angle θ_c.

At design, the LHC's instantaneous luminosity is in the range of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. An often used parameter is the integrated luminosity

$$L = \int \mathcal{L}dt. \tag{2.4}$$

It is a measure for the total number of interactions over a given time period and is usually given for one data taking run. The data recorded by the CMS experiments of the LHC run two, from 2016 to 2018, corresponds to an integrated luminosity of $150\,\mathrm{fb}^{-1}$.

2.1.4 Discoveries

The arguably most important discovery of the LHC in recent years was the Higgs boson in 2012 [20]. This discovery finally completed the particle content of the SM. This discovery also confirmed the Brout-Englert-Higgs mechanism, which explains how SM particles interact with the Higgs field and in this way gain their mass [21, 19].

As of March 2021, 59 new hadrons have been discovered at the LHC [22]. Hadrons are composite particles that are made out of quarks. There are hadrons made out of a quark-antiquark pair called mesons, and particles made out of three quarks called baryons, such as protons and neutrons. Among the 59 newly discovered particles are also tetraquarks and pentaquarks that consist of four and five quarks respectively.

Another discovery was made when quark-gluon plasma [23] was found by the ALICE experiment. This is the state of matter that is thought to have existed shortly after the Big Bang.

2.1.5 Ongoing Searches

Supersymmetry (SUSY) [24] is an extension to the SM. It offers potential solutions to some of the SM's problems, such as the unification of the three fundamental forces. Many analyses at the LHC focus on the search for SUSY particles and interactions, but until now there has not been experimental evidence for any of them [19].

Another area of interest is the search for dark matter [25]. As the galaxies seem to be too light to exist, there has to be something that gives them extra mass. This "matter" is not visible to us, as it does not interact with the electromagnetic force, therefore it is called dark matter. It is thought to make up 27% of the universe, whereas 5% are made of normal matter and the other 68% are made of dark energy. The dark matter particles are thought to interact with quarks and/or gluons and therefore also to be produced by the proton-proton collisions [19, 26, 27].

In the beginning of the universe, matter and antimatter should have been created equally, but today everything around us is comprised of normal matter. When matter and antimatter come

into contact, they annihilate to energy, therefore the universe should only contain leftover energy, but obviously this is not what is observed. This leads to the conclusion, that there must be a process that favours matter over antimatter. Therefore, searches are conducted to understand the difference in behaviour of matter and antimatter [28].

2.2 The CMS Experiment

2.2.1 Introduction

The CMS detector is one of the four detectors at the LHC at CERN [1] located at point 5. It is used to measure and identify particles from proton-proton collision and its subsystems facilitate this task. The innermost part, closest to the primary interaction point of the beam, is a silicon tracker that measures the path of charged particles bent by a superconducting solenoid magnet. This allows for the reconstruction of the momentum of charged particles. The energy of particles is independently measured by two calorimeter systems.

An important design requirement was to characterise events that involve the W and Z bosons, especially in leptonic decay channels, as these are very clean experimental signals. Therefore, emphasis has been on the efficient reconstruction of these leptons. Additionally, efficient tagging of hadron jets coming from b quarks and reconstruction of secondary vertices (SVs) were important design requirements [29]. Figure 2.3 shows a schematic image of the CMS detector with its sub-detectors.

2.2.2 Detector

Solenoid

For the momentum analysis of charged particles in the CMS detector a 3.8 T magnetic field is used. This is being provided by the superconducting solenoid magnet, around which the detector has been built [29]. This high magnetic field strength is achieved with a superconducting solenoid. The purpose for using such a strong magnetic field is to increase the muon trigger resolution and momentum tracking accuracy. The cables are made of a strand of the superconductor NbTi encased in aluminium for mechanical stability [31].

The solenoid has 2168 turns of wire with a length of 12.5 m and an inner diameter of 6.1 m. With a current of 20 000 A flowing through the cable, this results in

$$B = \mu_0 \mu_r \frac{NI}{L} \approx 4\mathrm{T}, \tag{2.5}$$

with $\mu_r = 1$ in a vacuum and $\mu_0 = 1.2566 \times 10^{-6}\,\mathrm{m\,kg\,s^{-2}\,A^{-2}}$. The cable is being kept at a temperature of 4.5 K [19, 31, 32].

Silicon Trackers

In order to calculate the momentum of a particle, its curvature in the magnetic field of the solenoid is measured. To do this, a silicon detector is being used. It operates similarly to a digital camera sensor that is stacked and arranged in a barrel with endcaps [29, 33]. The silicon sub-detector is made of silicon pixels nearest to the interaction point and silicon strips on the outside, as depicted in Fig. 2.3 [29, 34]. Once a charged particle passes through a pixel, it creates an electron-hole pair. This signal is then amplified. The silicon pixel detector is made of three layers of pixels to create a three dimensional track of the particle [35]. There are 65 million channels, each consuming about 50 µW. In order for the detector not to overheat, these pixels are mounted on cooling tubes [34]. Figure 2.4 shows a schematic image of a part of a pixel tile.

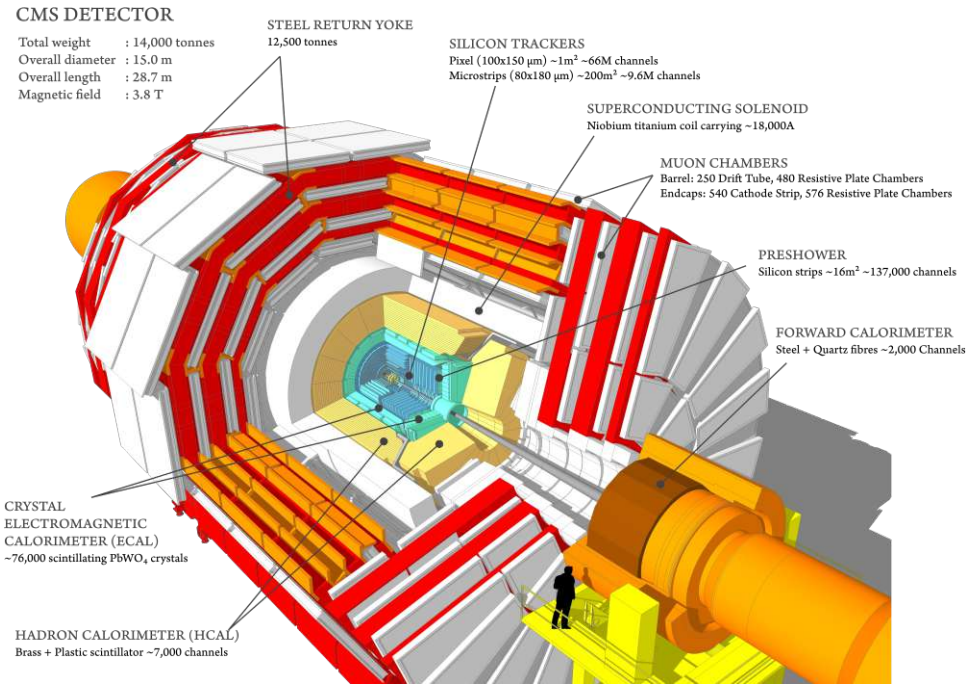


Figure 2.3: A schematic image of the CMS detector with its sub-detectors [30].

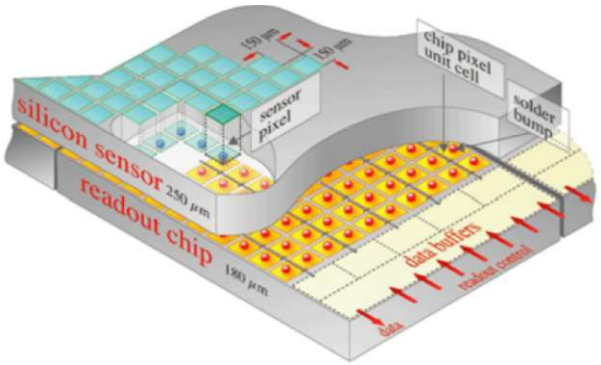


Figure 2.4: Schematic image of a silicon pixel sensor module [34].

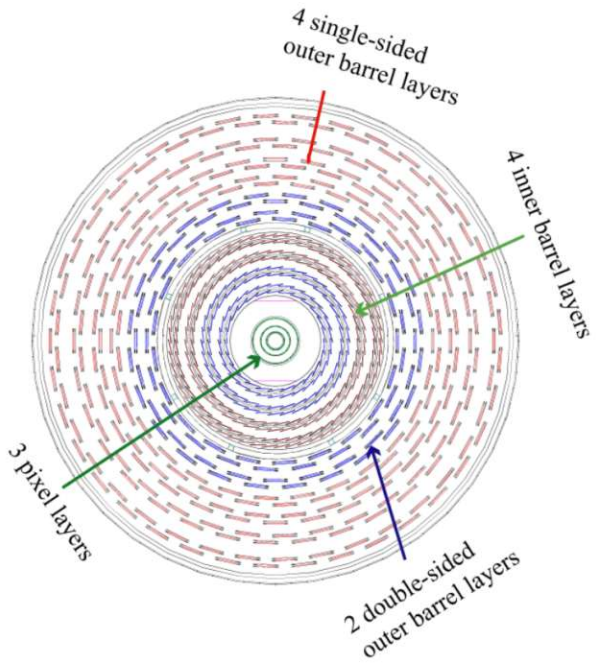


Figure 2.5: Schematic image of a cross-section cut of the arrangement of the silicon strips and silicon pixels[35].

The silicon strips work similarly to the silicon pixels. The barrel-part of the detector is layed out in ten layers as shown in Fig. 2.5, additionally there are nine layers of silicon strips on the endcaps [29, 35].
In order to reconstruct the tracks of the particles passing through the pixel detector, the measurements of the silicon trackers are accurate to 10 µm [33].

Calorimetry

The electromagnetic calorimeter (ECAL) [36] measures the energy of charged particles, such as electrons and photons. This is done with lead-tungstate crystals. They are highly transparent and scintillate when electrons and photons pass through them [37]. The light emitted by the crystals is proportional to the energy deposited. This scintillation light is then measured by silicon avalanche photodetectors.

The front face of the crystals in the barrel part is at a radius of 1.29 m. The crystals have a front face of 22 mm × 22 mm and a length of 230 mm, which corresponds to 25.8 times the radiation lengths (X_0), the distance at which a high-energy particle's energy is reduced to $\frac{1}{e}$ due to electromagnetic interactions.

The endcaps are made of crystals with a length of 220 mm, corresponding to 24.7 times X_0, with 3 X_0 from a preshower. In total, there are 61200 crystals in the barrel part, and 15000 crystals in the endcaps [37]. Figure 2.6 shows the relative energy resolution of the ECAL.
The hadron calorimeter (HCAL) [39, 40] measures the energy of charged and neutral hadrons and provides an indirect measurement of the presence of particles that interact at most weakly, because it is nearly hermetic. If an imbalance transverse to the direction of the beam line is measured, it can be deduced that invisible particles, such as neutrinos, have been produced.

The HCAL is built as a sampling calorimeter, which means that it is built of alternating layers of absorbing and scintillating materials. The absorber material is brass. The light of the scintillator plates is collected by optical fibres and measured by photodiodes. In the very forward part of the detector, where the particle flux is highest, steel absorber plates and quartz fibres as are used.

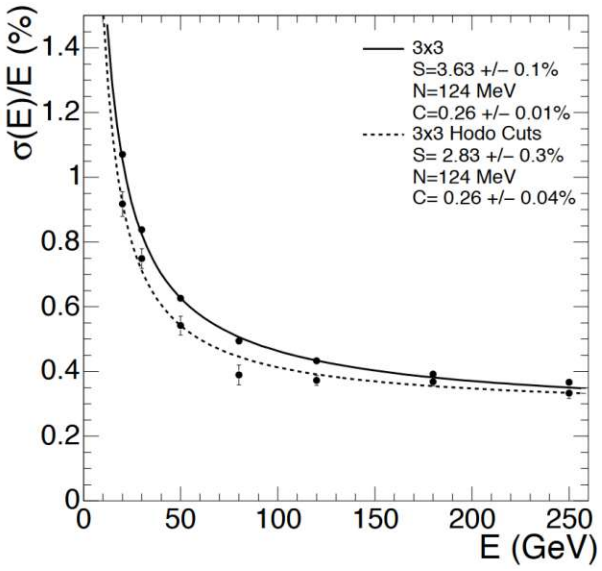


Figure 2.6: Relative ECAL energy resolution [38].

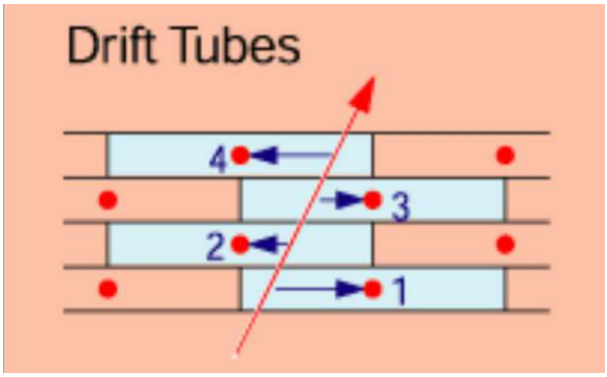


Figure 2.7: Illustration of muon drift tubes [42].

Muon System

As muons are able to penetrate several meters of iron without energy loss, the muon system is placed on the outside of the experiment, where they are likely the only particles left from the initial interaction [41]. The muon system is interleaved with the iron return yoke plates, which make up a large portion of the detector's mass and contain the solenoid's magnetic field.
The muon system consists of three different trackers. Drift tubes [42], illustrated in Fig. 2.7, consist of a chamber filled with gas and a stretched wire. When a muon passes through the chamber, electrons are knocked off the atoms in the gas, guided by the static EM field they end up at the positively charged wire, where they are detected. The drift tubes not only measure that a muon has passed through it's volume, but also the distance to the wire and position along the length of the wire. One drift tube chamber consists of twelve aluminium layers, these are arranged in three groups of four, containing up to 60 tubes each.
The next type of muon detector are the cathode strip chambers [43]. These consist of arrays of positively charged wires with negatively charged copper cathode strips. The wires and strips are arranged perpendicularly to each other, as illustrated in Fig. 2.8. When a muon passes through the the gas volume, it knocks off electrons of the gas atoms. The electrons move towards the anode wires, creating an avalanche of electrons. The positively charged ions move towards the copper cathode strips and away from the wires, therefore being detected on the strips. Because of the right angle between the strips and wires, two positional coordinates are obtained for each

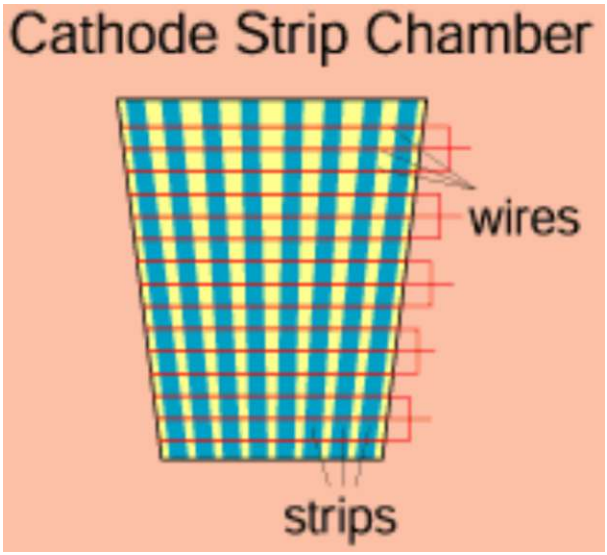


Figure 2.8: Illustration of a cathode strip chamber [43].

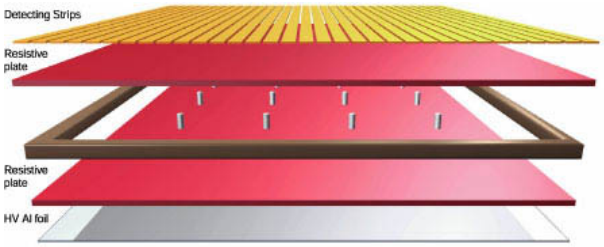


Figure 2.9: illustration of a resistive Plate Chamber [44].

passing particle. Additionally, due to the fast detection time, these chambers are suitable for triggering.
Another detection chamber is the resistive plate chamber [44]. They are fast detectors that are also used as a muon trigger system. They consist of two parallel plastic plates with a very high resistivity, these plates are separated by a gas volume. This is illustrated in Fig. 2.9. Once a muon passes through, the gas atoms in it's path inside the chamber are ionised. These free electrons are accelerated towards the anode, causing an avalanche of electrons. This signal is then picked up by the detection strips after a small, but precise, time delay. The pattern of hits on the strips gives a quick measure of the muon's momentum, which is then used by the trigger. Drift tubes and resistive plate chambers are used in the barrel region, in the endcaps cathode strip chambers and resistive plate chambers are used [41].

2.2.3 Trigger and Data Acquisition

When the LHC is running at peak instantaneous luminosity, about one billion collisions happen every second inside the CMS detector. There would not be enough capacity to store this large amount of data, considering that each event generates 1-2 MB of data. A trigger system reduces the rate to a few hundred Hz [45, 46].

This poses a challenge, as there are only 25 ns between proton bunch crossings, so new particles are generated even before the particles from the preceding bunch crossing have left the detector. The data collected by the subdetectors is, therefore, being stored on tape.

The first trigger stage (L1) is implemented in hardware and reduces the rate to approximately 100 kHz and uses the data from the calorimeters and the muon system. Events that pass through

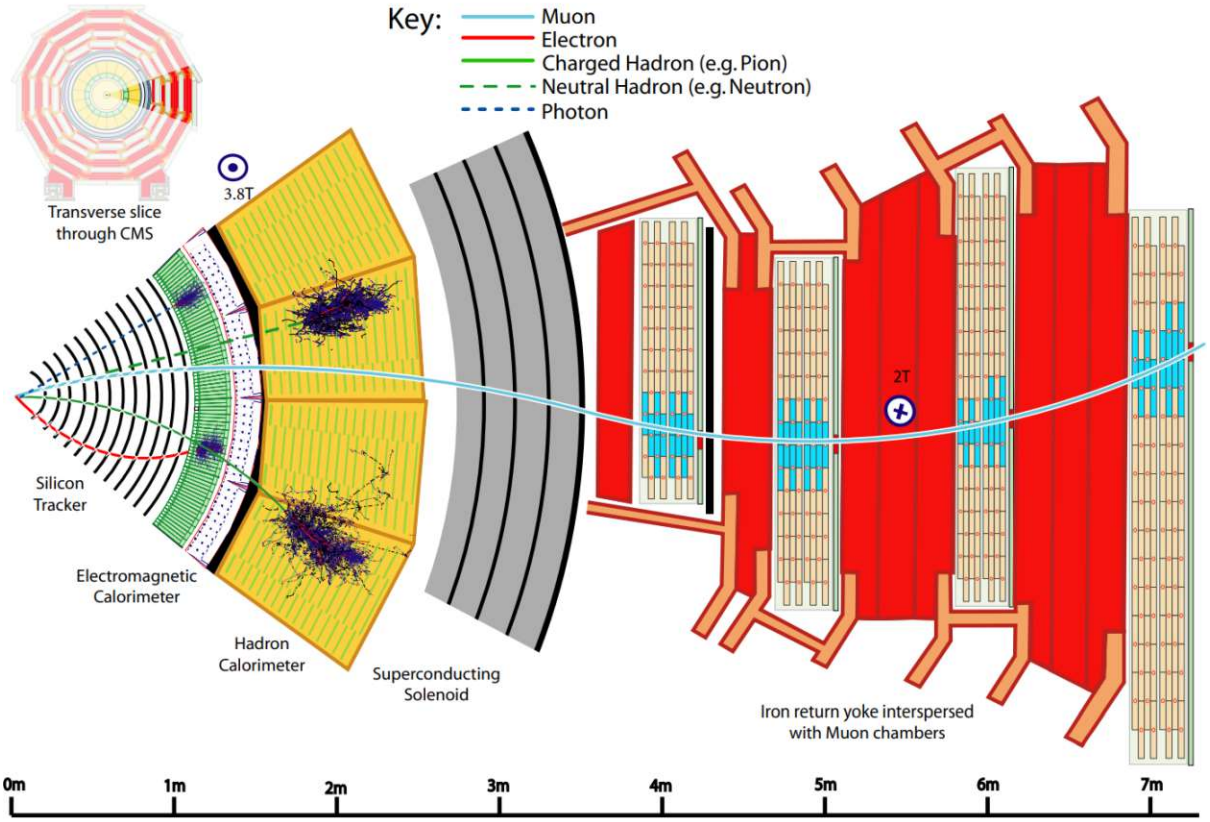


Figure 2.10: A sketch of a transverse slice through the CMS detector and how different particles pass through it [47].

this stage, are subsequently filtered by the high level trigger stage (HLT). The HLT output rate is, on average, 800 Hz at design, using a simplified version of the nominal event reconstruction.

2.2.4 Reconstruction

To reconstruct particles in the CMS detector, the particle flow (PF) [47] algorithm is used. With this algorithm, electrons, muons, neutral hadrons, charged hadrons and photons are reconstructed. It combines the information from all sub-detectors to identify the particles by using compatible detector signals to determine their type. This data is then used for further reconstructions and identifications, such as jet reconstruction and identification. Figure 2.10 shows a transverse slice through the detector and how different particles pass through it.

Electrons

Electron reconstruction is based on the combination of measurements from the ECAL with the inner tracking system [48]. Because of the relatively long distance between the collision point and the ECAL, some of the electron's energy is lost through bremsstrahlung. It was shown, that in a test beam about 97 % of it's energy were deposited in a 5×5 crystal array. Overall about one third to 86 % of the electron's energy is lost before it reaches the calorimeter, depending on the amount of material it passes through. To measure the initial energy accurately, it is therefore critical to collect all the radiated photons too.

In the barrel region, this is done via the hybrid algorithm. It uses a seed crystal, the crystal with the most energy deposited in the considered region that is larger than a predefined minimum $E_{\mathrm{T,seed}} > E_{\mathrm{T,seed}}^{\mathrm{min}}$. Along the transverse directions of the seed crystal arrays of 5×1 crystal are

added in a range of N_{steps}, as long as their energy exceeds a minimum energy of $E^{\text{min}}_{\text{array}}$. These clusters are then collected into a final global cluster called supercluster.

In the endcaps the multi-5×5 algorithm is used. The seed crystal is the crystal with the local maximal energy of it's four direct neighbours that again fulfils the requirement of $E_{T,seed} > E^{\text{min}}_{T,seed}$. Around these seed crystals, the energy is collected in 5×5 clusters, which may partly overlap. These clusters are then collected together to superclusters, if their energy exceeds a minimum [48]. Due to the large radiation losses of the electrons when they are curved in the magnetic field, the standard procedure of the Kalman filter [49] track reconstruction is not used, but a special algorithm.

This algorithm starts by "seeding". It selects the first two or three hits in the tracker, from where the track can be initiated. Because of the importance of seeding for the reconstruction efficiency, two complementary algorithms are used for this. One starts from the supercluster energy and position in the ECAL and estimates the electron track to select the electron seed from all reconstructed seeds. The other algorithm relies on tracks, that have been reconstructed using the general algorithm for charged tracks, extrapolates them towards the ECAL and matches them to a supercluster. Further steps are then taken to increase the efficiency, such as using a matching-momentum criterion. These selected electron seeds are then used to build the electron tracks. This is done iteratively - layer by layer - with the energy loss taken into account. In cases where several hits in a layer might be compatible with those predicted, several possible trajectory candidates are created. Over the whole track, only one missing hit is allowed. This procedure provides tracks up to the ECAL, where the fraction of momentum lost inside the tracker is calculated. Then, the tracks and the ECAL clusters are being associated with each other. There are criterions in place to obtain the highest reconstruction efficency while minimizing false positives. This leads to an overall efficiency for electron reconstruction for electrons from Z decay of ≈93 %.

Muons

To reconstruct muons, the tracks from the muon system and from the inner silicon tracker are treated separately at first [50]. The inner tracks are reconstructed by an algorithm based on Kalman-filters [49]. Contrary to electrons, this works for muons because of their much higher mass and their therefore lower radiation losses through bremsstrahlung. So called "standalone-muon tracks" are built by using the information from the muon system. Then the tracker muon tracks, that satisfy a transverse momentum of $p_T > 0.5\,\text{GeV}$ and a total momentum of $p > 2.5\,\text{GeV}$, are reconstructed by extrapolating tracker tracks from the tracker to the muon system. To be classified as a tracker muon, at least one muon segment has to match the extrapolated track. Global muons are muons that are built from the outside of the detector to the inside through the matching of standalone muon tracks of the muon system to inner tracks.

Because of the high efficiency of these reconstructions, about 99 % of muons can be reconstructed as either tracker or global muons, or both. Generally, muons with a low p_T are identified as tracker muons, as they often only reach the innermost muon segment. Late showering hadrons that pass the HCAL, might also be detected by the first muon station. Higher p_T muons usually pass through the whole muon system and are reconstructed as global muons.

These muons are then passed to the particle-flow algorithm [47]. There all the information from all subdetectors is combined. It applies a set of selection criteria to reconstructed candidates. These criteria are based on the quality parameters of the reconstruction. To calculate the momentum of the muons, the Tune-P algorithm is used. It selects the p_t measurement from a refit to reduce tails in the momentum resolution distribution caused by bat fits. The inner track is used if the p_t is smaller than 200 GeV, otherwise the track with the lowest χ^2 of it's fit is chosen. A muon that passes the particle-flow algorithm is excluded from being possibly reconstructed as another particle.

Photons

As illustrated in Fig. 2.10, photons only leave a hit in the ECAL, but no track in the tracker. Therefore, they are distinguished from electrons by their missing track and a difference in shower shape. Their energy is calculated similarly to the electrons. Bremsstrahlung photons are reconstructed and assigned to an electron by back-propagating their track to an electron trajectory. ECAL deposits with no compatible electron track are reconstructed as isolated photons if the ECAL cell energy distribution is compatible with the one expected from a photon shower. As hadrons might also leave some energy in the ECAL, photons are only reconstructed when there are no matching energy deposits in the HCAL [47].

Hadrons

As illustrated in Fig. 2.10, neutral hadrons do not leave tracks in the silicon tracker, but charged hadrons do. Both of them deposit most of their energy in the HCAL, but might also deposit some in the ECAL. Neutral hadrons are therefore reconstructed by the missing track and the HCAL shower, charged hadrons by their track in the silicon tracker and their energy deposits in the ECAL and HCAL. To calculate the momentum of charged hadrons, not only their track is used, but also the position of compatible ECAL and HCAL clusters [47].

Jets

Jets are showers of particles usually contained in a narrow cone. When two quarks are produced in a high energy collision, they travel in opposite direction in their center of mass frame. As they separate, the colour field is forming a tube in-between them, with an energy density of about $1\,\mathrm{GeV\,fm^{-1}}$. This quickly becomes enough energy stored in the field to create new quark-antiquark pairs, and therefore break the colour field into smaller "chunks" that are energetically more favourable [18]. Individually reconstructed particles are clustered together by the anti-k_t jet clustering algorithm [51] with distance parameter $R = 0.4$.

2.2.5 Simulation

Simulating data for particle physics enables a connection between theoretical predictions and experimental observations. Data of the CMS experiment is simulated by general-purpose event generators that implement a Monte-Carlo (MC) simulation [52].

Event generation is split into five steps. First, the hard scatter, which is the process with the highest momentum transfer of the event, is calculated in perturbation theory. This results in a probabilistic distribution of the outgoing partons, from input parton distribution functions. In the next step, the parton shower is simulated. The accelerated colour charges radiate gluons resulting in a cascade that is modelled. This is done until the energies are too low for the interactions to be modelled perurbatively [53].

In the third step, called hadronization, the partons from the second step are combined to colourless hadrons. As secondary interactions can take place in the protons, they can create secondary particles. These are in addition to the primary hard-scattering process and they form underlying events. In the fifth step, unstable hadrons that have been formed in the hadronization process decay [53]. Among the commonly used event generators are PYTHIA [54, 55], POWHEG [56, 57], HERWIG [58, 59] and MadGraph [60, 61].

To take the detector geometry and material into account, the GEANT4 [62] toolkit is used. This software also allows to consider the detector electronics by adding noise to the simulation.

The advantage of using simulation data is that we have the truth data of the generated events, this means that the parent particles of each particle are known. This enables the use of supervised learning techniques as applied in this thesis.

Chapter 3

Deep Learning

3.1 Introduction

Machine learning (ML) techniques have been used since the advent of computers [63] in the 1950s and have gained increased popularity since the 1980s and are heavily researched topics. These techniques involve computers to learn from data to carry out tasks without being explicitly programmed to do so. This makes many modern technologies like self driving cars possible, which would be too difficult to program manually. Typically ML techniques and algorithms can be split into their three main applications [64]:

- **Supervised learning**: The training data consists of input and output pairs ($\mathbf{X}$, $\mathbf{Y}$), with $\mathbf{X}$ a set of input parameters and $\mathbf{Y}$ the known output that the model should learn. The internal parameters of the model are optimised, by minimizing a figure of merit that quantifies the difference of the classifiers' prediction and the training label. Examples include the classification of images and the prediction of stock prices. Traditional techniques for supervised learning comprise of boosted decision trees [65], support vector machines [66] and artificial neural networks (ANN) [67].
- **Unsupervised learning**: This contains for example clustering, a method where unlabelled data is grouped to find new, meaningful information about its structure, so that data points, that share similar properties, are in the same group [68]. Typical examples are pattern recognition and fraud detection. Commonly used techniques include k-means clustering [69] and gaussian mixture models [70].
- **Reinforcement learning**: The model learns to make decisions in order to maximise a reward function that indicates the quality of the decisions. Therefore, the model learns the desired behaviour via trial and error. The underlying algorithms are often similar to the ones used in supervised learning. Popular applications for reinforcement learning are autonomous driving and path finding [71].

In particle physics supervised ML techniques are often referred to as multivariate analysis (MVA), because several variables are used simultaneously during the training of these algorithms. DL is a subgroup of ML and has become increasingly popular in recent years with the increasing capabilities of modern computing hardware and open source software frameworks such as tensorflow [72] and Keras [73].

Similarly to ML, DL can also be split into supervised, unsupervised and reinforcement learning. The key difference between ML and DL is how features get extracted. Typically ML approaches need to use handmade features and feature extraction techniques like Principle Component Analysis (PCA) and Recursive Feature Elimination (RFE) [74], as using the raw data as input

for its algorithms often does not lead to good results in many applications such as the processing of sound and image data.

Modern DL algorithms use a neural network (NN) architecture. NNs consist of various layers with different purposes. Some layers specialise in feature extraction while other layers can handle input data arrays of different length. This makes NNs extremely flexible and therefore they can be used to perform any tasks traditional ML techniques can with usually much better results. However, this often comes at the cost of larger computational requirements.

A big challenge in ML is the construction of good and reliable high level features. Thanks to their usually high capacities are NNs able to perform the feature extraction on their own, therefore DL enjoys great popularity in areas where feature extraction algorithms are lacking and impractical, such as in image and sound processing [75].

In this thesis DL was used to differentiate between prompt, non-prompt and fake leptons (see Sec. 4.2). Because the training data was generated from a MC simulation, the truth labels for these three classes are known. Therefore, this is a multi-classification problem.

3.2 Layer Types

A NN consists of an arrangement of different layers. This includes an input and an output layer, and a number of so called hidden layers that handle the processing. Each layer takes the output of the previous layer as its input.

Generally, the layers contained in a NN define the type. For example a neural network that contains convolutional layers is called convolutional neural network (CNN) and one that contains recurrent layers is called a recurrent neural network (RNN) [76]. But with most current DL frameworks, such as Keras [73], it is possible to combine these layers in almost any configuration.

3.2.1 Dense

A dense layer, also called fully connected layer, is a layer, where every output node is directly connected to every input node. It is the conceptionally simplest layer and very similar to a perceptron, that has been first theorised by Frank Rosenblatt in 1958 [67]. A simple perceptron can be used as a linear binary classifier [77]:

$$f(\vec{X}) = \begin{cases} 1, \text{ if } \vec{W} \cdot \vec{X} + b > 0 \\ 0, \text{ otherwise} \end{cases} \tag{3.1}$$

Where $\vec{X}$ is the input feature vector, $\vec{W}$ is the weight vector and b is the bias value.
This can be generalised to have more than one output [78]:

$$\vec{Y}(\vec{X}) = \boldsymbol{W} \cdot \vec{X} + \vec{b} \tag{3.2}$$

Where $\vec{X}$ is the input feature vector with a length of m, $\vec{Y}$ is the output vector with a length of n, $\boldsymbol{W}$ is the weight matrix with a dimensionality of n×m and $\vec{b}$ is the bias vector with a length of n.
A simple graph for this type of layer is given in Fig. 3.1.

A NN consisting of only one dense hidden layer already has the capacity to learn non-linear decision regions, which gives them a big advantage over the simple perceptron [79].

3.2.2 Simple RNN

RNNs behave differently than neural networks with only dense or convolutional [80] layers. Those networks are feed forward networks, as the information only propagates into the direction

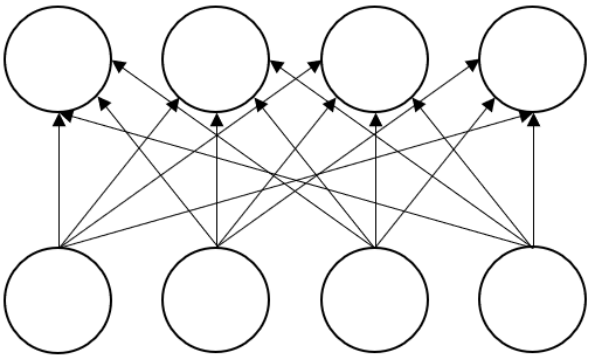


Figure 3.1: A graph that illustrates the dense layer with four input and output nodes.

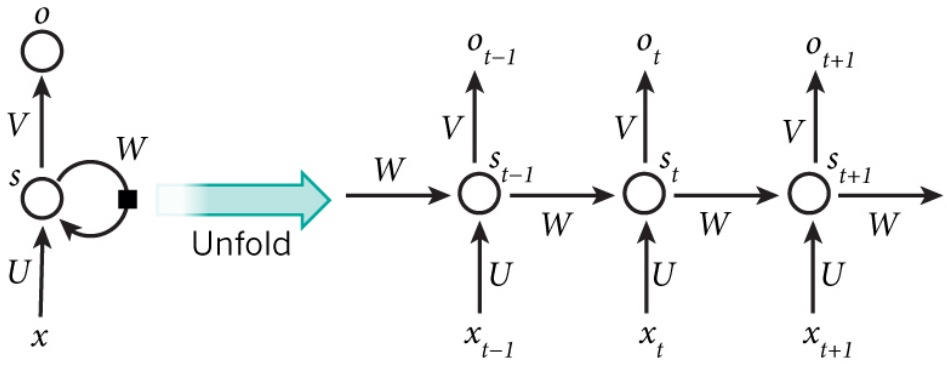


Figure 3.2: Loop structure of a RNN cell[82].

of the predicted output. An RNN feeds its output back into itself, whereas ANNs and CNNs only feed the information from the input towards the output. Therefore, its strength lies in its ability to process variable-length sequential data. Typical applications are in text analysis or weather forecasting [81]. In this case, they are used to sequentially read the particles near our lepton candidate.

The aim of RNNs is to store the information of an element and to reuse it as another feature for later elements in the sequence.
A single RNN cell works by feeding back information into itself for every input sequence, as illustrated in Fig. 3.2.

The Simple RNN is the simplest recurrent layer, as it simply combines the previous time steps output with the current time steps input, as shown in Fig. 3.3.

Because of their simple architecture, these simple RNNs suffer from the vanishing gradient problem, which will be explained in Sec. 3.5. This is due to a previous time steps input being only weakly correlated to a later time steps output, which means that the information of an earlier element is lost quickly and has only little impact on the processing of later elements. LSTMs and GRUs solve this problem with their more complicated cell architectures.

3.2.3 Long-Short term memory

The LSTM was proposed by Sepp Hochreiter in 1997 [3] and solves the problems of the simple RNN, as they were designed to avoid long-term dependency problems. They are based on the same principle as the simple RNN, but with a more complicated internal structure and an additional cell state, that is also passed from one time step to the next.
The following step-by-step explanation has been taken from [85] and uses the notation of Fig.

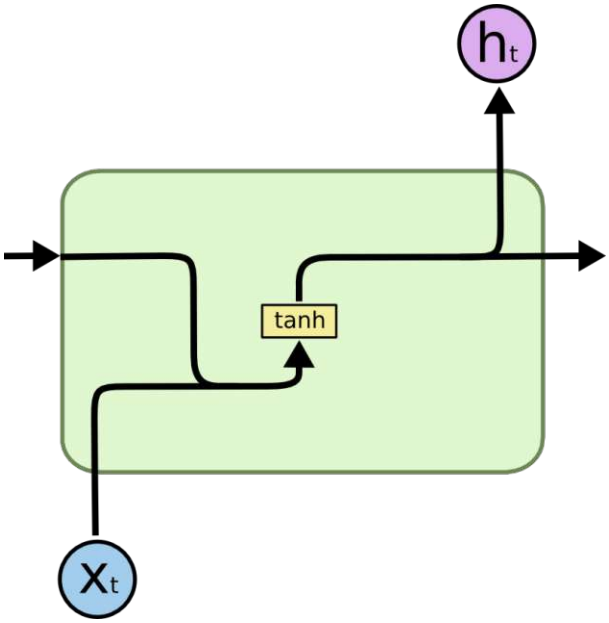


Figure 3.3: Schematic representation of a RNN cell [83].

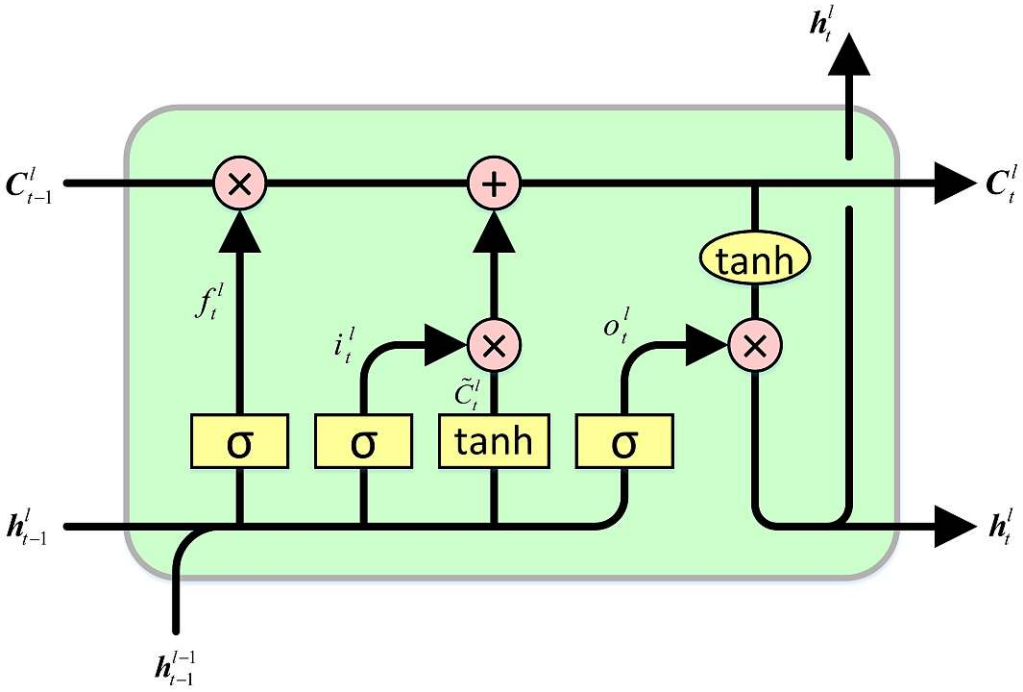


Figure 3.4: LSTM architecture[84].

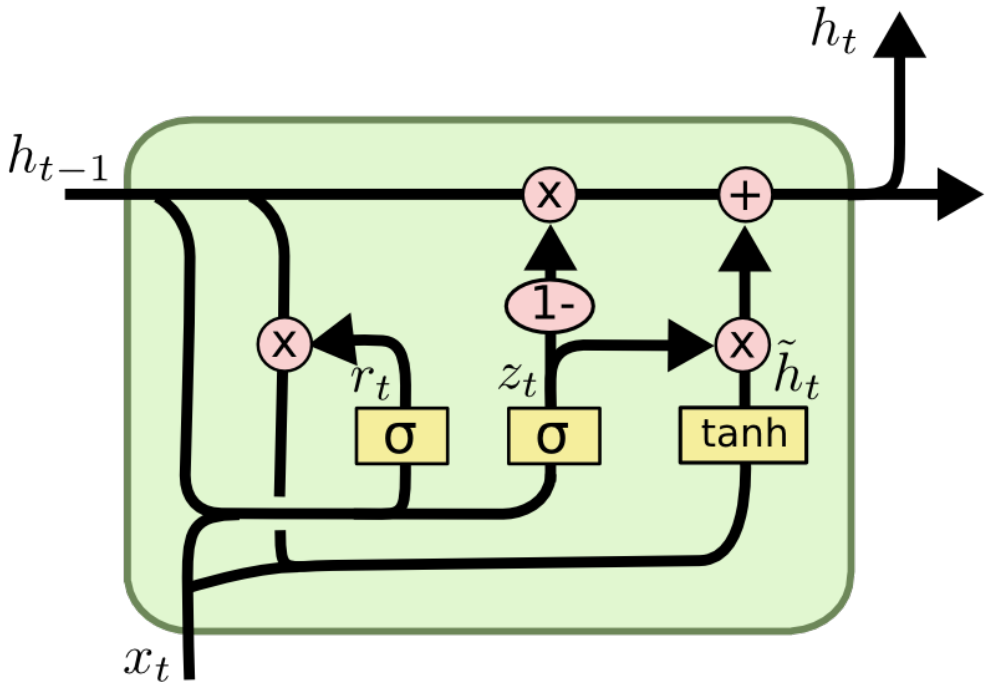


Figure 3.5: GRU architecture [85].

3.4:

First x_t and h_{t-1} are concatenated, then a decision is made on what information from the cell state is thrown away:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.3}$$

Then the information of what will be updated into the new cell state is calculated:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3.5}$$

So the new cell state becomes:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{3.6}$$

Finally, the output will be:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3.7}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{3.8}$$

C_t and h_t are then passed onto the next cell, and h_t is also used as input by the next layer.

3.2.4 Gated recurrent unit

The gated recurrent unit (GRU) has been proposed in 2014 [86]. Its operation principles are similar to the LSTMs, but it lacks the hidden memory state and also has some small changes that can be seen in Fig. 3.5:
According to the original paper [86] the performance of the GRU is comparable to the LSTM with a lower cost on system resources.

Following the same explanation as in the LSTM section, the GRU works as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{3.9}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{3.10}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \cdot h_{t-1}, x_t] + b_r) \tag{3.11}$$

$$h_t = (1 - z_t) \cdot h_t - 1 + z_t \cdot \tilde{h}_t \tag{3.12}$$

3.2.5 Bidirectional Wrapper

In Keras, there is the possibility to wrap the recurrent layer into the bidirectional layer. This wrapper trains the recurrent layer with the sequence in the positive and negative direction, afterwards the outputs of the forwards and backwards directions are typically concatenated and used as inputs for the next layer [87]. This was shown to lead to better results [88].

3.2.6 Batch Normalisation

Batch normalisation was proposed in 2015 by Sergey Ioffe and Christian Szegedy [89] in order to accelerate the training of a deep neural network. It works by applying a linear transformation to the ouput of the preceding layer, so that the mean output of a node becomes 0 and the standard deviation 1. This is done during training for every mini-batch. A mini batch is a small part of the training data that is evaluated and then used to update the networks internal parameters (see Sec. 3.5). Batch normalisation addresses the change in the distribution of each layer's inputs during training, which requires a much lower learning rate. Therefore, by using batch normalisation, the learning rate can be much larger and thus the training takes fewer training steps to complete. In the example given in the original paper [89], training with batch normalisation speeds up the training of a network by a factor of 5-14 for their image classification tasks and also improves the performance of these networks.

During the prediction, where the batch size might be one, normalising a layers output in this way would not be reasonable, therefore a moving average over the normalisation parameters of the last training batches is calculated and then used for the linear transformation [90].

3.2.7 Droput

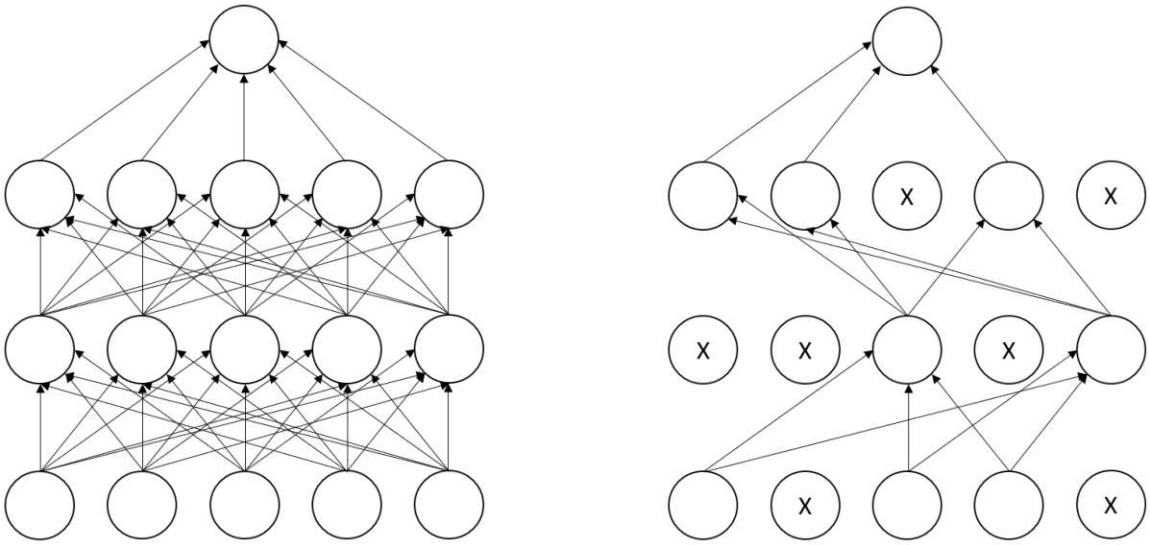
Dropout layers have first been theorized in 2014 [91] and are widely used as a simple, yet effective regularisation technique. A dropout layer randomly deactivates predefined amount of output nodes of the preceding layer, which is illustrated in Fig. 3.6. This results in the NN learning not to rely too much on individual neurons which often increases their performance on new data [92].

3.3 Activation Functions

Neural networks use activation functions on the output of each layer to introduce non-linearities. This is important as a dense NN would otherwise be a series of vector-matrix multiplications that could be simplified to be a linear function. Therefore, it would not have the capacity to perform the tasks for which they are used [93].

Traditionally, sigmoid activation functions have been used in hidden layers. These are functions that are similar to the hyperbolic tangent. As networks grow larger they often suffer from the vanishing gradient problem described in Sec. 3.6.2. To solve this, the ReLU function (rectified linear unit) is often used. It has the advantage that during training ReLU's gradient is very efficient to calculate, as the gradient is always either 0 or 1 [93]. One variation of the RelU function is the exponential linear unit (ELU), which is defined as

$$f(x) = \begin{cases} x \text{ if } x > 0 \\ e^x, \text{ otherwise} \end{cases} . \tag{3.13}$$



(a) Neural network without dropout.

(b) Neural network with dropout.

Figure 3.6: Illustration of a neural network with (a) and without (b) dropout activated.

3.3.1 Softmax

The softmax [94, 95] function is applied to the output of a NN for multiclassification tasks. It non-linearly transforms the output in a way that the per-class confidence scores add to one. This allows the final output to be interpreted as a probability:

$$s(\vec{v})_i = \frac{e^{v_i}}{\sum_{t=1}^{T} e^{v_j}} \text{ for } i = 1, ..., T \tag{3.14}$$

Where $\vec{v}$ is the predicted network output before softmax and T is the number of output classes.

3.4 Cross-entropy Loss

In order to evaluate the output of a network a loss function is needed. The loss function gives a quantitative measure on how much the predicted output of the network differs from the ground truth. The most commonly used loss function for classification in DL is the cross-entropy loss, defined in Eq. 3.15.

$$H(\vec{u}, \vec{v}) = -\sum_{t=1}^{T} u_t \ln v_t \tag{3.15}$$

Where $\vec{u}$ is a vector with the binary truth labels, $\vec{v}$ is a vector with the softmax ouput and T is the number of output classes. Depending on the intended use of the neural network, other loss functions have to be used, for example the mean squared error for regression tasks [96].

3.5 Gradient Descent

A NN is trained by minimising a suitable loss function. This is done via gradient descent, an iterative minimisation algorithm. The gradient for all internal parameters, weights and biases, is calculated with the backpropagation algorithm, which will be described in the next section. These parameters are then updated by a fraction of their gradient. This fraction is defined by the so-called learning rate. These steps are repeated until the gradient becomes small, which

means that a local minimum of the loss function was found. In practice the algorithm can also be stopped after a maximum number of iterations is reached [97].

One improvement on the standard gradient descent is the ADAM (adaptive moment estimation) [98] optimizer. During the training process the learning rate for each individual parameter is dynamically adapted by scaling their gradients inversely proportional to the L^2 norm of their past and current gradients. This leads to a convergence of the gradient descent algorithm with fewer iterations.

3.6 Backpropagation

There are two main ways to calculate the gradient of the loss function. Numerically by adding a small ϵ to each internal parameter of the network and revaluating it. In praxis, the analytical way, which is called back propagation, is much more computationally efficient, as many calculations can be stored and reused. This method was invented in 1976 by Seppo Linnainmaa [99]. It allows for the calculations to be distributed over many processing cores, as can be found in modern GPUs. This enables the training of very large NNs for various tasks such as self driving cars or the identification of leptons at the CMS experiment.

The calculation of the gradient via the backpropagation algorithm consists out of three main steps:

- **Forward pass**: In the forward pass the network is evaluated and all the intermediate values are stored.
- **Calculation of local gradients**: All local gradients are calculated independently.
- **Backward pass**: Local gradients are aggregated using the chain rule of derivation.

The backpropagation algorithm is the most integral part of DL, as it enables the training of large state of the art NNs. To give a better understanding of how this algorithm works, a minimal example for an arbitrary network of additions and multiplications, the operations found in NNs, is given in Figures 3.7-3.9. The blue cells represent the inital input values, the green cells represent internal operations. The goal is to calculate the gradient of the final cell h in dependence of the input values $a - d$.

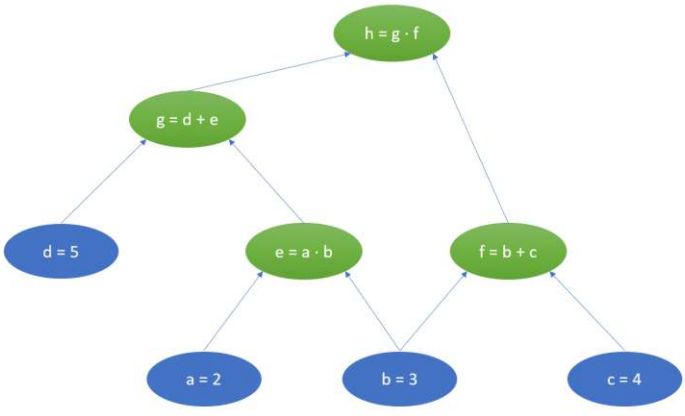


Figure 3.7: The layout of an arbitrary network for the minimal backpropagation example. The blue cells represent the input values, the green cells represent internal operations.

The network is evaluated and the intermediate results are stored:

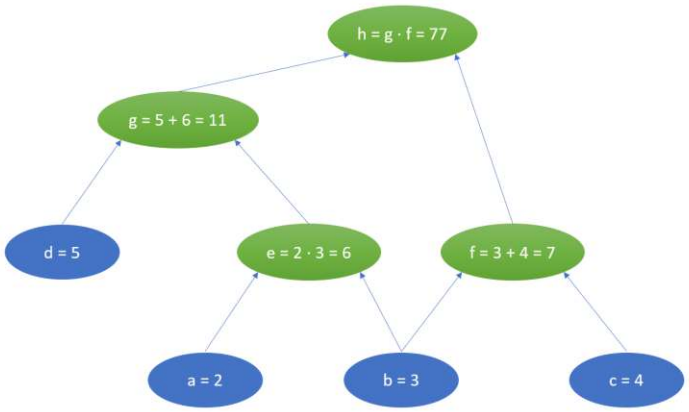


Figure 3.8: Forward pass: The network is evaluated and the intermediate values are stored.

The local gradients are calculated independently:

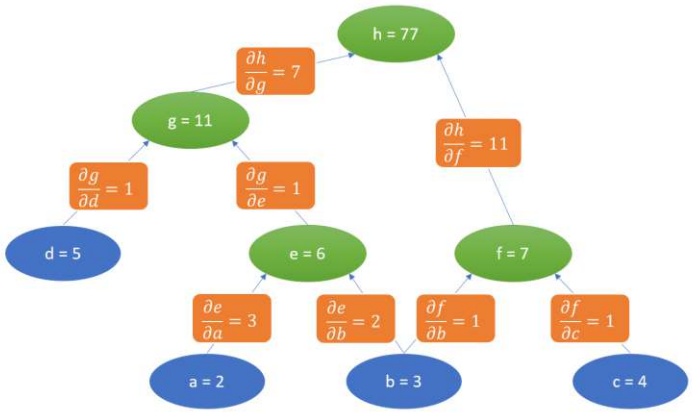


Figure 3.9: Calculation of local gradients: The local gradients are calculated independently (orange boxes).

Finally the local gradients are aggregated in the backward pass:

$$\begin{aligned}
\frac{\partial h}{\partial a} &= \frac{\partial h}{\partial g} \cdot \frac{\partial g}{\partial e} \cdot \frac{\partial e}{\partial a} = 21 \\
\frac{\partial h}{\partial b} &= \frac{\partial h}{\partial g} \cdot \frac{\partial g}{\partial e} \cdot \frac{\partial e}{\partial b} + \frac{\partial h}{\partial f} \cdot \frac{\partial f}{\partial b} = 25 \\
\frac{\partial h}{\partial c} &= \frac{\partial h}{\partial f} \cdot \frac{\partial f}{\partial c} = 11 \\
\frac{\partial h}{\partial d} &= \frac{\partial h}{\partial g} \cdot \frac{\partial g}{\partial d} = 7
\end{aligned} \tag{3.16}$$

3.6.1 Training

To train a network the data set is split into training and validation sets. The network is then trained on the training data set and after each iteration, also called epoch in this context, the network is evaluated with the validation data to validate that the training is converging. An example for the two resulting loss curves is shown in Fig. 3.10.

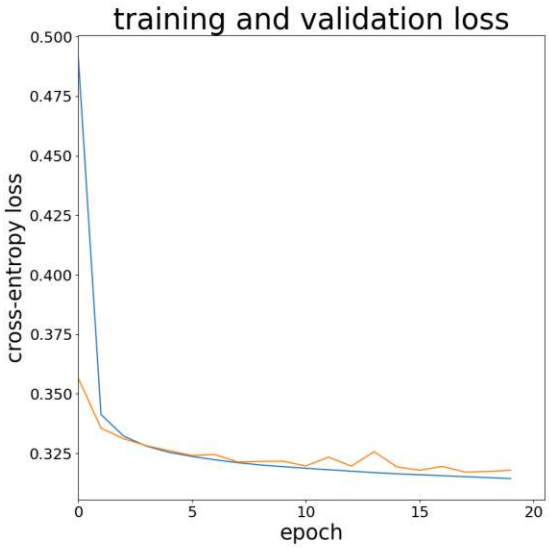


Figure 3.10: The training and validation loss curves of a typical training of the DeepLepton classifier. The network was trained for 24 hours and shows good convergence. The blue line represents the training loss and the orange line represents the validation loss.

3.6.2 Vanishing and Exploding Gradient

When the back propagation algorithm propagates away from the output layer, the gradients might get smaller and eventually approach zero. This leaves the weights of these layers almost unchanged. A possible cause is the sigmoid function being used as an activation function, as their gradients for large values are very small [100]. Alternatively, it could also be caused by the input having little correlation with the output, as it is often the case with simple RNN layers [101, 102]. As a consequence, the NN takes a lot longer to train.

The opposite effect is called exploding gradient. It happens when the aggregated gradients become very large, which causes the gradient descent algorithm to diverge and is often caused by the ReLU activation function, but can easily be solved by using a batch normalisation layer after each computational layer [100].

Chapter 4

Lepton Identification

4.1 Important processes with lepton final states

4.1.1 Drell-Yan process

The Drell-Yan process is the creation of a lepton pair in hadron collisions via an intermediate vector boson [103].This process was described by Sidney Drell and Tung-Mow Yan in 1970.

A quark-antiquark pair from the initial state annihilate to a virtual photon γ^* Fig. 4.1 (a) or a Z boson Fig. 4.1 (b). The Z boson has a probability of 10.1% to decay into a lepton-antilepton pair (e^+e^-, $\mu^+\mu^-$ and $\tau^+\tau^-$ in equal parts) [104]. Alternatively the Z boson can decay into various hadrons with a probability of 69.9%. The branching ratio into neutrinos is 20%.

In Fig. 4.2 the dimuon invariant mass distribution for events recorded in 2017 and 2018 at the CMS experiment is shown. An important feature of this spectrum are the peaks due to mass resonances with SM particles. The invariant mass peak corresponding to the Z boson mass is located at 91.2 GeV. This "standard candle" can be easily reconstructed even in dense proton-proton environments and is used for the calibration of particle detectors [105, 106].

In addition to being important for the detector calibration, this process also has a high significance for searches of new physics beyond the SM and for quantifying SM parameters. Due to the DY process having a clean final state, it is measured with high efficiency [106, 107]. Therefore the DY process is important for the accurate measurements of the production rates and masses of the Z and W bosons, as well as measuremts of the protons' parton distribution functions [18].

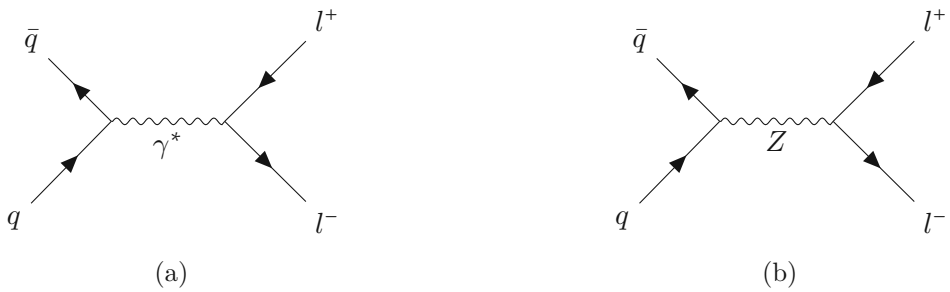


Figure 4.1: Feynman diagrams of the Drell-Yan process with the exchange of a virtual photon (a) and with a Z boson (b).

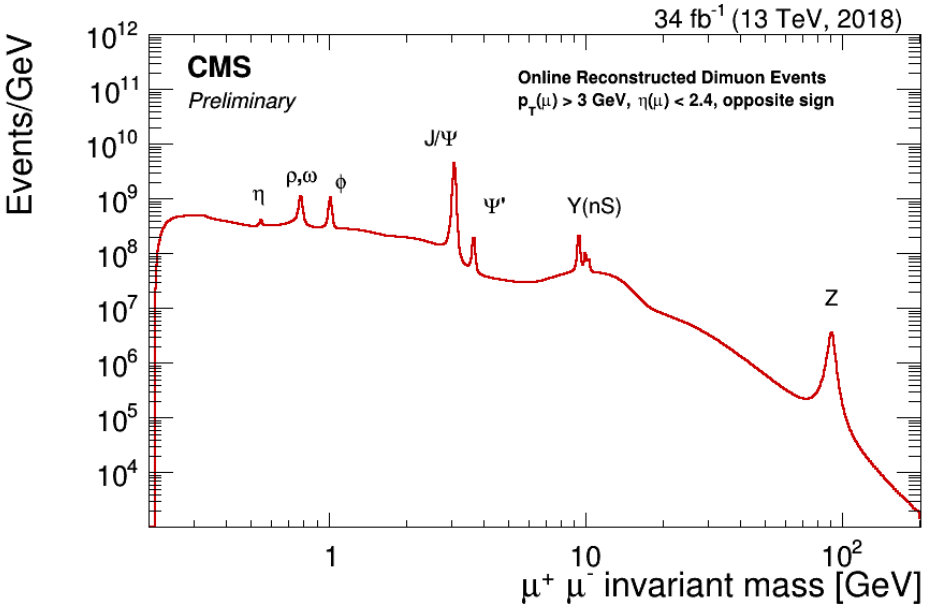


Figure 4.2: The dimuon invariant mass distribution for events recorded in 2017 and 2018 at the CMS experiment [108].

4.1.2 Top quark pair production

Another important process is top quark pair production. These are events where a $t\bar{t}$ pair is produced in the initial proton-proton collision through an interaction of the strong force. Leading order (LO) Feynman diagrams are shown in Fig. 4.3. A $t\bar{t}$ pair can be produced through the annihilation of a quark-antiquark pair (a) or through gluon-gluon fusion (b, c, d) [110]. The top quark is the heaviest fundamental particle with an invariant mass of 172.76 GeV [104]. It was discovered in 1995 at the Tevatron particle accelerator at Fermilab [111]. Due to its high mass and short lifetime of 4×10^{-25} s it decays before travelling a significant distance and before it hadronizes [112].

The top quark decays almost exclusively into a W^+ boson and a bottom quark. It could in principle decay into a down or strange quark, but these decays are suppressed by the very small values of the corresponding matrix elements of the Cabibbo-Kobayashi-Maskawa (CKM) [113, 114] matrix $|V_{td}|$ and $|V_{td}|$, given in Eq. 4.1 with the most current values revised in March 2020 [115], compared to the large value of $|V_{tb}|$. The CKM matrix is a unitary matrix that indicates the strength of interaction of the up, charm and top quarks with the down, strange and bottom quarks via the W boson. Due to the GMI mechanism [116] decays of the top quark via the Z boson are also suppressed.

$$|V_{CKM}| = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.041 \pm 0.0014 \\ 0.008 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{bmatrix} \quad (4.1)$$

The bottom quark hadronises, which leads to a jet in the subsequent event. The W boson decays into either a lepton and a neutrino with a probability of about 33% or into a hadron with a probability of 67% [104]. Therefore a top-antitop quark pair can either decay dileptonically ($l^+\nu_l$ and $l^-\bar{\nu}_l$) with a probability of about 11%, into a single lepton ($l^+\nu_l$ or $l^-\bar{\nu}_l$ and $q\bar{q}'$) with a probability of about 44% or into two quark-antiquark pairs with a probability of about 45%.This results in a minimum of either two, three or four hadron jets in an event.

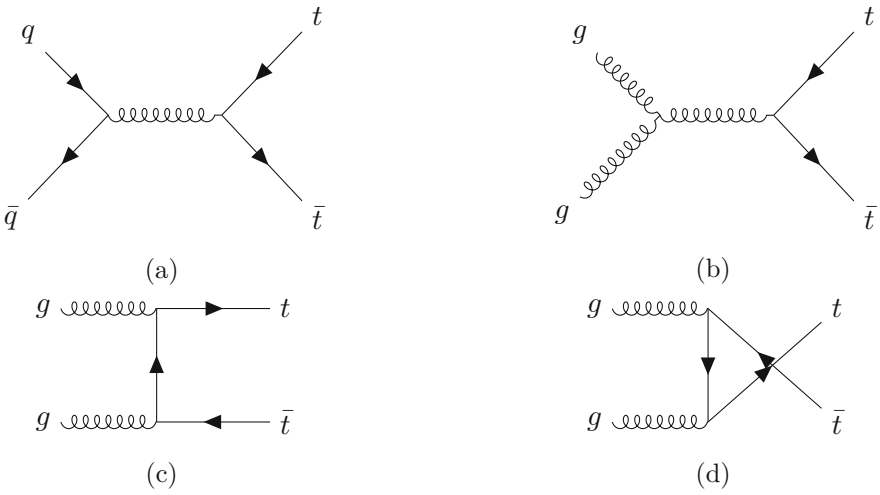


Figure 4.3: Four LO processes of top quark production, one by quark-antiquark annihilation (a) and the other three by gluon-gluon fusion (b, c, d). Diagrams after [109].

4.2 Lepton Classes

We classify leptons into three categories: Prompt, non-prompt and fake leptons. Prompt leptons are the signal category, while non-prompt and fake leptons are considered as background. This was also done in this thesis. Differentiating these theree classes with a high efficiency is important for many analyses to do accurate measurements with a high statistical significance.

Prompt leptons emerge from the primary interaction point from the decay of the vector bosons Z and W in processes. In this thesis leptons originating from DY and $t\bar{t}$ are considered as prompt. Non-prompt leptons originate from a secondary vertex (SV). This is an interaction point that occurs due to the decay of relatively long lived particles that travelled a significant distance in the detector. Often these are B-mesons that decay leptonically, or pions and kaons that decay before hitting the calorimeters [117]. As non-prompt leptons originate from SVs that are displaced from the primary vertex (PV), they differ from prompt leptons in their distributions of the impact parameter and isolation. The impact parameter is a lepton feature, that is the shortest distance of the reconstructed track from the primary vertex. A histogram plot for the impact parameter d_z is shown in Fig. 4.4. Another important feature for differentiating prompt and non-prompt leptons is the isolation. It provides a measure on the number of other particles in the vicinity around the lepton. The isolation number is usually high for lepton candidates produced inside hadronic jets and much smaller for prompt leptons. The relative isolation is defined as the p_T-sum of all tracks within a cone $R_{\text{cone}} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, centered on the lepton, divided by the lepton transverse momentum p_{T_lep} [118]:

$$I_{rel} = \frac{\sum_{i=1}^{N} p_{T,i}}{p_T, lep} \tag{4.2}$$

η is the pseudorapidity, a measure of angle along the beam axis, and ϕ is the polar angle perpendicular to the beam axis.

Fake leptons are lepton candidates that have been misidentified by the CMS detector and reconstruction software as leptons [119]. Therefore, these are in a separate category. Their defining feature is the absence of a lepton candidate at the generator level. Fake leptons have the following sources.

As electron reconstruction is based on the particle tracks in the silicon tracer and the corresponding energy deposits in the ECAL, the track of a charged hadron can be falsely linked with an ECAL shower caused by a photon. The distributions of the energy deposited by electrons

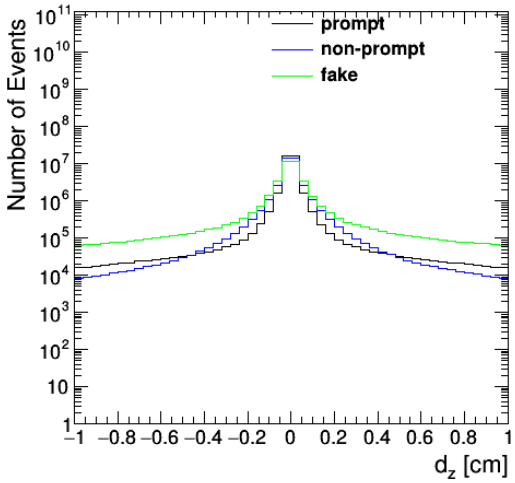


Figure 4.4: Distribution of the longitudinal impact parameter d_z in the electron channel using the DeepLepton training data from the year 2016.

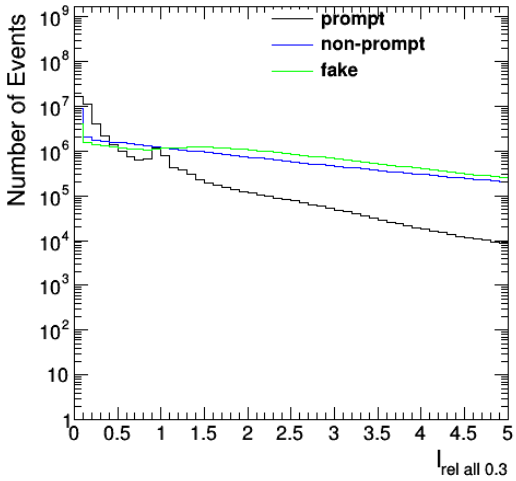


Figure 4.5: Histogram of the relative isolation distribution for $R_{\text{cone}} = 0.3$. The underlying data comes from the training data for electrons of the year 2016 in $t\bar{t}$ processes.

electron	muon
$d_{xy} < 0.02$	$d_{xy} < 0.02$
$d_z < 0.1$	$d_z < 0.1$
$\eta < 2.5$	$\eta < 2.4$
	if $p_T \leq 25$:
$I_{rel} \cdot p_T < 5$	$I_{rel} \cdot p_T < 5$
	if $p_T > 25$:
$I_{rel} < 0.02$	$I_{rel} < 0.02$
	candidate is particle flow muon
	candidate is global muon

Table 4.1: Prompt electron and muon requirements used in the cut-based lepton ID for the stops compressed search.

and photons is very similar, but the so called shower shape is a parameter that can be used to differentiate them. As electrons have a low mass, they radiate bremsstrahlung photons tangential to their track in the magnetic field of the solenoid. This results in a different shower shape for electrons and photons [48].

The muon reconstruction is based on the tracks in the muon system in combination with a matching track in the silicon tracker. Almost all hadrons are absorbed by the HCAL. Some particles manage to "puch through" the HCAL and cause hits in the muon system. Additionally a track in the muon system can be linked to an inner track by chance. A parameter that is commonly used to differentiate these fake muons from prompt and non-prompt muons is the segment compatibility, a measure for the match of hits in the individual muon chambers.

4.3 Cut-based lepton identification

The stops compressed cut based lepton ID is a cut-based lepton identifier used in a search for events that contain decays of the top squark [120], the supersymmetric partner of the top quark. One important part of the event selection of this search is the existence of exactly one low p_T electron or muon. In addition to this the lepton has to satisfy further criteria. For a current (at the time of writing) version of this search they are listed in Tab. 4.1 for electrons. In the results section these cuts will be used as a lepton selection to compare the performance of the isolation cut with the DeepLepton classifier to determine the theoretical improvements the usage of the DeepLepton classifier would bring.

4.4 A conventional multivariate lepton ID

A ML based approach to prompt lepton identification is the TTH lepton MVA. It was originally developed and used for the discovery of the $t\bar{t}H$ process [121]. It is implemented as a lepton feature in the analysis data framework and assigns every lepton a scalar value between 0 and 1 that can be interpreted as a confidence to differentiate between prompt and not-prompt leptons. The underlying architecture is based on BDTs and uses the same global lepton features as the classifier developed in this thesis.

Chapter 5

DeepLepton

The DeepLepton identification algorithm, that has been developed in this thesis, is based on the DeepJet framework [4], which is used to identify high p_T hadron jets. DeepLepton exploits the information of particles in the vicinity of the lepton. Both use TensorFlow [72] with the Keras [73] API as a frontend interface and the Python programming language as a basis. The training process i.e. prediction and backpropagation can be highly parallelised, therefore TensorFlow has the option of training the DL networks directly on a graphics processing unit (GPU). The GPU training is enabled by CUDA [122] which was developed by NVIDIA. The GPU used for training is a NVIDIA Tesla V100-GPU.

5.1 Overview and usage of Particle Flow Candidates

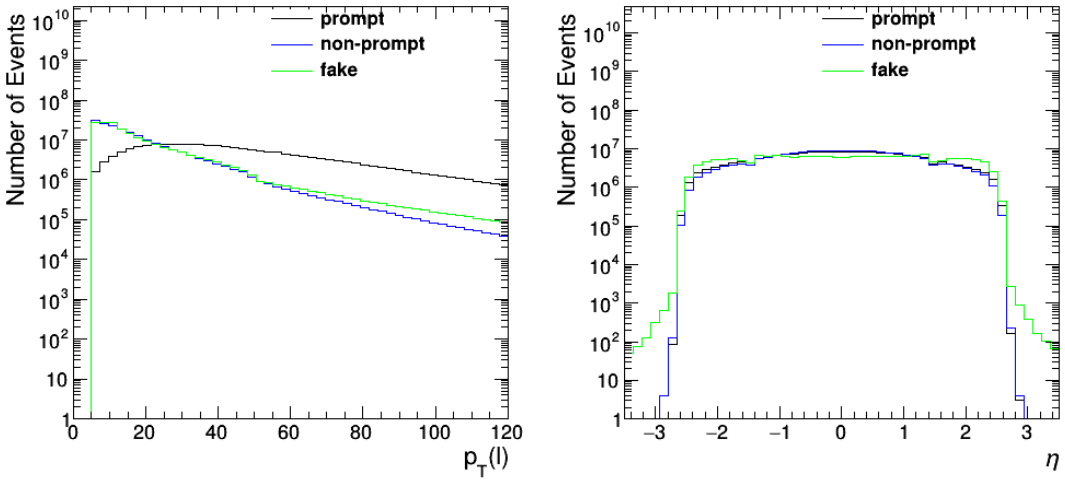
Lepton classification gradually improved during the course of the last decade. These approaches range from simple cut-based IDs to algorithms using ML and DL, such as the TTH lepton MVA. However, none of the previous techniques use the information of the other reconstructed particles in the vicinity of the lepton, the PF candidates, except for the isolation parameters.

Technically, the lepton candidate is represented by a list of features. The features of the PF candidates are exploited by also feeding them into the network. This poses a challenge, as not all leptons have the same amount of particles in their vicinity, which cannot be handled easily by ML techniques or DL without recurrent layers. Therefore, recurrent sub-networks are used, with lists of features for the PF candidates that are fed sequentially into these sub-networks. The output of these sub-networks is then concatenated with the output of a sub-network that processed the global lepton features. This forms the input for the final sub-network that outputs a probability for each lepton class. A more detailed discussion of the network's architecture can be found in Sec. 5.5.

The usage of the PF candidates gives the DeepLepton classifier an advantage in performance over previous algorithms, which will be shown in Sec. 6.1.

5.2 Training features

This section will discuss important features of lepton candidates that are often used for the classification of leptons. Additionally, features of hadrons, photons and secondary vertices will be explained, as the DeepLepton classifier implemented in this thesis also uses the particles in the vicinity of the lepton in an effort to gain more information about the lepton's creation. The detailed functionality will be explained in Sec. 5.1 and a full list of all features used for training with a short description can be found in Sec. 5.1, Sec. 5.2, Sec. 5.3 and Sec. 5.4.



(a) (b)

Figure 5.1: Distributions of p_T (a) and η (b) in the electron channel using the DeepLepton training data from the year 2016.

5.2.1 Electron and muon features

Electrons and muons are treated differently during reconstruction, resulting in slightly different lepton candidate features. I begin with discussing the features that are common to both lepton classes.

The most important features are p_T and η, the lepton candidate's transverse momentum and the pseudorapidity, which is defined as

$$\eta = -\ln\left(\tan(\frac{\theta}{2})\right), \tag{5.1}$$

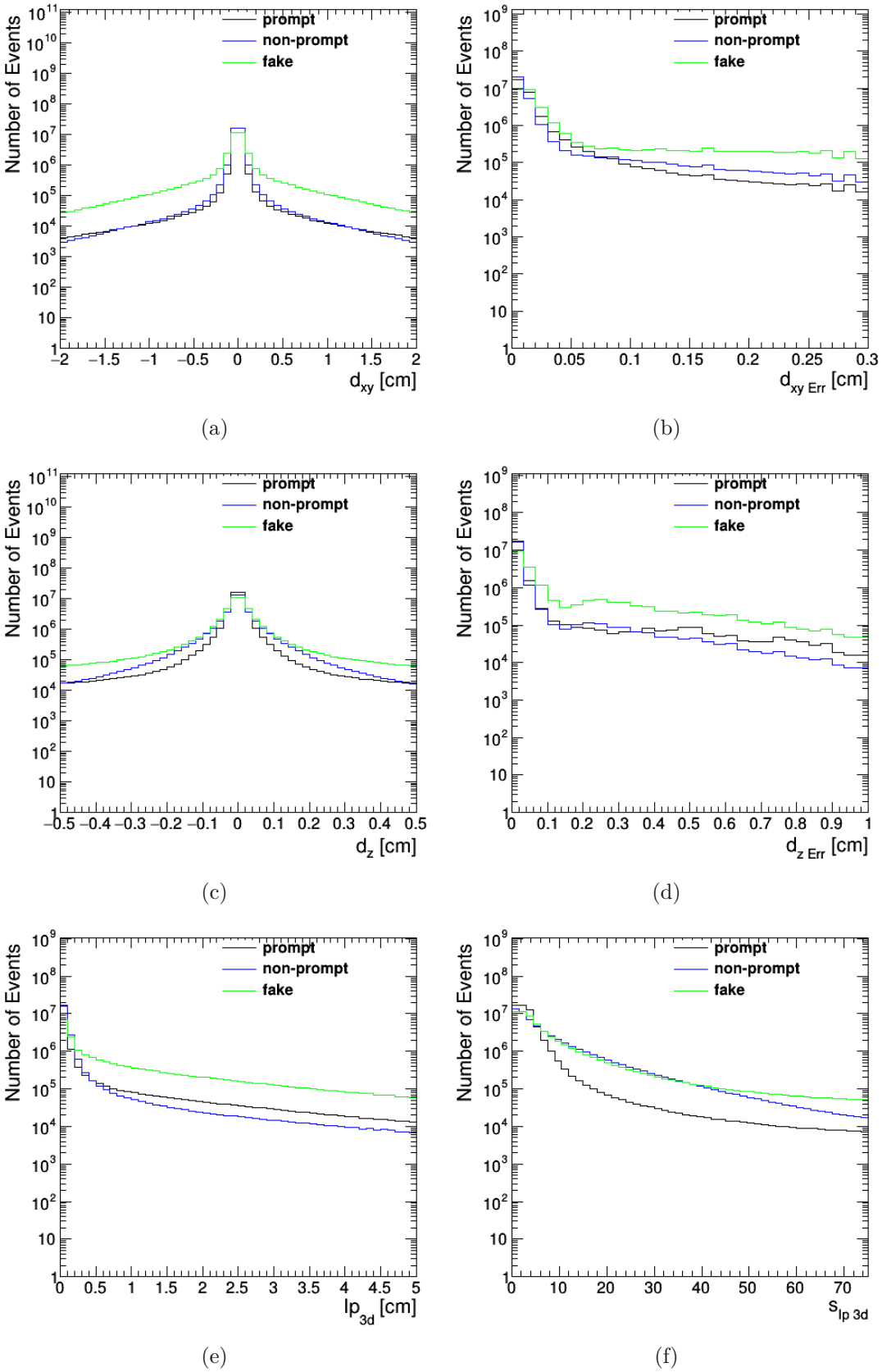
where θ is the lepton candidate's angle to the beam axis. A key advantage of η over θ is, that a angular difference $\Delta\eta$ between two particles is Lorentz invariant if their masses are negligible. Distributions of these two parameters are shown in Fig. 5.1.

A group of important features is related to the impact parameter d. They are shown in Fig. 5.2. d_{xy} represents the closest distance of the reconstructed track to the PV perpendicular to the proton beams and d_z is the closest distance parallel to the proton beam. The parameters $d_{xy\,Err}$ and $d_{z\,Err}$ correspond to their respective uncertainties. The 3D impact parameter Ip_{3d} gives the closest distance of the fitted track to the PV. The unit for these parameters is cm. Additionally, the significance of the 3D impact parameter $s_{Ip\,3d}$ is also used. In general, the distributions of these parameters are sharper for prompt leptons, while they are broader for non-prompt and fake leptons.

A second group of parameters that is important in lepton classification are the measures of the lepton isolation. The relative isolation is defined as the p_T-sum of all tracks within a cone $R_{\text{cone}} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, centered on the lepton candidate ℓ, divided by the lepton candidate's transverse momentum $p_{T,\ell}$ [118]:

$$I_{rel} = \frac{\sum_{i=1}^{N} p_{T,i}}{p_T, \ell} \tag{5.2}$$

There are versions of this parameter with a $R_{\text{cone}} = 0.3$ and $R_{\text{cone}} = 0.4$, with all particles in the vicinity of the lepton $I_{rel\,all}$ and with only charged particles in the vicinity of the particle $I_{\text{rel charged}}$. And "mini" versions of the isolation, $I_{\text{rel mini all}}$ and $I_{\text{rel mini charged}}$, which were



(a) (b) (c) (d) (e) (f)

Figure 5.2: Distributions for the lepton variables related to the impact parameter d in the electron channel using the DeepLepton training data from the year 2016.

introduced to improve the efficiency of IDs based on isolation for very high p_T muons. These function by varying the cone size R_{cone} between 0.2 and 0.05 as a function of p_T between 50 and 200 GeV. Similar to the isolation are the relative momentum of the lepton to the closest jet $p_{t\,rel jet}$ and the relative isolation in the matched jet $I_{\text{rel jet}}$. Examples for the distribution of these variables are given in Fig. 5.3. The distributions for prompt leptons are less broad than the distributions of non-prompt and fake leptons, except for the distribution of the $p_{T\,rel\,jet}$ variable, where the opposite is true. This is to be expected, as non-prompt and fake leptons are mostly found within jets.

5.2.2 Electron Features

Other important electron features are related to the ECAL shower shape of the electron, such as $\sigma_{i\eta i\eta}$ [123] and R_9, which correspond to the width and shape of the ECAL shower respectively. R_9 is defined as the ratio of energy deposits within ECAL crystal areas

$$R_9 = \frac{E_{3\times3}}{E_{\text{supercluster}}}, \tag{5.3}$$

where $E_{3\times3}$ is the energy deposited in a 3×3 crystal array and $E_{\text{supercluster}}$ is the energy deposited in the local supercluster [124]. Their distributions are shown in Ref. 5.4. These are especially useful when segregating prompt from fake electrons.

Furthermore, features corresponding to the electrons energy are used. Their distributions are given in Fig. 5.5, H/E is the fraction of the HCAL energy deposit over the ECAL energy deposit and $\Delta\eta_{\text{SC}}$ is the difference in pseudorapidity between the electron and the local supercluster. The variable E_{Corr} gives the ratio of the calibrated energy over the miniAOD energy and $E^{-1} - p^{-1}$ represents the difference of the inverse energy and the inverse momentum. The energy error E_{err} is the error of the cluster-track combination. A common feature of these five distributions is that the shapes for prompt and non-prompt are mostly very similar while the distributions for fake electrons are noticeably broader.

The features $\text{ID}_{\text{cut Based}}$ and ID_{noIso} are both identifiers introduced in 2017 and are a cut-based boolean selector and based on a MVA respectively. ID_{noIso} does not use the isolation parameters. Their distributions are shown in Fig. 5.6.
The electron lepton feature of note is the number of lost hits $N_{\text{hits lost}}$ in the tracker. The distribution of this feature is given in Fig. 5.7. As it can be expected, fake electrons have more lost hits than prompt and non-prompt electrons.

5.2.3 Muon Features

Further muon features of note include the two IDs $\text{ID}_{\text{low}p_{\text{T}}}$ and $\text{ID}_{\text{medium}}$. $\text{ID}_{\text{low}\,p_{\text{T}}}$ is an MVA ID with the aim of identifying prompt leptons with high efficiency and $\text{ID}_{\text{medium}}$ is a cut-based muon ID that requires the muon to be reconstructed as either a global-muon or as an arbitrated tracker-muon and satisfies additional track quality requirements. It was designed to be an efficient selector for prompt muons [125]. Distributions for these two features are shown in Fig. 5.8.

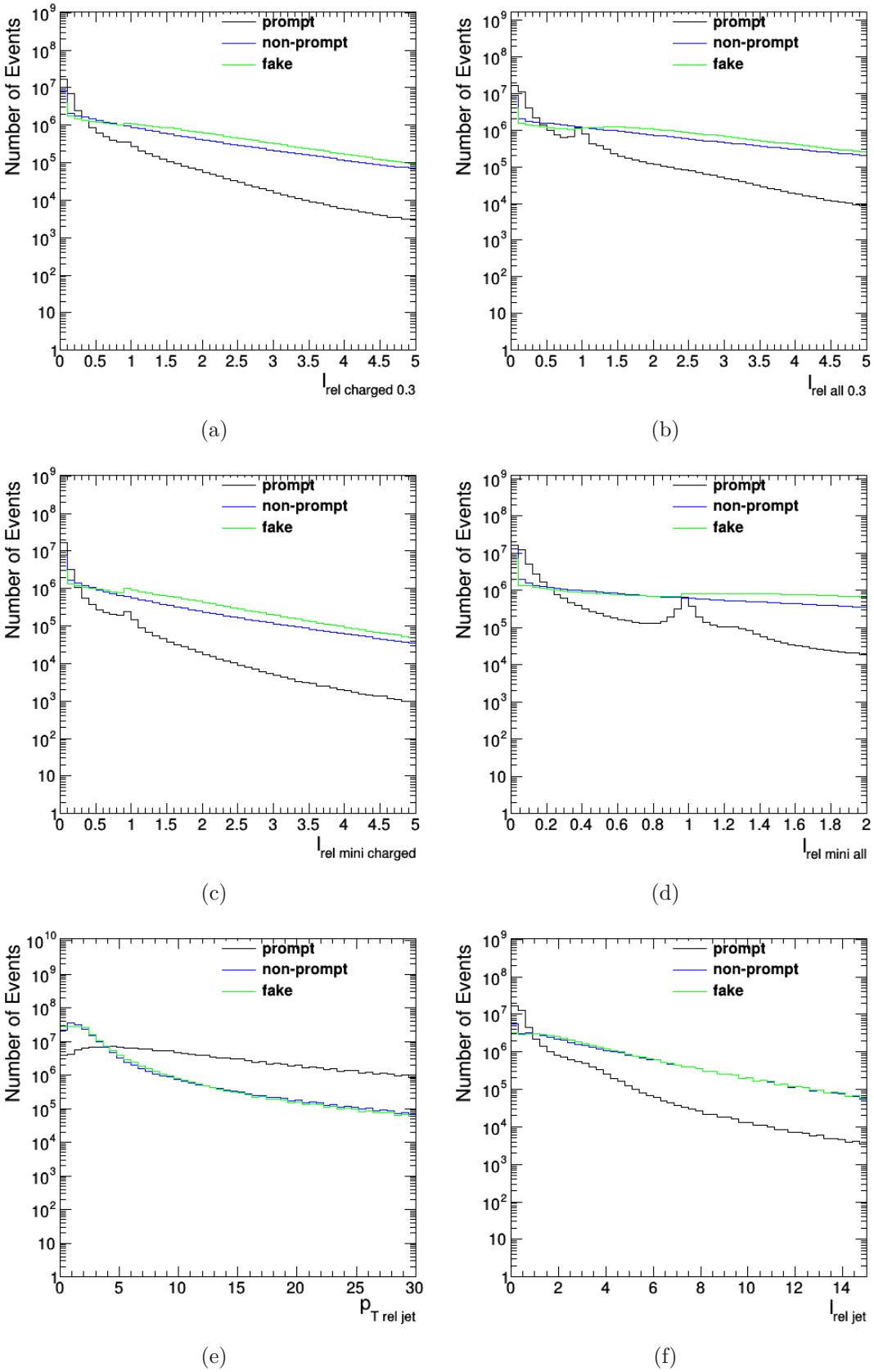
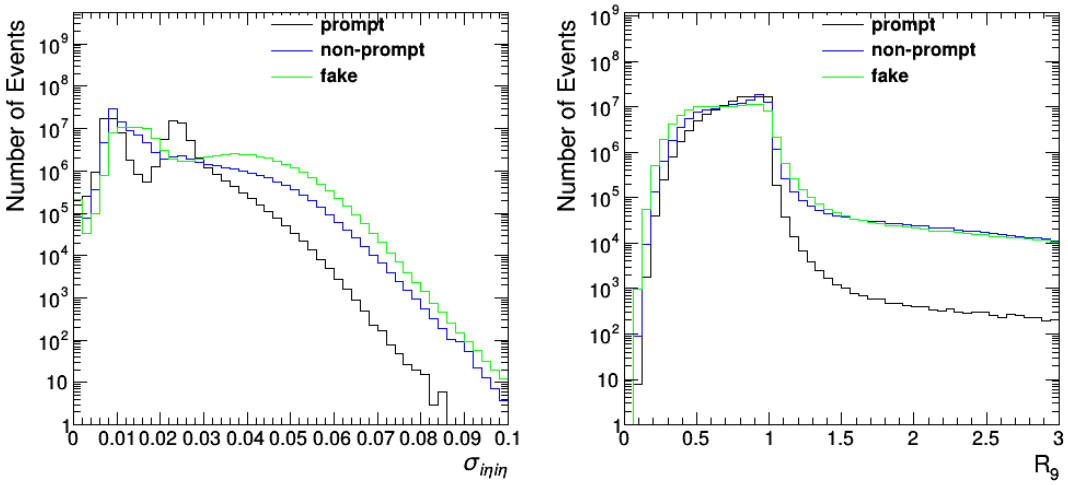
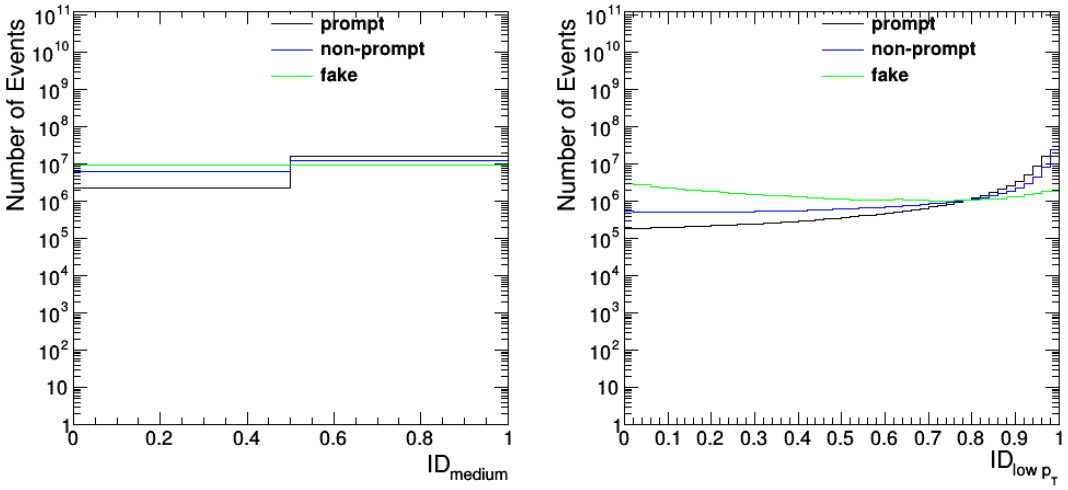


Figure 5.3: Distributions for the lepton variables $I_{\text{rel charged}}$ (a), $I_{\text{rel all}}$ (b), $I_{\text{rel mini all}}$ (c), $I_{\text{rel mini charged}}$ (d), $p_{\text{T rel jet}}$ (e) and $I_{\text{rel jet}}$ (f) in the electron channel using the DeepLepton training data from the year 2016.



(a) (b)

Figure 5.4: Distributions for the lepton variables $\sigma_{i\eta i\eta}$ (a) and R_9 (b) in the electron channel using the DeepLepton training data from the year 2016.



(a) (b)

Figure 5.8: Distributions for the lepton variables $\mathrm{ID_{medium}}$ (a) and $\mathrm{ID_{low\,p_T}}$ (b) in the muon channel using the DeepLepton training data from the year 2016.

As muons are detected inside the silicon tracker and in the muon system, parameters regarding the track fit quality and compatibility of the inner and outer track are used to separate fake leptons. These include the so called segment compatibility, a measure of the compatibility of hits in the individual muon chambers. Additionally, the number of matched muon stations N_{stations} and the number of tracker layers with hits $N_{\mathrm{tracker\,layers}}$ are used. These numbers are typically higher for non-fake leptons. $p_{T\,err}$ represents the uncertainty of the transverse momentum of the muon track. The distributions of these variables are shown in Ref. 5.9.

5.2.4 Particle Flow Candidate Features

As mentioned before, the PF candidates form an integral part of the DeepLepton classifier. They share many of the features of the leptons with some additional features. PF candidates only

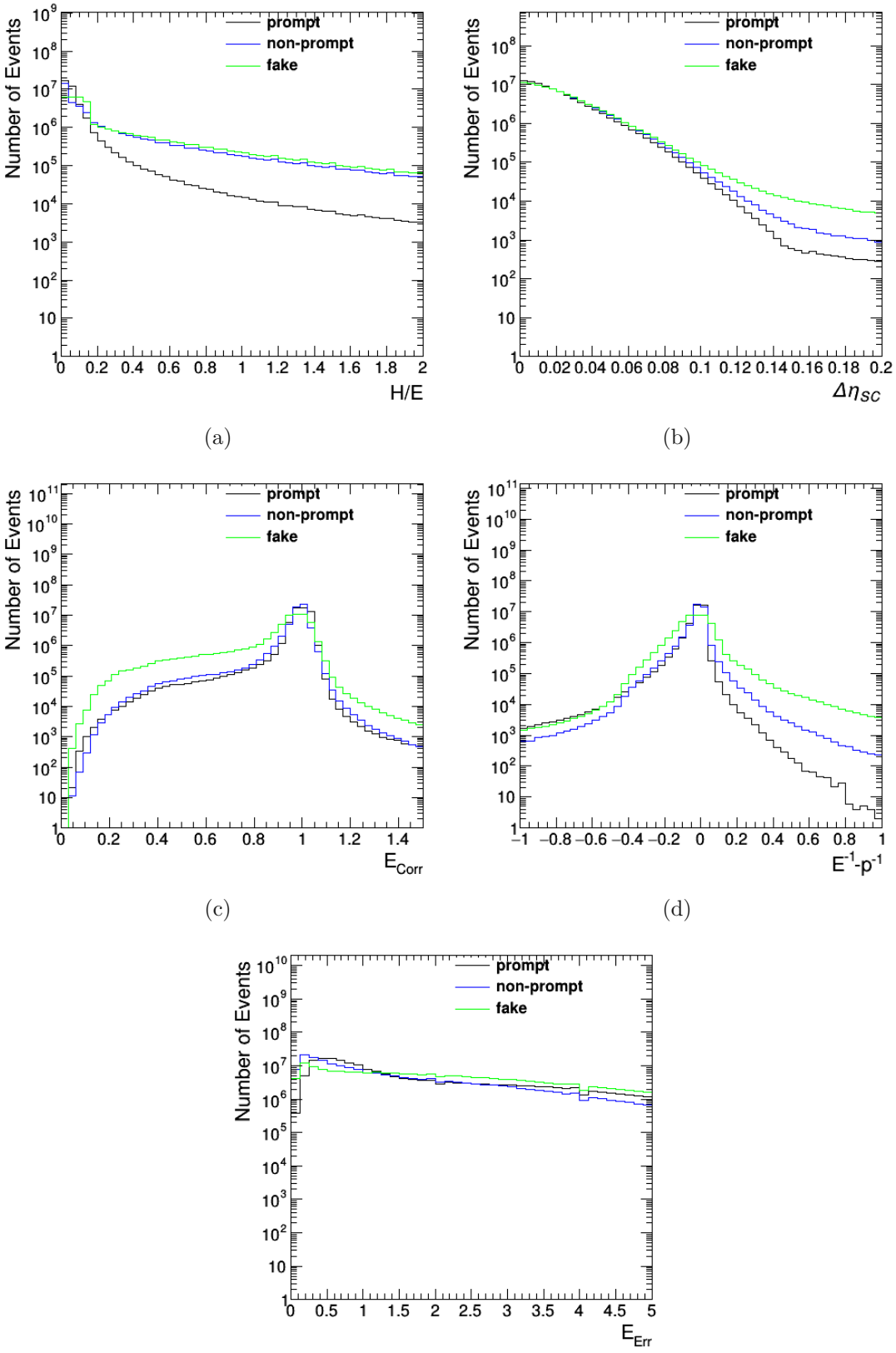
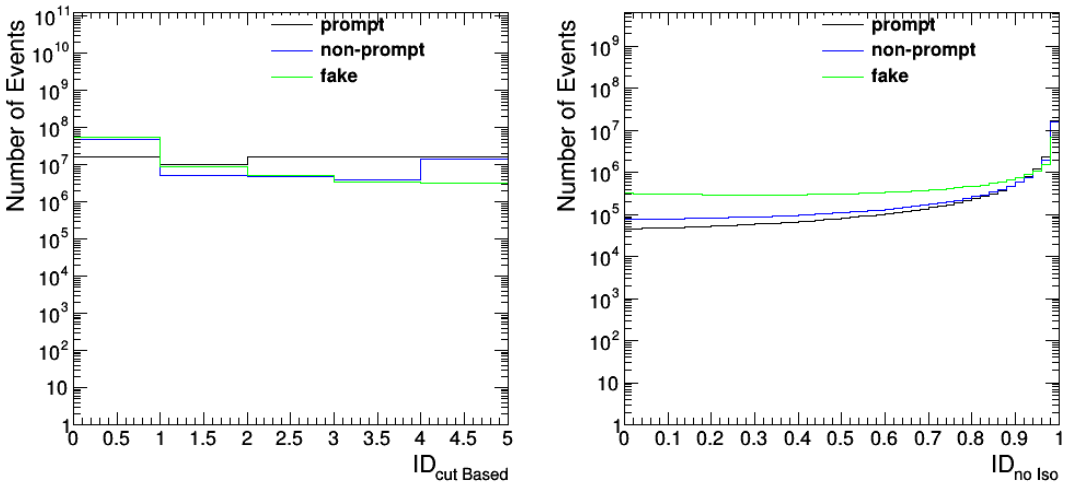


Figure 5.5: Distributions for the lepton variables H/E (a), $\Delta\eta_{\mathrm{SC}}$ (b), E_{Corr} (c), $E^{-1} - p^{-1}$ (d) and E_{err} (e) in the electron channel using the DeepLepton training data from the year 2016.



(a) (b)

Figure 5.6: Distributions for the lepton variables $\mathrm{ID}_{\mathrm{cut\,Based}}$ (a) and $\mathrm{ID}_{\mathrm{noIso}}$ (b) in the electron channel using the DeepLepton training data from the year 2016.

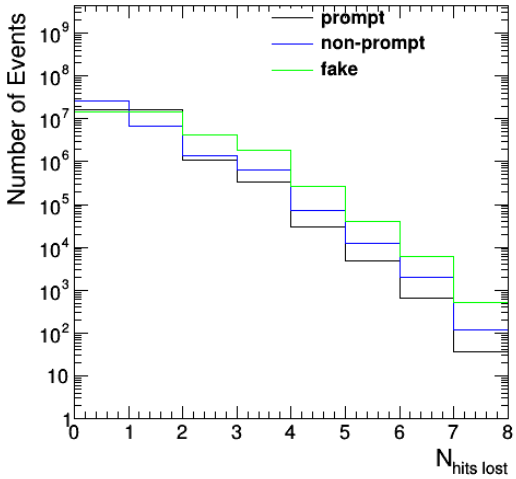
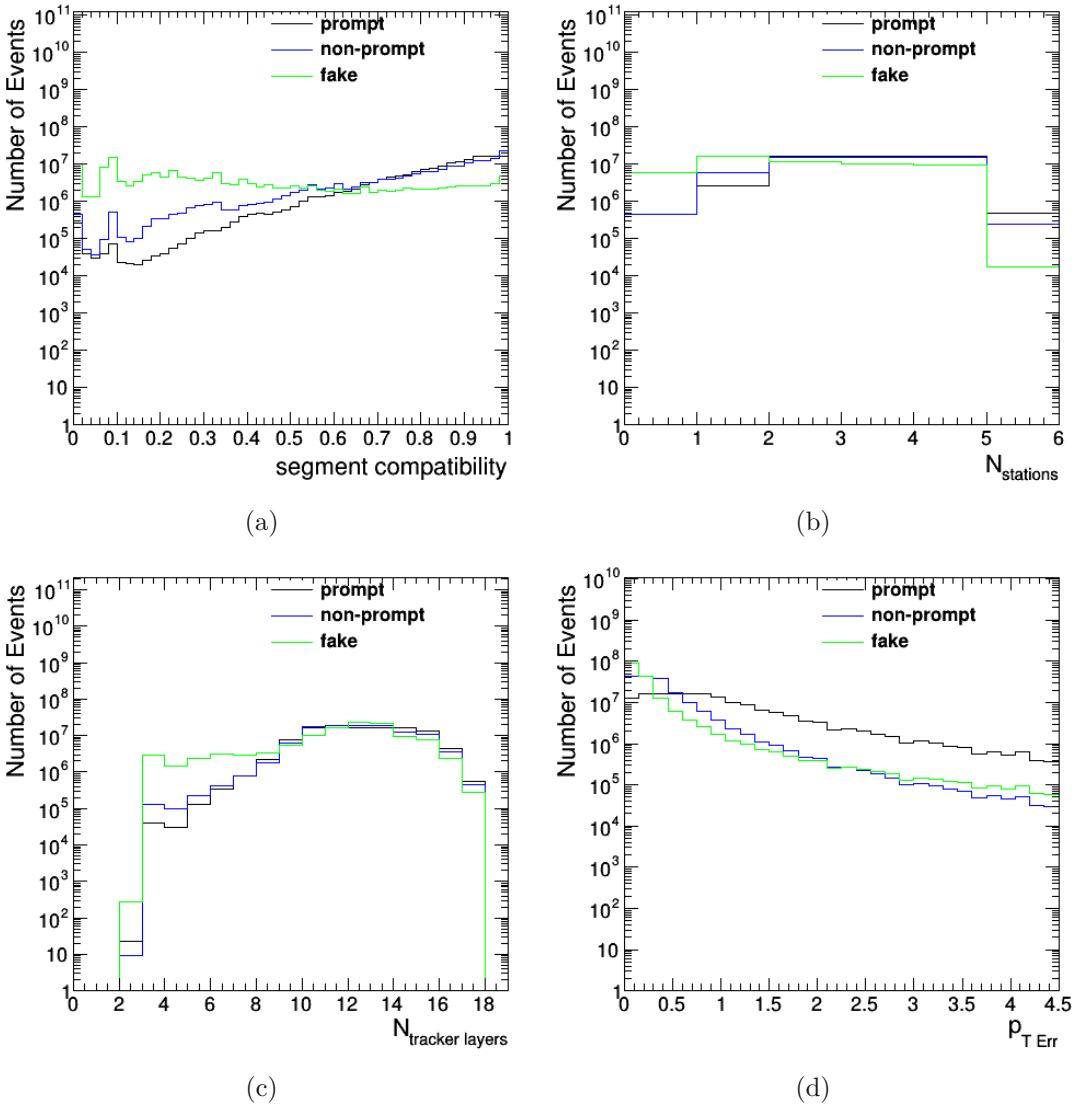


Figure 5.7: Distribution for the lepton variable $N_{\mathrm{hits\,lost}}$ in the electron channel using the DeepLepton training data from the year 2016.



(a) (b)

(c) (d)

Figure 5.9: Distributions for the lepton variables segment compatibility (a), N_{stations} (b), $N_{\text{tracker layers}}$ (c) and $p_{T\,err}$ (d) in the muon channel using the DeepLepton training data from the year 2016.

Feature	Description
ϕ	polar angle
charge	electric charge
d_{xy}	perpendicular impact parameter, in cm
d_z	longitudinal impact parameter, in cm
$d_{xy\,Err}$	d_{xy} uncertainty, in cm
$d_{z\,Err}$	d_z uncertainty, in cm
I_{p3d}	3D impact parameter, in cm
$s_{Ip\,3d}$	3D impact parameter significance
$I_{\text{rel all}\,0.3}$	relative isolation $\Delta R = 0.3$, all particles
$\text{ID}_{\text{rel charged}\,0.3}$	relative isolation $\Delta R = 0.3$, charged particles
$I_{\text{rel mini all}}$	mini relative isolation, all particles
$I_{\text{rel mini charged}}$	mini relative isolation, charged particles
$I_{\text{rel jet}}$	relative isolation in matched jet
$p_{\text{T rel jet}}$	relative momentum of the lepton with respect to the closest jet after subtracting the lepton
ID_{PDG}	PDG code assigned by the event reconstruction (not by MC truth)
$\text{ID}_{\text{cut Based}}$	cut-based ID Fall17 V2
$\text{ID}_{\text{no Iso}}$	mva ID without isolation
R_9	ECAL shower shape
$\sigma_{i\eta i\eta}$	ECAL shower width
H/E	ratio of HCAL and ECAL energy deposit
$\Delta\eta_{\text{SC}}$	η difference between electron and local supercluster (SC)
E_{Corr}	ratio of the calibrated energy/miniaod energy
$E^{-1} - p^{-1}$	difference of the inverse energy and momentum
E_{err}	energy error of the cluster-track combination
$N_{\text{hits lost}}$	number of missing tracker hits
convVeto	pass conversion veto
`vidNestedWPBitmap`	VID compressed bitmap
`dr03EcalRecHitSumEt`	non-PF ECAL isolation within a ΔR_{cone} of 0.3 with electron $p_T > 35\,\text{GeV}$
`dr03HcalDepth1TowerSumEt`	non-PF HCAL isolation within a ΔR_{cone} of 0.3 with electron $p_T > 35\,\text{GeV}$
`dr03TkSumPt`	non-PF track isolation within a delta ΔR_{cone} of 0.3 with electron $p_T > 35\,\text{GeV}$

Table 5.1: Electron training features.

features include the so called PUPPI weights, the pileup per particle identification weights. They indicate the probability of the particle being from pileup, which means the particle originates from another collision in the same bunch crossing. Furthermore, χ^2 measures of the track fit of the charged types of PF candidates, electrons, muons and charged hadrons, are used.

Two features that are used by all types of PF candidates are the absolute of the relative transverse momentum and the absolute of the angular difference of the PF candidate to the lepton in question. These features are also used for the SVs, which also use the position of the vertex, the pointing angle of the decays, the decay length and the number of degrees of freedom of the fit as features.

The multiplicities of PF candidates and SVs in the vicinity of the lepton are shown in Fig. 5.10. These distributions show a big similarity to the distributions of the relative isolation parameters and the prompt leptons have a lot fewer particles in their vicinity.

5.2.5 Training Feature Lists

The following tables, Tab. 5.1, Tab. 5.2, Tab. 5.3, Tab. 5.4 and Tab. 5.5, show the full lists of training features, the most important ones have been discussed in the previous sections. The explanations have been taken from [126] and from the nanoAOD ROOT-branch descriptions. Plots for the electron and muon input variables for all trainings are given in Appendix B.

5.3 Selection of MC Data Samples

For the generation of training data, MC samples were used. These contain data from the years 2016, 2017, and 2018, by using the samples from the `RunIISummer16miniAODv3`, `RunIIFall17miniAODv2` and `RunIIAutumn18miniAODv1` campaigns.

The data is taken for two types of training. One type will take prompt leptons from DY samples and non-prompt and fake leptons from QCD samples. These training data sets will be referred to as DYvsQCD. The other training data sets are comprised of samples that include events with top quark pair production and will be referred to as $t\bar{t}$. In these samples all leptons corresponding to their class are used. A summary for all training samples can be found in Tab. 5.6. Networks

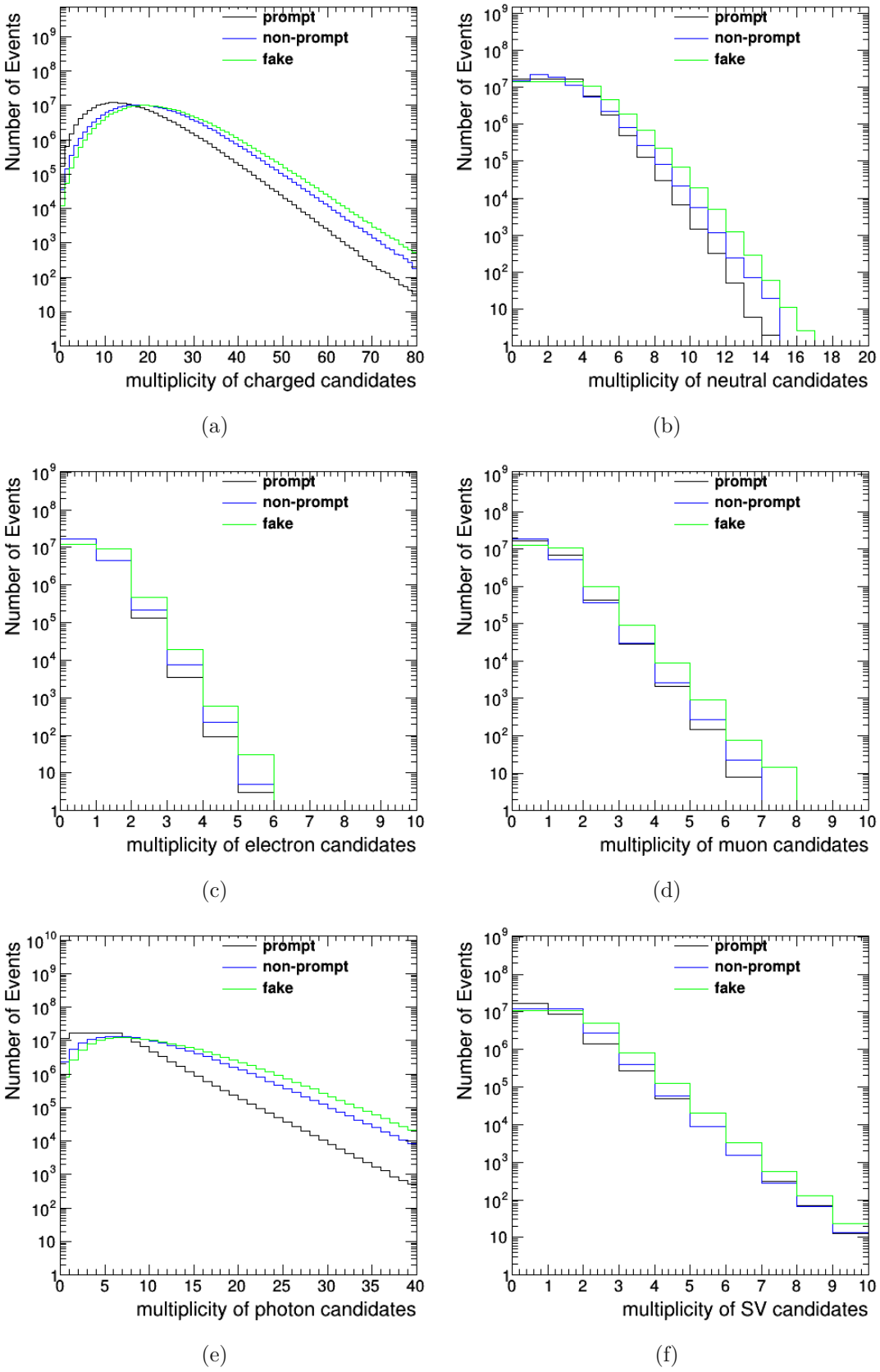


Figure 5.10: Distributions for the PF candidate multiplicities of charged hadrons (a), neutral hadrons (b), electrons (c), muons (d) and photons (e) in the electron channel using the DeepLepton training data from the year 2016.

Feature	Description
ϕ	polar angle
charge	electric charge
d_{xy}	perpendicular impact parameter, in cm
d_z	longitudinal impact parameter, in cm
$d_{xy\,Err}$	d_{xy} uncertainty, in cm
$d_{z\,Err}$	d_z uncertainty, in cm
$I_{p_{3d}}$	3D impact parameter, in cm
$s_{Ip\,3d}$	3D impact parameter significance
$I_{\text{rel all}\,0.3}$	relative isolation $\Delta R = 0.3$, all particles
$I_{\text{rel charged}\,0.3}$	relative isolation $\Delta R = 0.3$, charged particles
$I_{\text{rel all}\,0.4}$	relative isolation $\Delta R = 0.4$, all particles
$I_{\text{rel mini all}}$	mini relative isolation, all particles
$I_{\text{rel mini charged}}$	mini relative isolation, charged particles
$I_{\text{rel jet}}$	relative isolation in matched jet
$I_{\text{rel tk}}$	tracker-based relative isolation $\Delta R = 0.3$ for high p_T
$p_{\text{T rel jet}}$	relative momentum of the lepton with respect to the closest jet after subtracting the lepton
$p_{T\,rel}$	tuneP relative p_T, tuneP p_T/p_T
$p_{T\,err}$	p_T Error of the muon track
$N_{\text{tracker layers}}$	number of layers in the tracker
N_{station}	number of matched stations with default arbitration (segment & track)
segment compatibility	compatibility of inner and outer muon track
$\text{ID}_{\text{medium}}$	cut-based ID, medium WP
$\text{ID}_{\text{low p}_\text{T}}$	low pt muon ID score

Table 5.2: Muon training features.

Feature	Description
p_T	transverse momentum
η	pseudorapidity
phi	polar angle
charge	charge
m	mass
$p_{T\,rel}$	absolute difference in transverse momentum to lepton candidate
ΔR	angular difference to lepton candidate
d_0	distance to PV
$d_{0\,Err}$	error of d0
d_z	longitudinal impact parameter
$d_{z\,Err}$	error of d_z
$\text{PUPPI}_{\text{weight}}$	pileup per particle identification weight
$\text{PUPPI}_{\text{weight no},\ell}$	pileup per particle identification weight without the lepton
χ^2_{trk}	normalized χ^2 of the track fit
χ^2_{vtx}	χ^2 of the vertex fit
$N_{\text{hits lost}}$	lost inner hits
`pvAssocQuality`	primary vertex association quality
`trkQuality`	track quality mask

Table 5.3: Charged hadron, electron and muon PF candidate training features.

Feature	Description
p_T	transverse momentum
η	pseudorapidity
phi	polar angle
$p_{T\,rel}$	absolute difference in transverse momentum to lepton candidate
ΔR	angular difference to lepton candidate
$\text{PUPPI}_{\text{weight}}$	pileup per particle identification weight
$\text{PUPPI}_{\text{weight no},\ell}$	pileup per particle identification weight without the lepton

Table 5.4: Neutral hadron and photon PF candidate training features.

Feature	Description
p_T	transverse momentum
η	pseudorapidity
phi	polar angle
m	mass
$p_{T\,rel}$	absolute difference in transverse momentum to lepton candidate
ΔR	angular difference to lepton candidate
x	secondary vertex X position, in cm
y	secondary vertex Y position, in cm
z	secondary vertex Z position, in cm
D_{len}	decay length in cm
$s_{D\,len}$	decay length significance
D_{xy}	2D decay length in cm
$s_{D\,xy}$	2D decay length significance
P_{angle}	pointing angle, i.e. $\arccos(P_{\text{SV}} * (P_{\text{SV}} - P_{\text{PV}}))$
χ^2	reduced χ^2, i.e. χ/N_{dof}
N_{dof}	number of degrees of freedom

Table 5.5: SV PF candidate training features.

Year	Particle	Training	Sample Table	Signal/Background
2016	electron	DYvsQCD	Tab. A.1	Signal
2016	electron	DYvsQCD	Tab. A.3	Background
2016	electron	$t\bar{t}$	Tab. A.4	Signal+Background
2016	muon	DYvsQCD	Tab. A.1	Signal
2016	muon	DYvsQCD	Tab. A.2	Background
2016	muon	$t\bar{t}$	Tab. A.4	Signal+Background
2017	electron	DYvsQCD	Tab. A.5	Signal
2017	electron	DYvsQCD	Tab. A.7	Background
2017	electron	$t\bar{t}$	Tab. A.8	Signal+Background
2017	muon	DYvsQCD	Tab. A.5	Signal
2017	muon	DYvsQCD	Tab. A.6	Background
2017	muon	$t\bar{t}$	Tab. A.8	Signal+Background
2018	electron	DYvsQCD	Tab. A.9	Signal
2018	electron	DYvsQCD	Tab. A.11	Background
2018	electron	$t\bar{t}$	Tab. A.12	Signal+Background
2018	muon	DYvsQCD	Tab. A.9	Signal
2018	muon	DYvsQCD	Tab. A.10	Background
2018	muon	$t\bar{t}$	Tab. A.12	Signal+Background

Table 5.6: Training sample summary.

for each year, training type and lepton flavour are trained separately. In total, 12 networks were trained.

5.4 Data Preprocessing

5.4.1 Data Processing with CRAB

The CMS Remote Analysis Builder (CRAB)[127] is an utility to process data and MC samples of the CMS experiment. It enables the use of computational resources that are linked to the CMS experiment and are located around the world. With this tool the samples for the preprocessing are generated.

5.4.2 Step 1 - Selection

Step 1 is the internal name for the first preprocessing step, after the samples have been generated with CRAB. The data that is input into this step is event-based, which means that all particles of the event are saved. Therefore step 1 consists of two main loops. The first one iterates through all events and finds the leptons. The second loop then writes out all particles in the vicinity, the PF candidates, and secondary vertices (SV) in a ΔR-cone of 0.5 around the lepton and sorts them into prompt, non-prompt and fake data files.

5.4.3 Step 2 - Mix and Sort

In classification tasks with Deep Learning it is important that the training data is relatively balanced in the amount of entries per class. For slightly unbalanced data re-weighting the loss function can be a solution. In this thesis we are interested in the network performance of prompt versus non-prompt and fake leptons. Therefore the number of prompt leptons is chosen to be equal to the number of non-prompt and fake leptons. This is achieved by randomly selecting a lepton of one of the three classes until one class runs out of leptons, so that the number matched $N_{\text{prompt}} = N_{\text{non-prompt}} + N_{\text{fake}}$.

variable	bins
p_T	3.5, 4.35, 5.42, 6.74, 8.4, 10.45, 13.02, 16.20, 20.17, 25.10, 31.24 38.89, 48.40, 60.24, 74.99, 93.34, 116.18, 144.61, 180.01, 224.05, 278.88 347.13, 432.07, 537.8, 669.40, 833.22, 1037.11, 1290.91, 1606.80, 2000
η	-2.5, -2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2, 2.5

Table 5.7: Bins for normalisation.

In addition to this, the PF candidates are sorted by their relative transverse momentum to the lepton and the secondary vertices are sorted by the angular distance to the particle. This sorting is important, as the recurrent layers of the deep neural network need their inputs ordered to perform their task of extracting information.

Conversion

As Keras, the deep learning framework used in this thesis, needs numpy [128] arrays as training input, the ROOT [129] files, that have been used until this point, need to be converted. This is done with the DeepJetCore [4] repository. In addition to converting the data from ROOT to numpy, it also zero-pads the data of the particles around the lepton.

DeepJetCore was developed for the identification of hadron jets and the datastructure of jets with their individual particles and leptons with their vicinity particles is very similar. Both consist of global variables and sorted lists of different flavour particles, therefore, a lot of the pre-existing code can be reused.

Because the distributions of the lepton variables p_T and η vary between the three lepton classes, they are normalised to a reference class. The reference class was chosen to be the prompt class. To get similar distributions for each class, the leptons were put into p_T and η bins (see Tab. 5.7). Then leptons are discarded from the other two classes, so that the ratio of the number of particles in its bins matches the one of the reference class. This is done, so that the network does not learn to discriminate based on these variables.

5.5 Network Layout and Training settings

The DeepLepton classifier makes use of a large NN with LSTMs and dense layers. The NN architecture is given in Fig. 5.11. The integers in the blue boxes stand for a dense layer with that amount of nodes. The numbers in front of LSTM indicate the number of cells in the LSTM layers. They were used in conjunction with the bidirectional wrapper. Because it usually is good practice to normalize the input data, as this often mitigates training problems and usually leads to higher training accuracies, a batch normalisation layer has also been placed as a first layer to normalise the input data. For regularisation dropout layers with a dropout rate of 50% were placed after each dense layer. To accelerate the training a batch normalisation layer is also placed after each dropout layer. The ELU function was used as an activation function for the dense layers. For the training the cross-entropy loss was used and the ADAM [98] optimiser was used to accelerate the training further.

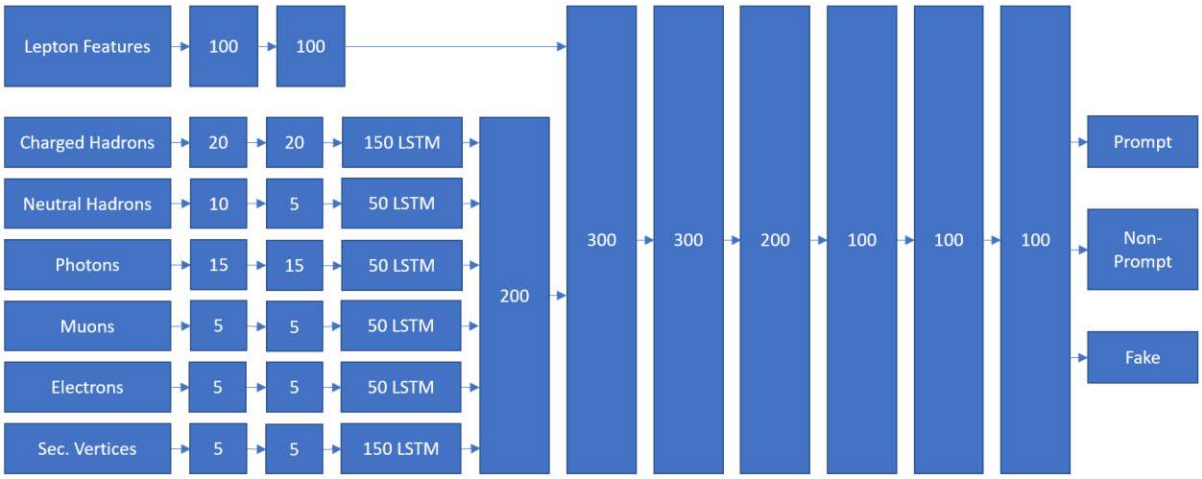


Figure 5.11: The standard DeepLepton neural network layout with LSTM as recurrent cells. The numbers in the boxes indicate the number of nodes per dense layer, and the numbers in front of LSTM indicate the number of parallel cells. Additionally batch-normalisation and dropout layers were used after each dense and recurrent layer of the figure to improve convergence.

Chapter 6

Results and Discussion

While no preselection was applied to the leptons used for the training of the networks, one is applied to all leptons used for evaluation of the DeepLepton classifier, which is defined in Tab. 6.1. One area of improvement of the DeepLepton classifier is over lepton isolation cuts.

6.1 Network performance

6.1.1 Discriminator shapes

Discriminator shapes show the distribution of the network's predicted output for a certain output class with respect to the true input class. The lepton classes are provided by a MC simulation, as per the requirement of supervised learning. Therefore, the discriminator shapes can be plotted for each of the three true classes. An example for the 2016 muon DYvsQCD training is shown in Fig. 6.1. These discriminator histograms show that there is a good separation between the classes. The discriminator histogram plots for all trainings can be found in the appendix Appendix C.

6.1.2 ROC and Efficiency

To give a measure of quality of a classifier, receiver operating characteristics (ROC) [130] curves are used. They allow for a clear comparison between classifiers. The ROC gives the true positive rate as a function of the false positive rate of the classifier. They are defined as:

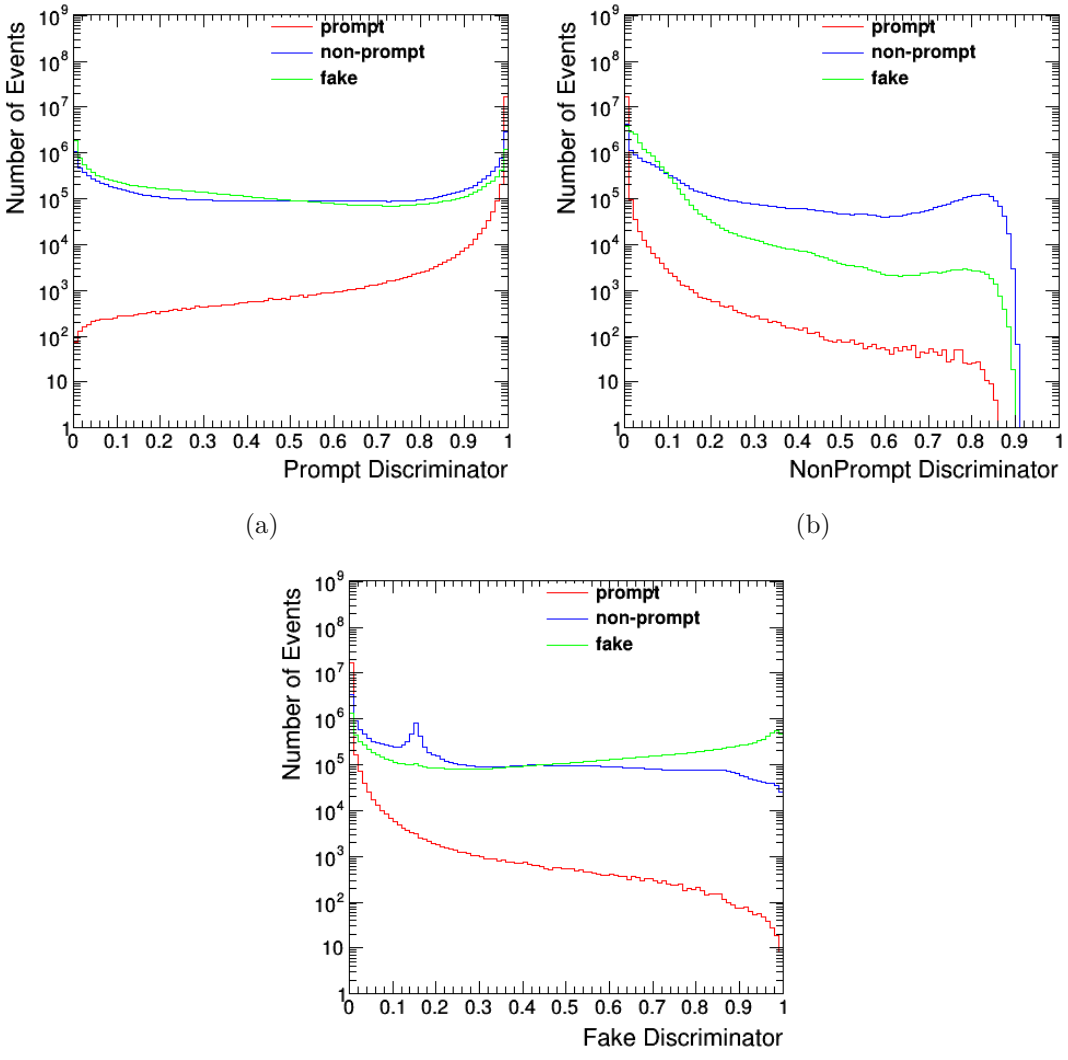
$$TPR = \frac{N_{\text{true positive}}}{N_{\text{true positive}} + N_{\text{false negative}}} \tag{6.1}$$

$$FPR = \frac{N_{\text{false positive}}}{N_{\text{false positive}} N_{\text{true negative}}} \tag{6.2}$$

An example, that compares the ROCs of the DeepLepton classifier with the TTH lepton MVA and the stops compressed cut based lepton ID's isolation cut is given in Fig. 6.2.

electron	**muon**
$d_{xy} < 0.02$	$d_{xy} < 0.02$
$d_z < 0.1$	$d_z < 0.1$
$\eta < 2.5$	$\eta < 2.4$
	candidate is particle flow muon
	candidate is global muon

Table 6.1: Prompt electron and muon requirements used in the preselection for the evaluation of the DeepLepton classifier.



(a) (b)

(c)

Figure 6.1: Discriminator value histograms of the 2016 muon DYvsQCD training from a prediction on the `DYJetsToLL_M50_LO`, `DYJetsToLL_M50_LO_ext2` samples and samples in Tab. A.2. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.

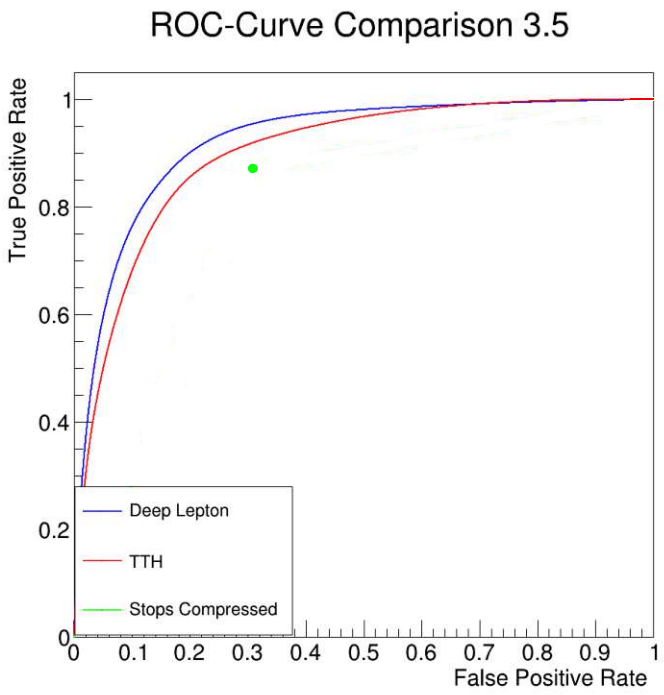


Figure 6.2: ROC of the 2016 muon DYvsQCD training from a prediction on the `DYJetsToLL_M50_LO`, `DYJetsToLL_M50_LO_ext2` samples and samples in Tab. A.2 for the p_T-bin 3.5-7.5 GeV.

A ROC curve also allows for the selection of the working point. This means, that it is possible to choose a desired false positive rate and receive the corresponding true positive rate. The true positive rate is then referred to as signal efficiency, the amount of signal leptons that are correctly classified as signal. Whereas the false positive rate is referred to as background efficiency, the amount of background leptons that are falsely classified as signal. Ideally, the signal efficiency is close to one, whereas the background efficiency stays close to zero.

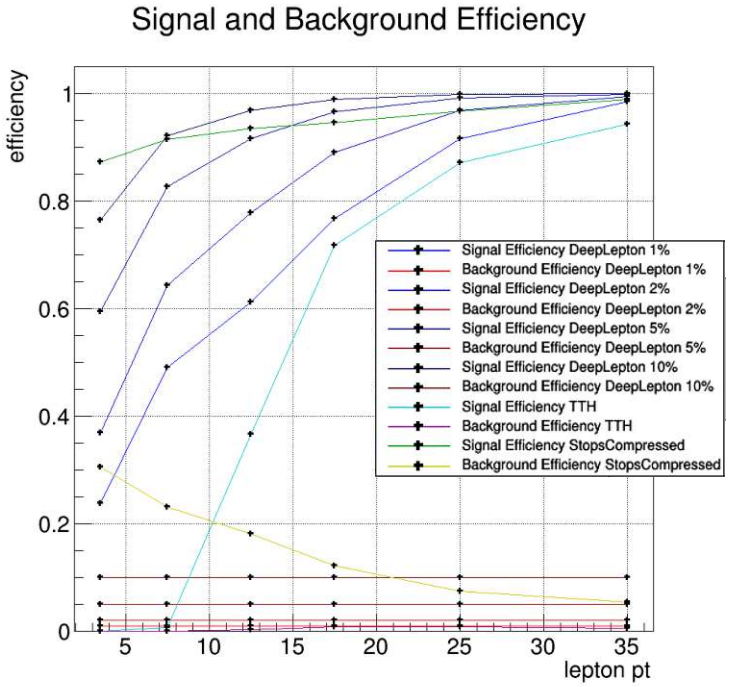
The corresponding signal and background efficiencies are calculated for each individual p_T bin. As an example the resulting efficiency curves for the 2016 muon DYvsQCD training are shown in Figure 6.3 for 1%, 2% 5% and 10% targeted background efficiency, once for the low p_T range from 3.5 to 30 GeV and once for the high p_T range up to 300 GeV.

6.1.3 Permutation Variable Importance

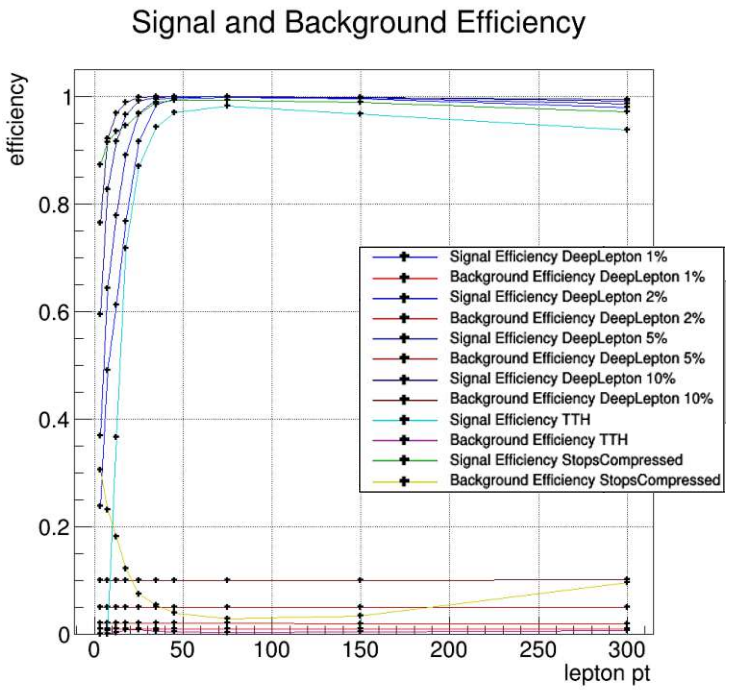
The permutation variable importance [131] is a conceptually simple way to compare the relative importances of input features. In principle this works by shuffling the input data column of one variable at a time to make sure that the shuffled column has the same distribution as before, but is decorrelated from the rest of the data. Then the classifiers performance is re-evaluated and the increase in loss compared to the baseline is calculated. This increase is interpreted as a measure for the importance of this variable. The differences are then linearly scaled to a maximum of one.

The permutation variable importance was measured for the data year 2016 for all 4 trainings. Figure 6.4 and Fig. 6.5 show the permutation importance of all training variables for electrons and muons. To achieve these plots, 20000 leptons were taken from the training sample, the preprocessing step 2 output, and input into the network to calculate the loss difference per shuffled variable. With 20000 leptons the evaluation took about 24 hours per image.

For the electron trainings the variable $\mathrm{ID}_{\mathrm{no\,Iso}}$ is clearly the most important feature. This is due

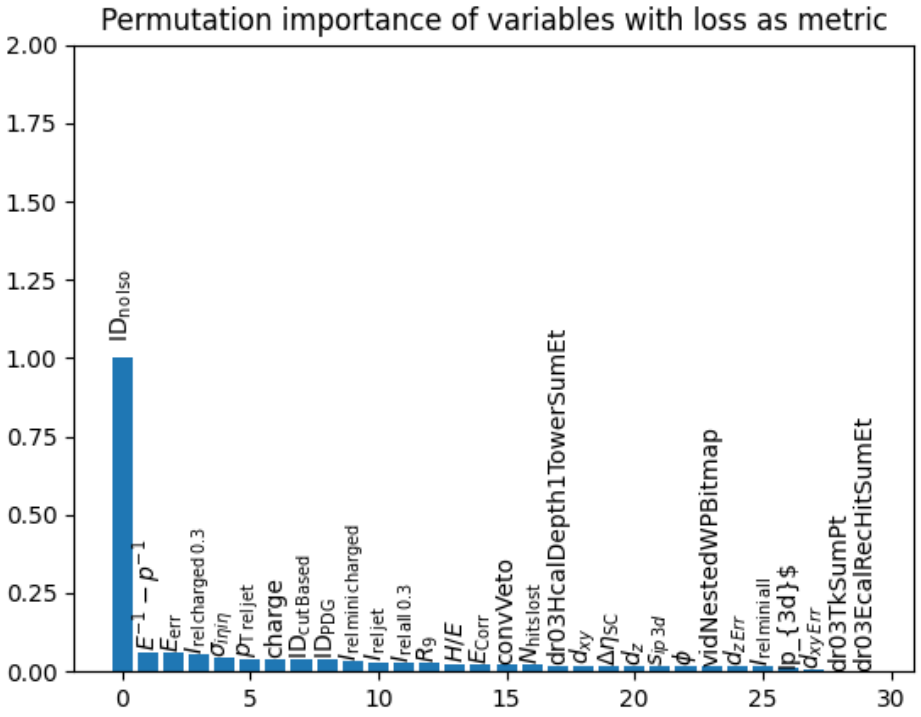


(a)

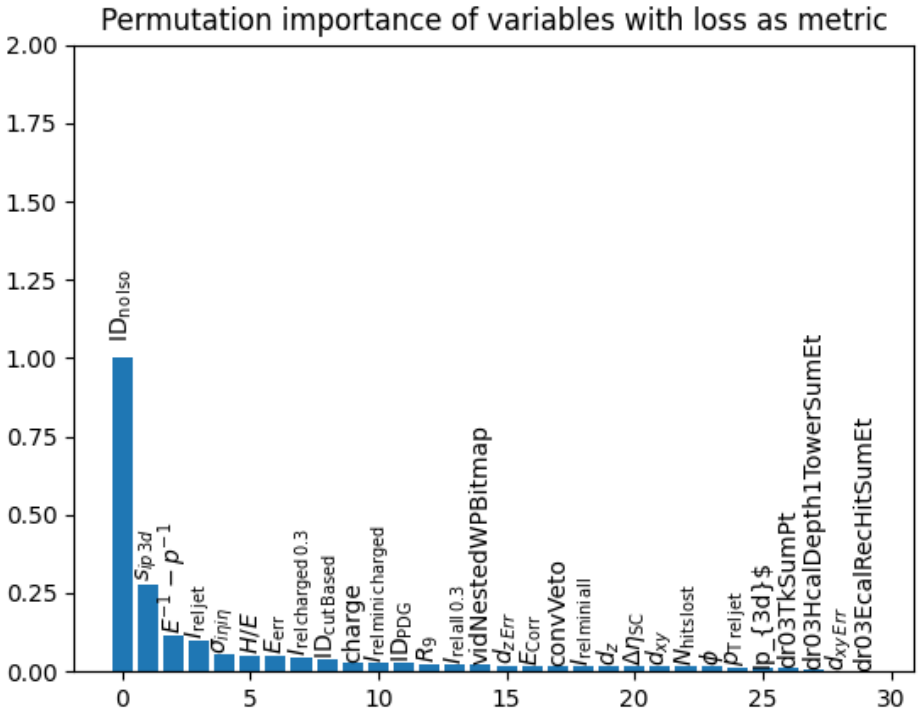


(b)

Figure 6.3: Efficiency curves of the 2016 muon DYvsQCD training from a prediction on the `DYJetsToLL_M50_LO`, `DYJetsToLL_M50_LO_ext2` samples and samples in Tab. A.2 for the p_T ranges 3.5-30 (a) and 3.5-300 (b) GeV. The background efficiencies of the DeepLepton classifier are locked to 1%, 2% 5% and 10%. The working point of the TTH MVA is set to 0.9, as per the recommendation on CERN's internal wiki page.

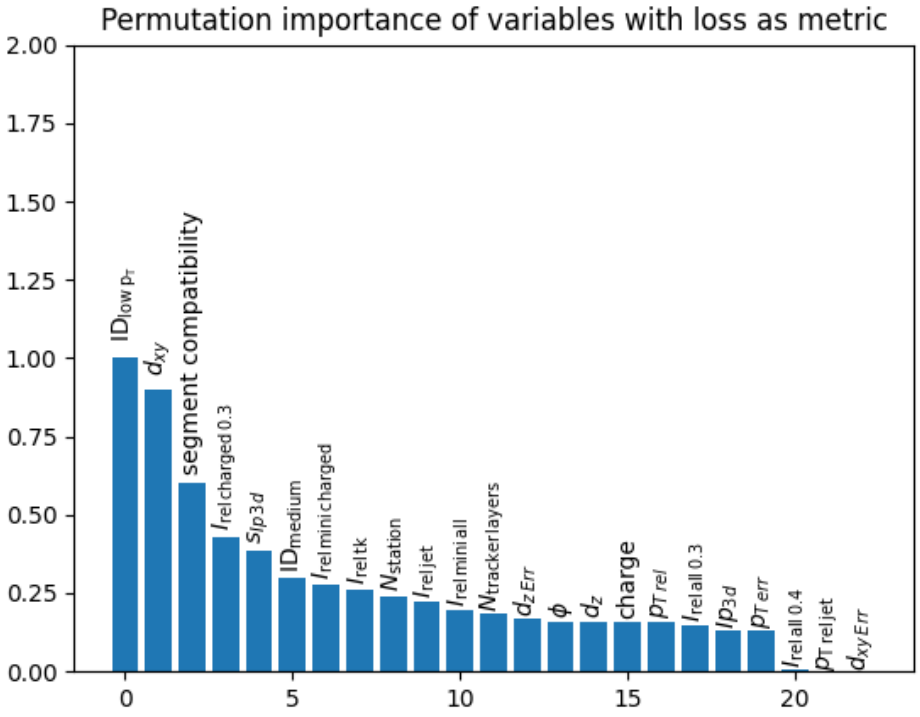


(a) DYvsQCD.

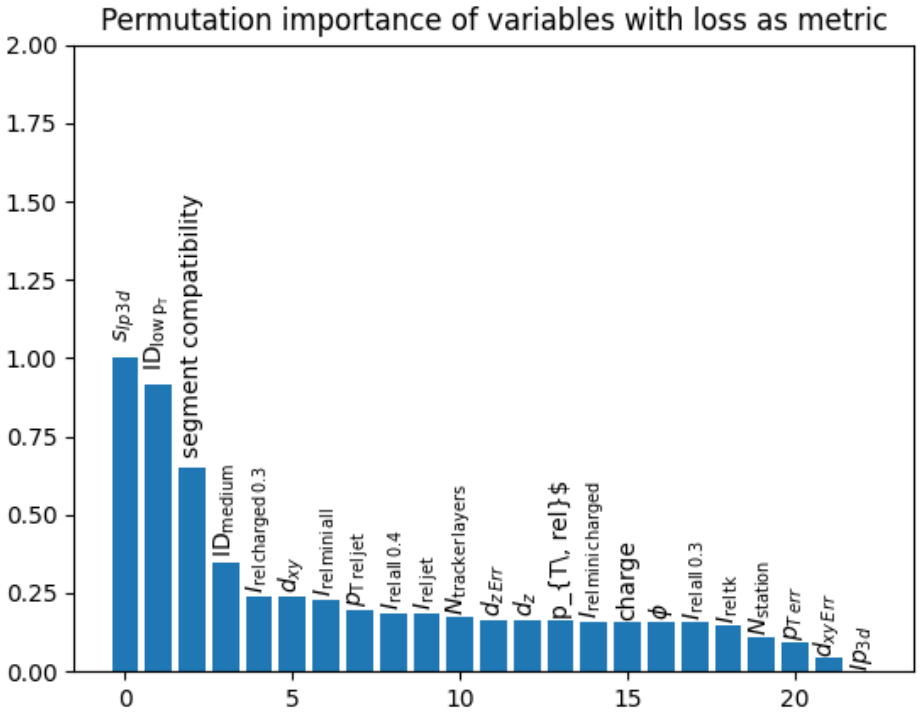


(b) Top.

Figure 6.4: Permutation importance for 2016 electron DYvsQCD and $t\bar{t}$ trainings.



(a) DYvsQCD.



(b) Top.

Figure 6.5: Permutation importance for 2016 muon DYvsQCD and $t\bar{t}$ trainings.

to it being MVA variable that was trained on the other features, meaning it already contains their information. Therefore, most of the information of the DeepLepton identifiers gain likely comes from the inclusion of the PF candidates. For the electron Top training the significance of the impact parameter to the event vertex $s_{Ip\,3d}$ is also of relative high importance. This is due to the $t\bar{t}$ samples containing much more non-prompt leptons and this variable being suitable for this separation.

For the both muon trainings in Fig. 6.5 the six most important features are the same. The $\mathrm{ID}_{\mathrm{low\,p_T}}$, a MVA trained for the identification of muons with a low p_T, the impact parameter d_{xy}, the segment compatibility of the muon system, the relative isolation in a cone of $\Delta R = 0.3$ for charged particles $I_{rel\,mini\,charged}$ and the cut-based medium working point muon ID $\mathrm{ID}_{\mathrm{medium}}$. This leads to the conclusion that a combination of these variables should be considered in new cut-based muon Ids. On the contrary, it is interesting, that although the perpendicular impact parameter d_{xy} is of relative high importance, the three dimensional impact parameter Ip_{3d} is not.

6.1.4 Comparison between LSTM, GRU, Simple RNN

The LSTMs form an integral part of the NN. Therefore, it is important to use the most suitable recurrent cell architecture. Therefore, the comparison between different recurrent cell architectures was made. For this the LSTMs of the standard NN used in this thesis Fig. 5.11 are exchanged with GRUs and simple RNN cells, otherwise the network and training settings are kept the same. In addition to this, a training without the recurrent blocks of the network gives a good estimation of the information that is gained through the inclusion of the PF candidates.

First, the average time of evaluating a single lepton was measured by predicting each lepton with each network, logging these times and averaging them. The average times over 10000 muons are shown in the second column of Tab. 6.2. This table also shows the space that the trained model takes up in RAM rounded to 100 KiB. Due to limitations of the CMS software environment the leptons were predicted on a CPU.

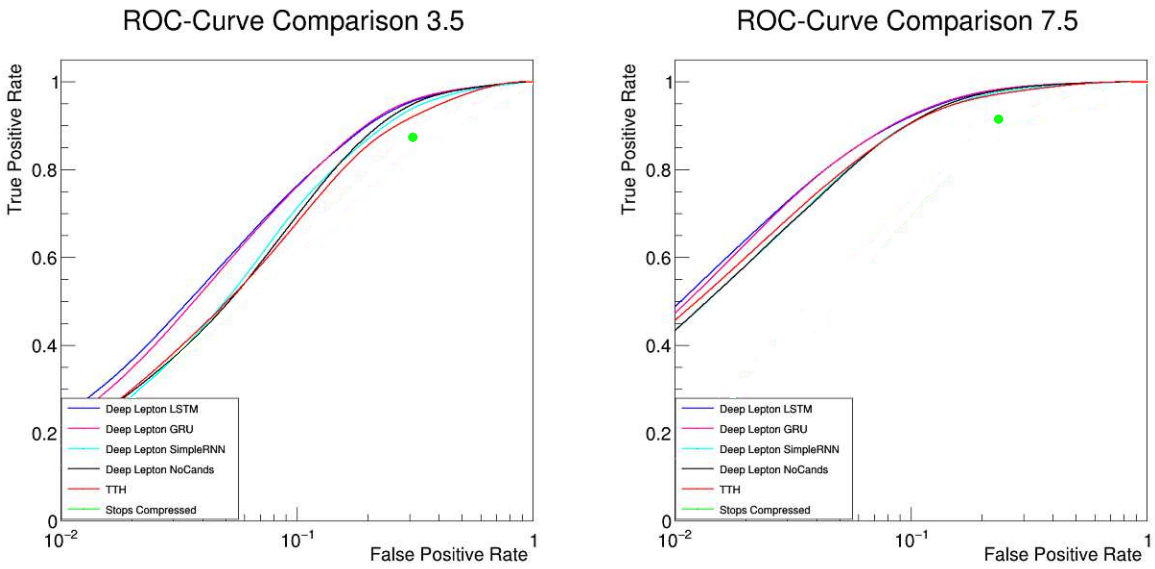
Celltype	prediction time [s]	memory usage [KiB]
LSTM	0.07787	164500
GRU	0.07668	136000
Simple RNN	0.07284	49900
No Cands	0.06643	4500

Table 6.2

The time to predict does not change a lot for each type of recurrent cell, but is unsurprisingly lower for the network without the recurrent block. In terms of memory consumption, the network with LSTM cells is the largest followed by the network with GRU cells.

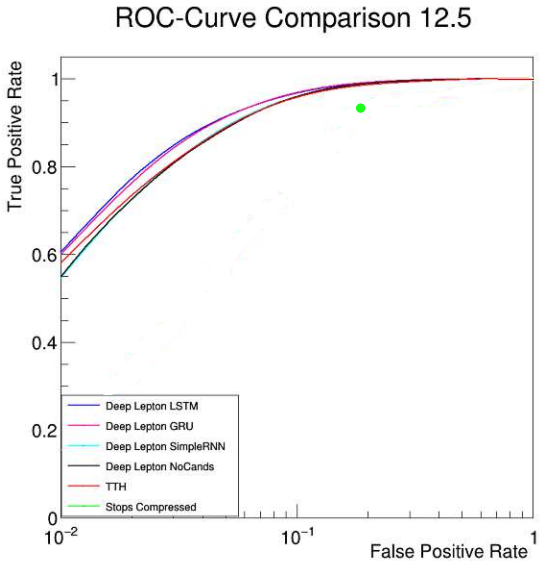
ROC comparisons for the four trained networks for p_T bins from 3.5 to 25 GeV are shown in Fig. 6.6, additionally this also shows the comparison with the ROC of the TTH MVA and the single cut-based working point of the stops compressed search. These figures show that the network with LSTMs is marginally better than the network with GRUs and the network with simple RNN cells does not perform better than the network that does not use the PF candidates, both show a very similar performance to the TTH MVA. This leads to the conclusion that the gated architectures of the LSTM and GRU are essential for extracting the extra information out of the PF candidates.

A comparison between LSTM and GRU can also be found in [132, 133]. Both come to the conclusion, that given a large enough training dataset, the LSTM performs better at the cost of a longer runtime. Due to the recurrent cells being an important but computationally small part

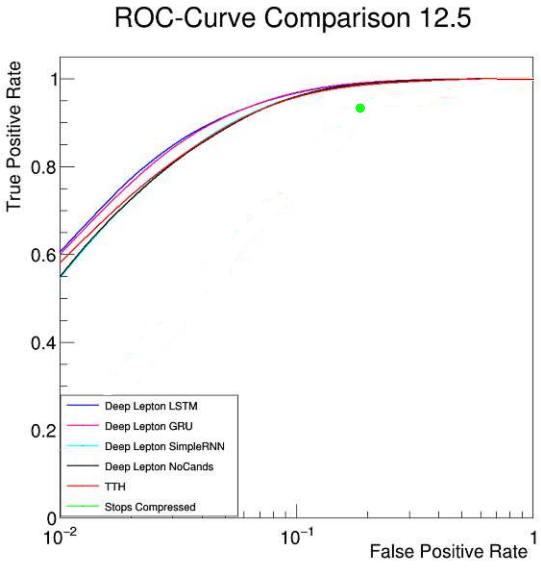


(a) p_T bin 3.5 – 7.5

(b) p_T bin 7.5 – 12.5



(c) p_T bin 12.5 – 17.5



(d) p_T bin 17.5 – 25

Figure 6.6: ROC comparison between networks with LSTM, GRU simple RNN cells and no recurrent block, the TTH MVA and the single stops compressed working point.

p_T bin	TPR	stops FPR	DeepLepton FPR	FPR factor
		muon		
3.5-5	0.848	0.212	0.135	0.637
5-12	0.868	0.126	0.065	0.516
12-20	0.873	0.071	0.032	0.451
20-30	0.884	0.048	0.018	0.375
30-50	0.926	0.053	0.012	0.226
50-∞	0.965	0.079	0.010	0.127
		electron		
5-12	0.911	0.192	0.099	0.516
12-20	0.913	0.143	0.064	0.447
20-30	0.917	0.134	0.049	0.366
30-50	0.945	0.217	0.079	0.364
50-∞	0.972	0.460	0.169	0.367

Table 6.3: Ratio of the DeepLepton and stops compressed FRPs for the fixed stops compressed TPR for each p_T bin, evaluated on a top quark pair production sample from the year 2016 (`TT_pow`).

p_T bin	FPR improvement factor
3.5-5	0.637
5-12	0.516
12-20	0.451
20-30	0.375
30-50	0.364
50-∞	0.367

Table 6.4: Conservative summary of the improvements of the DeepLepton classifier over the stops compressed isolation cut for each p_T bin.

of the network, the slight performance increase of the LSTMs is worth the marginally longer runtime of the network.

6.2 Improvements of the Stops Compressed Search

In order to gauge the improvements the usage of the DeepLepton classifier would bring to a real search on the ongoing version of the stops compressed search [120] a background estimation is performed. To achieve this task, the relative improvements on the background efficiency, the FPR, for electrons and muons for each p_T bin needs to be measured. These are shown in Tab. 6.3. The improvement in FPR lies in the range of 0.637 to 0.127.

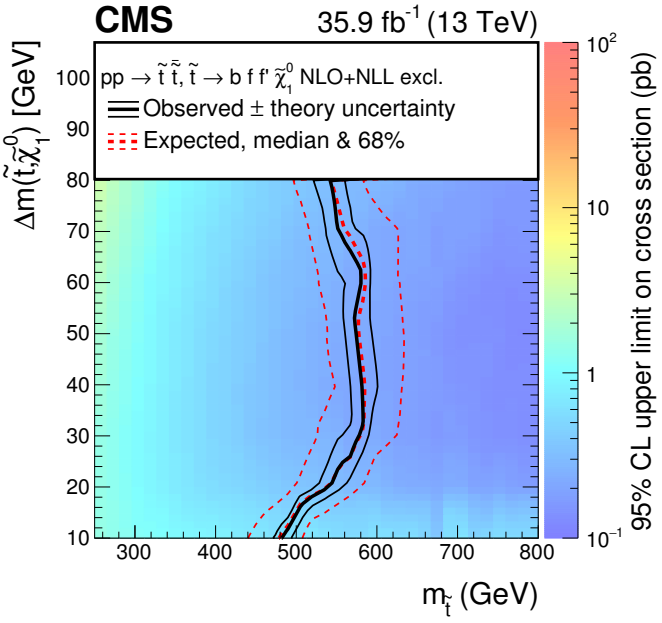
The algorithm used in the stops compressed search to measure the exclusion limits does not consider the background efficiencies of electrons and muons separately. In order to achieve a conservative estimate of the improvements, the worse background factor was chosen for each p_T bin. This summary is shown in Tab. 6.4.

By scaling the number of non-prompt background events with these numbers the effect of the DeepLepton classifier can be estimated. As a comparison, Fig. 6.7 shows the exclusion limit for the chargino mediated decay of the stop squark as a function of the stop mass $m_{\tilde{t}}$ and the stop-neutralino mass difference Δm for the stops compressed search at a 95% confidence level for the default stops compressed lepton selection with the isolation cut (a) and the DeepLepton

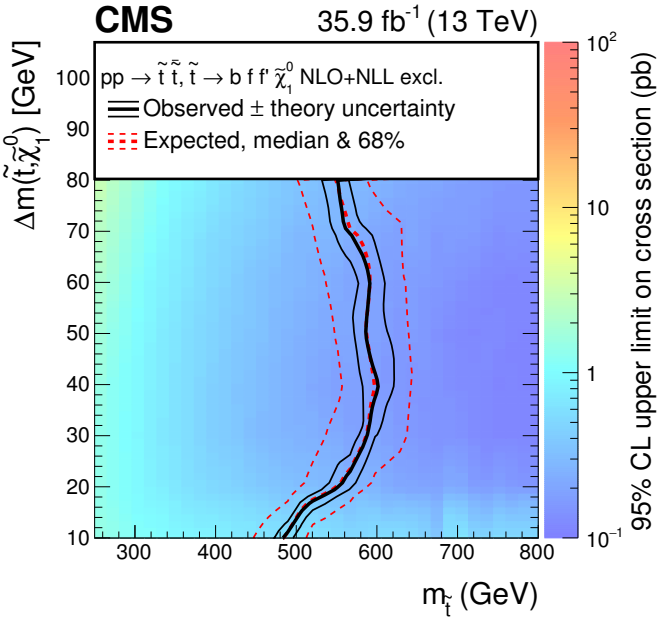
classifier (b). The observed/expected limits are represented by the solid black/dashed red lines. These were derived from the expected stop squark pair production cross section whereas the thin lines represent the variations due to theoretical/experimental uncertainties and the thick lines represent the central values.

In short, all possible combinations of $m_{\tilde{t}}$ and Δm left of the thick lines can be excluded.

By comparing the limits in Fig. 6.8 a shift of the limit lines by 5 to 10 GeV to the right can be observed. This is an improvement as more combinations of $m_{\tilde{t}}$ and Δm could be excluded by the usage of the DeepLepton classifier.



(a)



(b)

Figure 6.7: The exclusion limit for the chargino mediated decay of the stop squark as a function of the stop mass $m_{\tilde{t}}$ and the stop-chargina mass difference Δm for the stops compressed search at a 95% confidence level for the default stops compressed lepton selection with the isolation cut (upper) and the DeepLepton classifier (lower) instead of the isolation cut. The observed/ expected limits are represented by the solid black/dashed red lines. These were derived from the expected stop squark pair production cross section. The thin lines represent the variations due to theoretical/experimental uncertainties and the thick lines represent the central values.

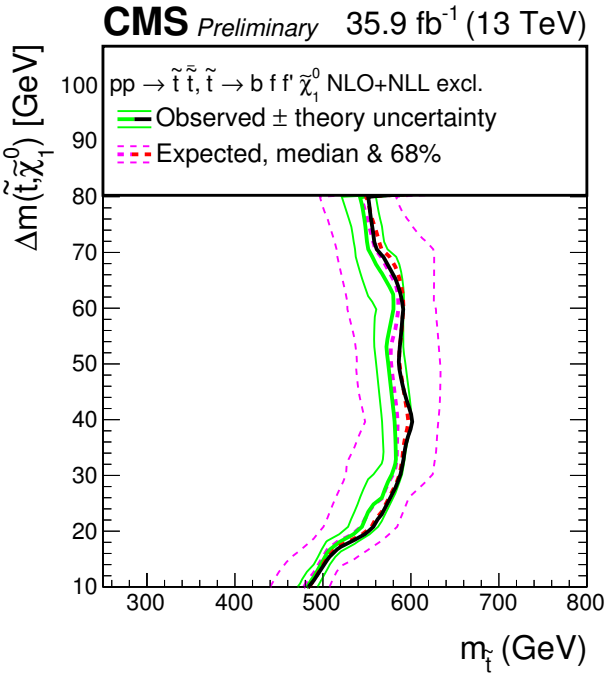


Figure 6.8: A comparison of the exclusion limits of Fig. 6.7 superimposed. The magenta and green lines represent the observed and expected limits for the isolation cut, the black and red lines represent the observed and expected limits for the limits of the DeepLepton classifier. For better visibility the variation lines for the limits of the DeepLepton classifier are not shown.

Chapter 7

Summary and Conclusion

In this thesis I trained and evaluated the novel lepton classification algorithm DeepLepton for two types of training, namely DYvsQCD and $t\bar{t}$ for the MC data years 2016, 2017 and 2018 for electrons and muons. In total, twelve networks were trained. The addition of particles in the vicinity of the lepton candidate increased the classifier's efficiency over traditional methods. This was shown in the improved exclusion limits for the stops compressed search by the DeepLepton classifier due to the improvement in the suppression of non-prompt backgrounds. Furthermore, the permutation variable importance was calculated for each training of the 2016 data year. The results show the reliance on the $\mathrm{ID}_{\mathrm{no\,Iso}}$ variable of the electron networks and the relative similarities between the DYvsQCD and $t\bar{t}$ trainings in variable importance.

To evaluate the impact of the LSTM layers and to compare them with alternative recurrent layers, networks where the LSTM layers were swapped with GRUs and Simple RNN layers were trained. Additionally, a network without the recurrent block was trained. The results showed, that the LSTMs outperform their alternatives and that their computational cost during prediction due to their complicated internal structure is minimal.

The DeepLepton classifier shows that the inclusion of more information into the algorithm increases it's performance. This leads to the conclusion that future searches should focus on the use of ML and DL techniques and to exploit all data available to train these algorithms.

In the future, analyses and searches should move away from statistically limited approaches to methods that can effectively exploit more information that are therefore only systematically limited. These ML based techniques will be able to effectively use more data and will unify across multiple objects such as leptons and jets. This will lead to an improved understanding of rare and complicated events.

Chapter 8

Acknowledgements

During the time of work on my master's thesis I gained a lot of inside into the physics at the LHC and techniques for the analysis of multiple terabytes of data. This work would not have been possible without certain people.

Firstly I want to thank Robert Schöfbeck for his support and guidance through this thesis. Due to the Corona restrictions and multiple lockdowns it was unfortunately not possible to hold meetings regularly in person. Therefore, I thank him for his efforts to regularly conduct online meetings and quick responses to the many questions I had.

Furthermore I want to thank Dietrich Liko for his help and knowledge on setting up the various software environments, his effort in helping me with various technical problems and for hosting some DeepLepton meetings on his beautiful terrace. I also want to thank Priya and Janik for taking time to help me generate the limit plots.

Finally, I want to thank my parents, my sister and my girlfriend Carolina for their support and giving strength during while working on my thesis and during the pandemic.

Appendix A

Samples

Short Name	Sample name
DYJetsToLL_M50_LO	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_LO_ext2	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_ext2	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M10to50_LO	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M10to50	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M5to50_HT70to100	DYJetsToLL_M-5to50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT100to200	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT100to200_ext	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT200to400	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT200to400_ext	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT400to600	DYJetsToLL_M-5to50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT400to600_ext	DYJetsToLL_M-5to50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M5to50_HT600toInf	DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT70to100	DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT100to200_ext	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT200to400	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600_ext	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT600to800	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT800to1200	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT1200to2500	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT2500toInf	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8

Table A.1: Samples for 2016 DY.

Short Name	Sample name
QCD_Mu_pt20to30	QCD_Pt-20to30_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt30to50	QCD_Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt50to80	QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt80to120	QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt80to120_ext1	QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt120to170	QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt170to300	QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt170to300_ext1	QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt300to470	QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt300to470_ext1	QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt470to600	QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt470to600_ext1	QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt600to800	QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt600to800_ext1	QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt800to1000	QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt800to1000_ext1	QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt1000toInf	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8
QCD_Mu_pt1000toInf_ext1	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8

Table A.2: Samples for 2016 QCD muons.

Short Name	Sample name
QCD_EMEnriched_pt20to30	QCD_Pt-20to30_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt30to50	QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt30to50_ext1	QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt50to80	QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt50to80_ext1	QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt80to120	QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt80to120_ext1	QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt120to170	QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt120to170_ext1	QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt170to300	QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_EMEnriched_pt300toInf	QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt15to20	QCD_Pt_15to20_bcToE_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt20to30	QCD_Pt_20to30_bcToE_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt30to80	QCD_Pt_30to80_bcToE_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt80to170	QCD_Pt_80to170_bcToE_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt170to250	QCD_Pt_170to250_bcToE_TuneCUETP8M1_13TeV_pythia8
QCD_bcToE_pt250toInf	QCD_Pt_250toInf_bcToE_TuneCUETP8M1_13TeV_pythia8

Table A.3: Samples for 2016 QCD electrons.

Short Name	Sample name
TTTo2L2Nu_noSC_pow	TTTo2L2Nu_noSC_TuneCUETP8M2T4_13TeV-powheg-pythia8
TTTo2L2Nu_pow_CP5	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
TTTo2L2Nu_pow	TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
TTToSemilepton_pow	TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
TTToSemilepton_pow_CP5	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8
TT_pow	TT_TuneCUETP8M2T4_13TeV-powheg-pythia8
ST_schannel_4f_NLO	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1
ST_schannel_4f_CP5	ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8
ST_tchannel_antitop_4f_pow	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
ST_tchannel_antitop_4f_pow_CP5	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tchannel_top_4f_pow	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
ST_tW_antitop_NoFullyHad_5f_pow	ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1
ST_tW_antitop_5f_pow_ext1	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
ST_tW_antitop_5f_pow	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4
ST_tW_antitop_5f_pow_CP5	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tW_top_NoFullyHad_5f_pow	ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1
ST_tW_top_NoFullyHad_5f_pow_ext	ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1
ST_tW_top_5f_pow_ext1	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
ST_tW_top_5f_pow	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4
ST_tW_top_5f_pow_CP5	ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tWll_5f_LO	ST_tWll_5f_LO_13TeV-MadGraph-pythia8
ST_tWnunu_5f_LO	ST_tWnunu_5f_LO_13TeV-MadGraph-pythia8
THQ_LO	THQ_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1
THW_LO	THW_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1
tZq_ll_NLO	tZq_ll_4f_13TeV-amcatnlo-pythia8
tZq_nunu_NLO	tZq_nunu_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1
TTTT_NLO	TTTT_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8
TTWW_NLO	TTWW_TuneCUETP8M2T4_13TeV-madgraph-pythia8
TTWZ_NLO	TTWZ_TuneCUETP8M2T4_13TeV-madgraph-pythia8
TTZZ_NLO	TTZZ_TuneCUETP8M2T4_13TeV-madgraph-pythia8
TTW_LO	ttWJets_13TeV_madgraphMLM
TTWJetsToLNu_NLO	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8
TTWJetsToQQ_NLO	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8
TTZToLLNuNu_NLO_ext2	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8
TTZToLLNuNu_NLO_ext3	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8
TTZToLL_M1to10_NLO	TTZToLL_M-1to10_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
TTZToQQ_NLO	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
ttZJets_LO	ttZJets_13TeV_madgraphMLM-pythia8
ttHToNonbb_pow	ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
ttHTobb_pow	ttHTobb_M125_13TeV_powheg_pythia8
TGG	TGGJets_leptonDecays_13TeV_MadGraph_madspin_pythia8

Table A.4: Samples for 2016 $t\bar{t}$.

Short Name	Sample name
DYJetsToLL_M10to50_LO	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT100to200_LO	DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT100to200_LO_ext1	DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT200to400_new_pmx_LO	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT200to400_LO	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT200to400_LO_ext1	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT400to600_LO	DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT400to600_LO_ext1	DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT600toInf_LO	DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT600toInf_LO_ext1	DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_NLO	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M50_LO	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_LO_ext1	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT70to100_LO	DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT100to200_LO	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT100to200_LO_ext1	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT200to400_LO	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT200to400_LO_ext1	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600_LO	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600_LO_ext1	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT600to800_LO	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT800to1200_LO	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT1200to2500_LO	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT2500toInf_LO	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8

Table A.5: Samples for 2017 DY.

Short Name	Sample name
QCD_Mu_pt15to20	QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt20to30	QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt30to50	QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt50to80	QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt80to120	QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt120to170	QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt170to300	QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt300to470	QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt470to600	QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt600to800	QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt800to1000	QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt1000toInf	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8

Table A.6: Samples for 2017 QCD muons.

Short Name	Sample name
QCD_EMEnriched_pt20to30	QCD_Pt-20to30_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt30to50	QCD_Pt-30to50_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt50to80	QCD_Pt-50to80_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt80to120	QCD_Pt-80to120_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt120to170	QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt170to300	QCD_Pt-170to300_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt300toInf	QCD_Pt-300toInf_EMEnriched_TuneCP5_13TeV_pythia8
QCD_bcToE_pt20to30_new_pmx	QCD_Pt_20to30_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt20to30	QCD_Pt_20to30_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt30to80	QCD_Pt_30to80_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt80to170	QCD_Pt_80to170_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt170to250	QCD_Pt_170to250_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt250toInf	QCD_Pt_250toInf_bcToE_TuneCP5_13TeV_pythia8

Table A.7: Samples for 2017 QCD electrons.

Short Name	Sample name
ST_schannel_4f_NLO	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8
ST_schannel_4f_NLO_PS	ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8
ST_tchannel_antitop_4f_incl_pow	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8
ST_tchannel_top_4f_incl_pow	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8
ST_tchannel_antitop_5f_pow_PS	ST_t-channel_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tchannel_antitop_5f_pow_PS_old_pmx	ST_t-channel_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tchannel_top_5f_pow_old_pmx	ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8
ST_tchannel_top_5f_pow	ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8
ST_tW_antitop_NoFullyHad_5f_pow_PS	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tW_antitop_incl_5f_pow	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8
ST_tW_antitop_incl_5f_pow_PS	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tW_top_NoFullyHad_5f_pow	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8
ST_tW_top_NoFullyHad_5f_pow_PS	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tW_top_incl_5f_pow	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8
ST_tW_top_incl_5f_pow_PS	ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
ST_tWll_5f_LO	ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
ST_tWnunu_5f_LO	ST_tWnunu_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
TTJets_NLO	TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
TTJets_LO	TTJets_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_semilepFromT_LO	TTJets_SingleLeptFromT_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_semilepFromTbar_LO	TTJets_SingleLeptFromTbar_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_dilep_LO	TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_semilepFromT_genMET150_LO	TTJets_SingleLeptFromT_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_semilepFromTbar_genMET150_LO	TTJets_SingleLeptFromTbar_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_dilep_genMET150_LO	TTJets_DiLept_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8
TTTo2L2Nu_pow	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
TTTo2L2Nu_pow_PS	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
TTToSemiLeptonic_pow	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
TTToSemiLeptonic_pow_PS	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8
TTHH_LO	TTHH_TuneCP5_13TeV-madgraph-pythia8
TTZH_LO	TTZH_TuneCP5_13TeV-madgraph-pythia8
TTWH_LO	TTWH_TuneCP5_13TeV-madgraph-pythia8
TTWW_LO	TTWW_TuneCP5_13TeV-madgraph-pythia8
TTWZ_LO	TTWZ_TuneCP5_13TeV-madgraph-pythia8
TTZZ_LO	TTZZ_TuneCP5_13TeV-madgraph-pythia8
TTJets_HT600to800_LO	TTJets_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_HT800to1200_LO	TTJets_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_HT1200to2500_LO	TTJets_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8
TTJets_HT2500toInf_LO	TTJets_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8

Table A.8: Samples for 2017 $t\bar{t}$.

Short Name	Sample name
DYJetsToLL_M50_NLO	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M50_LO	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M10to50_LO	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT70to100_LO	DYJetsToLL_M-4to50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT100to200_LO	DYJetsToLL_M-4to50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT200to400_LO	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT400to600_LO	DYJetsToLL_M-4to50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M4to50_HT600toInf_LO	DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT70to100_LO	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT100to200_LO	DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT200to400_LO	DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600_LO	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT400to600_LO_ext2	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT600to800_LO	DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT800to1200_LO	DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT1200to2500_LO	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
DYJetsToLL_M50_HT2500toInf_LO	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8

Table A.9: Samples for 2018 DY.

Short Name	Sample name
QCD_Mu_pt15to20	QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt20to30	QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt30to50	QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt50to80	QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt80to120	QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt80to120_ext1	QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt120to170	QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt120to170_ext1	QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt170to300	QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt300to470	QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt300to470_ext1	QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt470to600	QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt470to600_ext1	QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt600to800	QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt800to1000	QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD_Mu_pt1000toInf	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8

Table A.10: Samples for 2018 QCD muons.

Short Name	Sample name
QCD_EMEnriched_pt20to30	QCD_Pt-20to30_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt30to50	QCD_Pt-30to50_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt50to80	QCD_Pt-50to80_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt80to120	QCD_Pt-80to120_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt120to170	QCD_Pt-120to170_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt170to300	QCD_Pt-170to300_EMEnriched_TuneCP5_13TeV_pythia8
QCD_EMEnriched_pt300toInf	QCD_Pt-300toInf_EMEnriched_TuneCP5_13TeV_pythia8
QCD_bcToE_pt15tp20	QCD_Pt_15to20_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt20to30	QCD_Pt_20to30_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt30to80	QCD_Pt_30to80_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt80to170	QCD_Pt_80to170_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt170to250	QCD_Pt_170to250_bcToE_TuneCP5_13TeV_pythia8
QCD_bcToE_pt250toInf	QCD_Pt_250toInf_bcToE_TuneCP5_13TeV_pythia8

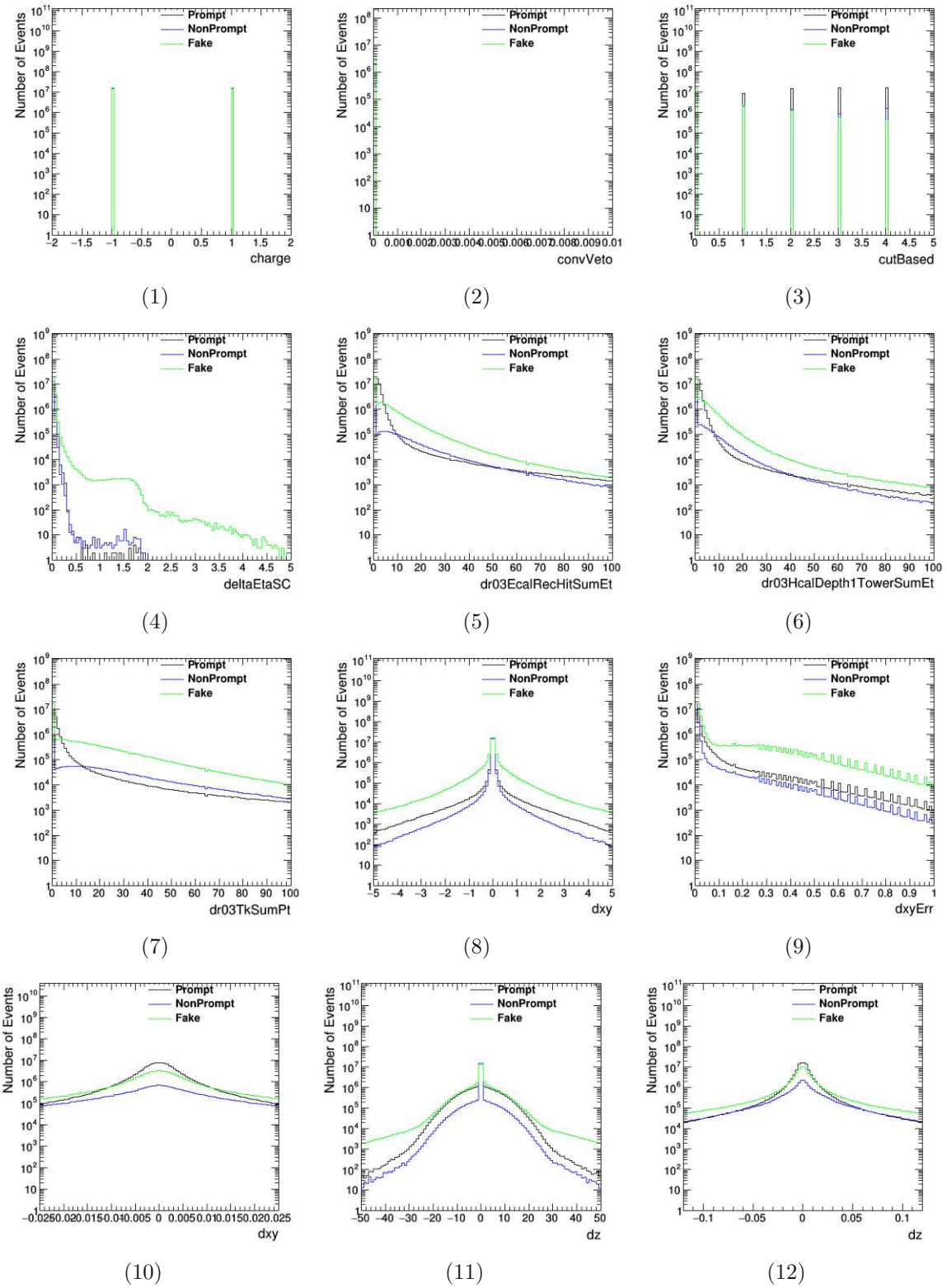
Table A.11: Samples for 2016 QCD electrons.

Short Name	Sample name
ST_schannel_LO	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8
ST_tchannel_antitop_4f_pow	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
ST_tchannel_top_4f_pow	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
ST_tW_antitop_pow	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8
ST_tW_top_pow	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8
ST_tWll_LO	ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
ST_tWnunu_LO	ST_tWnunu_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
TGJets_lep_NLO	TGJets_leptonDecays_TuneCP5_13TeV-madgraph-pythia8
TT_LO	TTJets_TuneCP5_13TeV-madgraphMLM-pythia8
TT_dilep_NLO	TT_DiLept_TuneCP5_13TeV-amcatnlo-pythia8
TTTo2L2Nu_pow	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8

Table A.12: Samples for 2018 $t\bar{t}$.

Appendix B

Input Variable Plots



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11) (12)

Figure B.1: Input feature plots before conversion 2016 electron DYvsQCD (1/4).

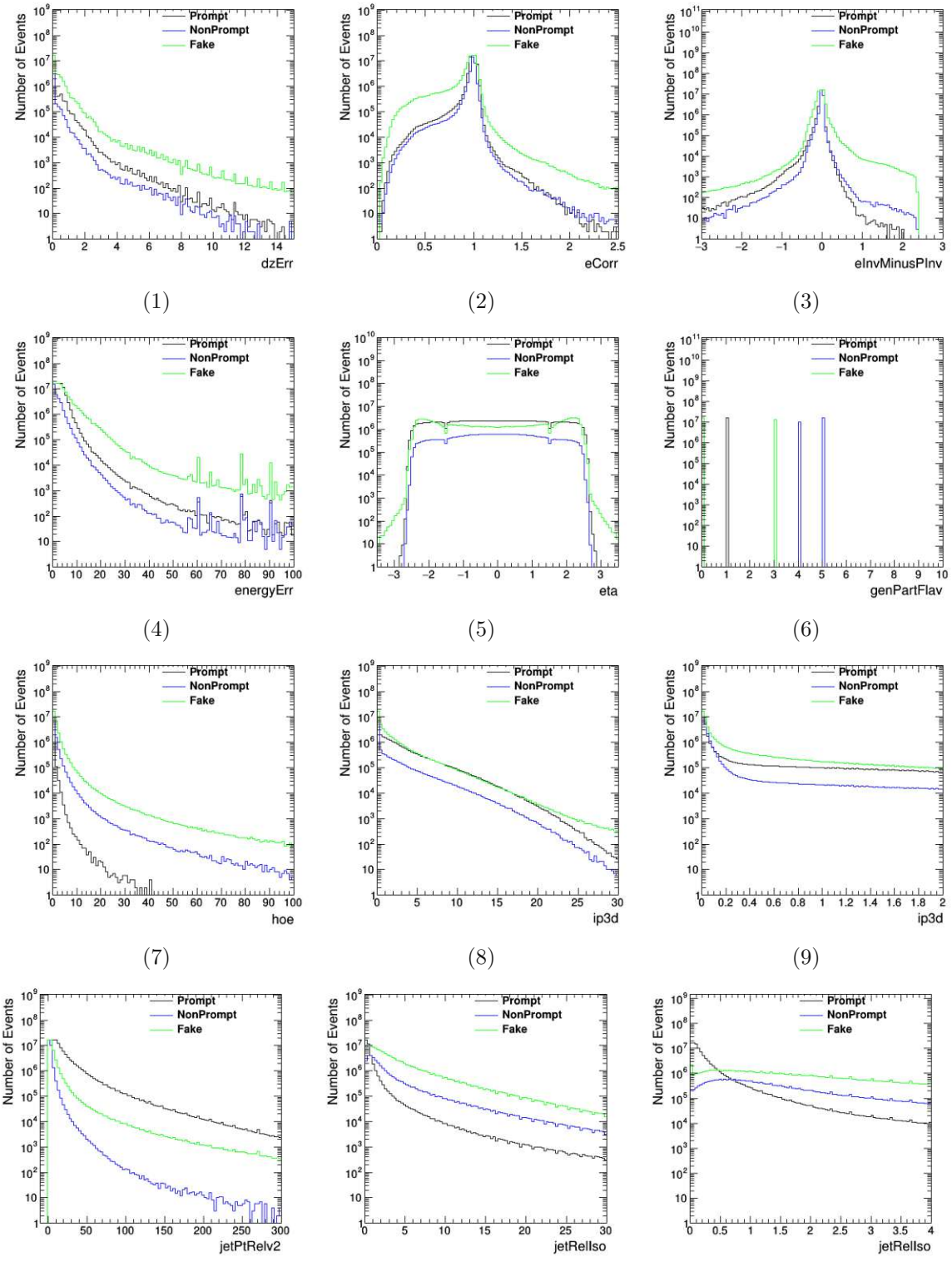


Figure B.2: Input feature plots before conversion 2016 electron DYvsQCD (2/4).

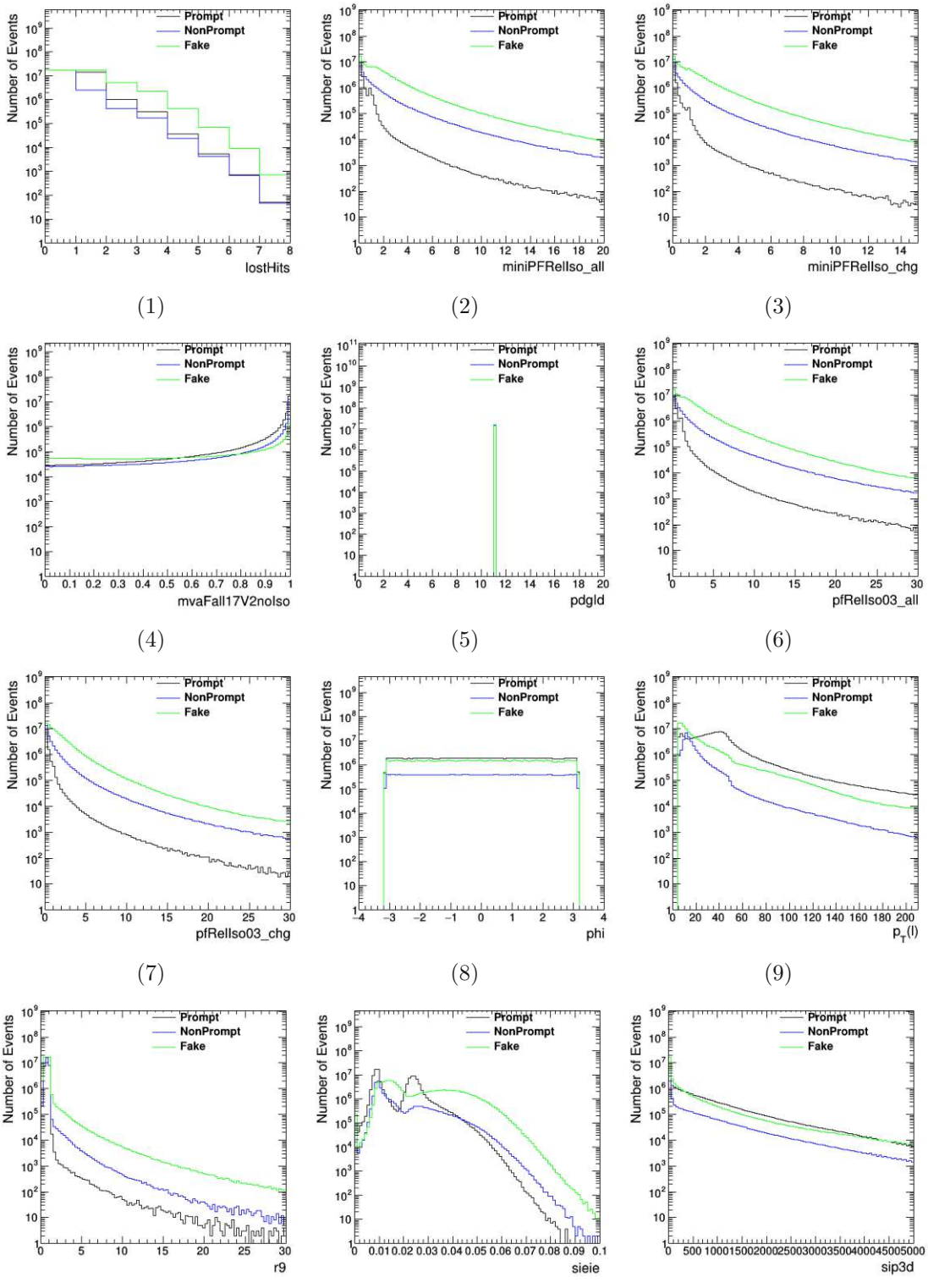


Figure B.3: Input feature plots before conversion 2016 electron DYvsQCD (3/4).

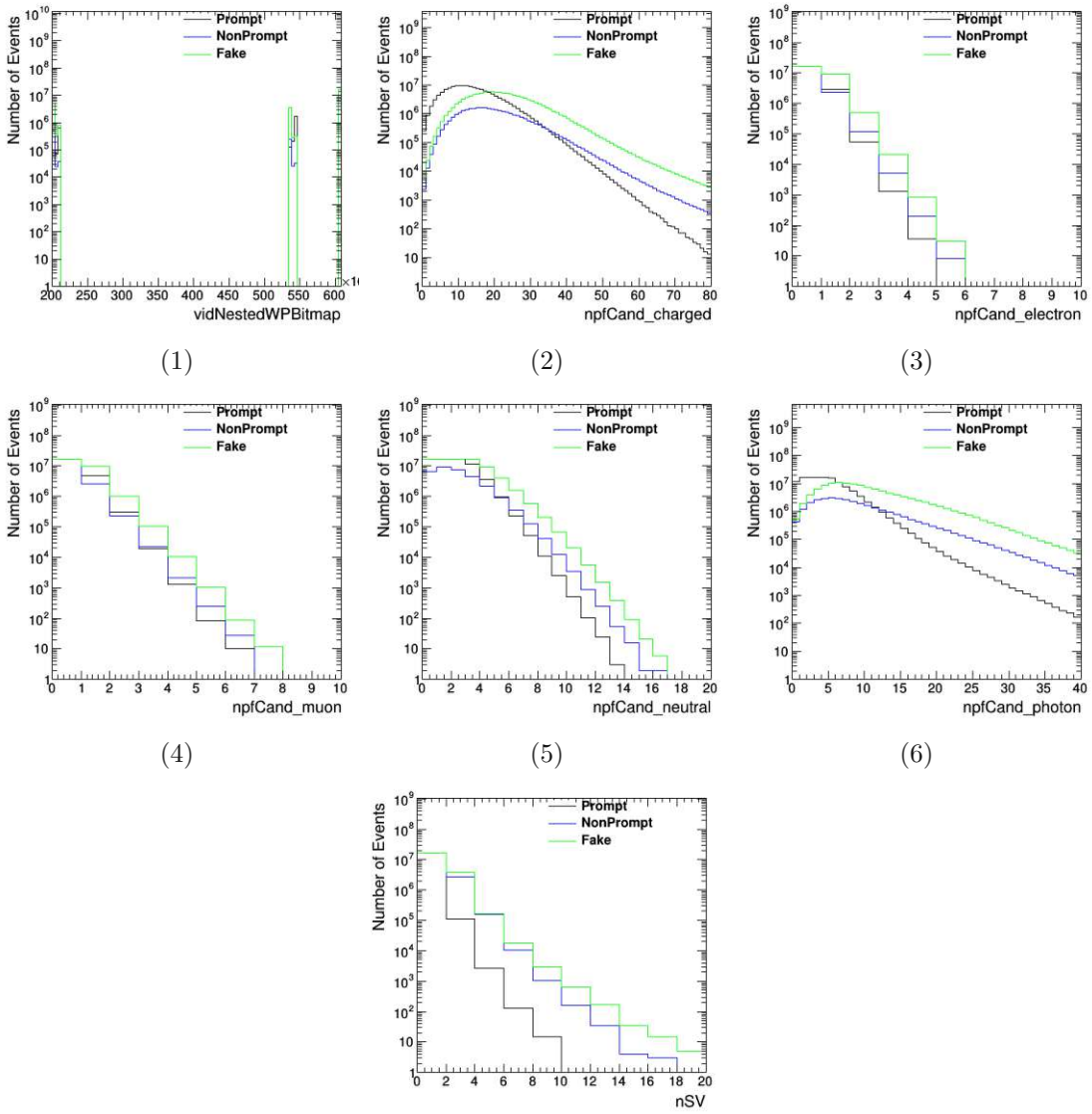


Figure B.4: Input feature plots before conversion 2016 electron DYvsQCD (4/4).

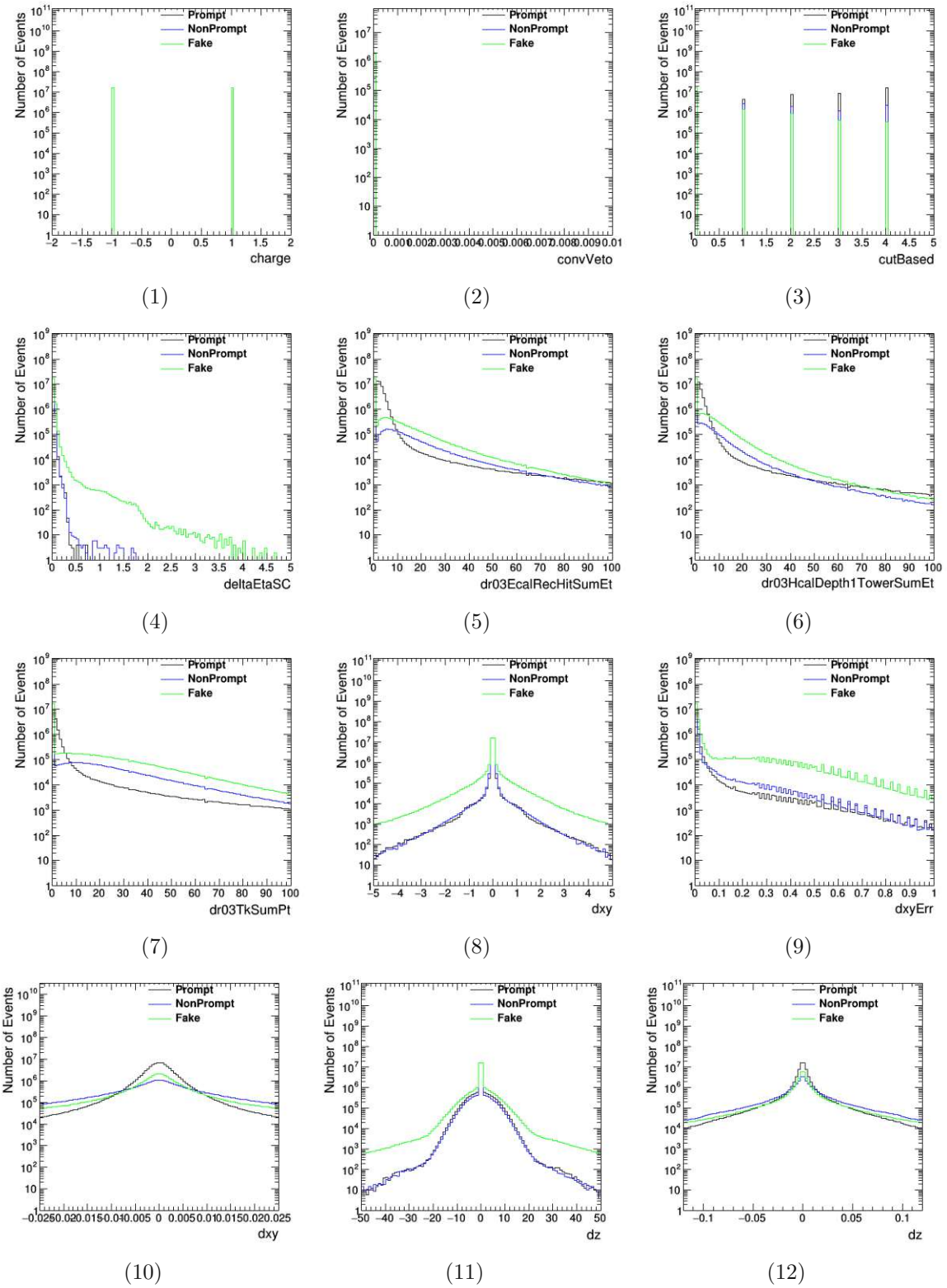


Figure B.5: Input feature plots before conversion 2017 electron DYvsQCD (1/4).

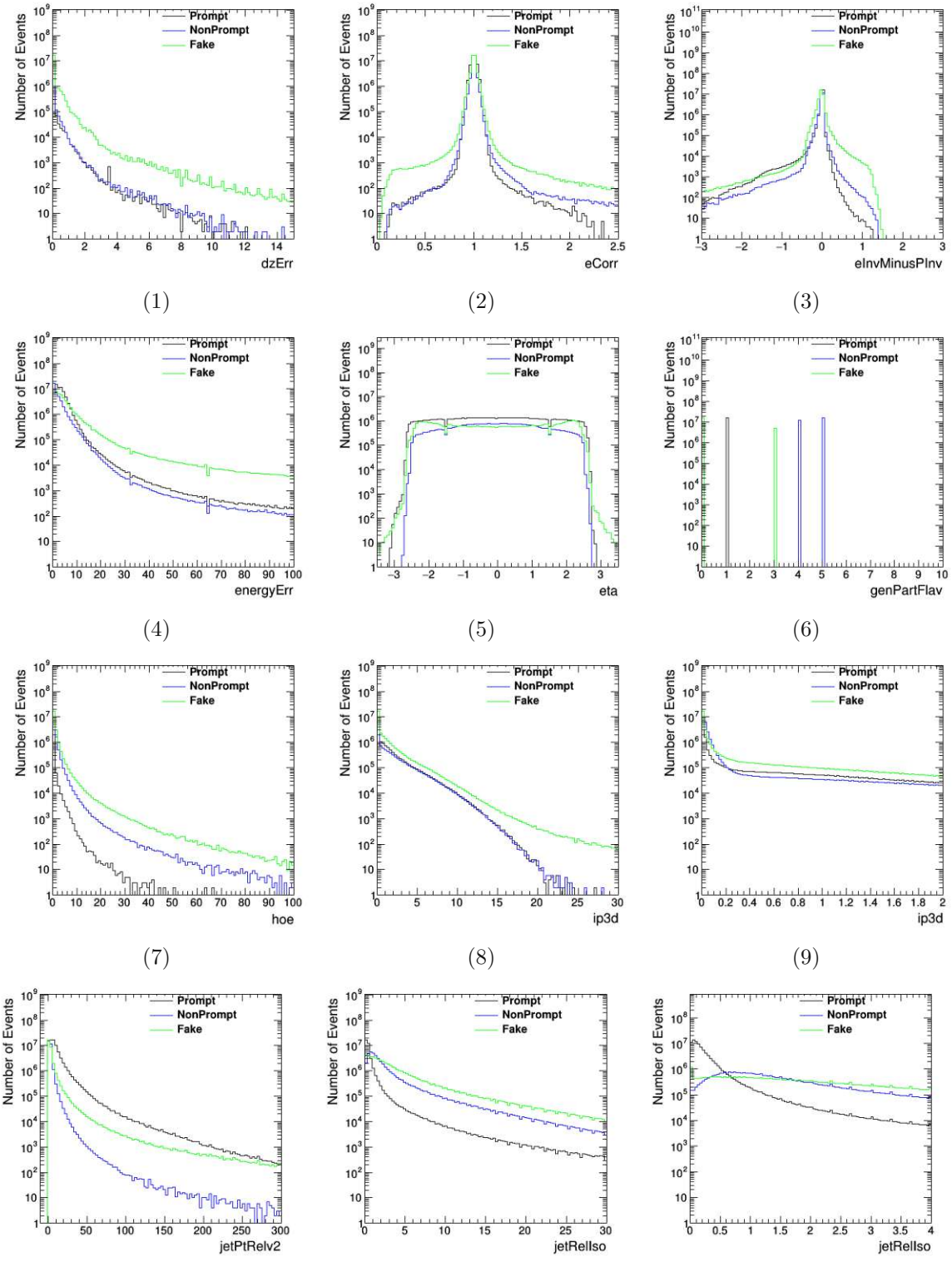


Figure B.6: Input feature plots before conversion 2017 electron DYvsQCD (2/4).

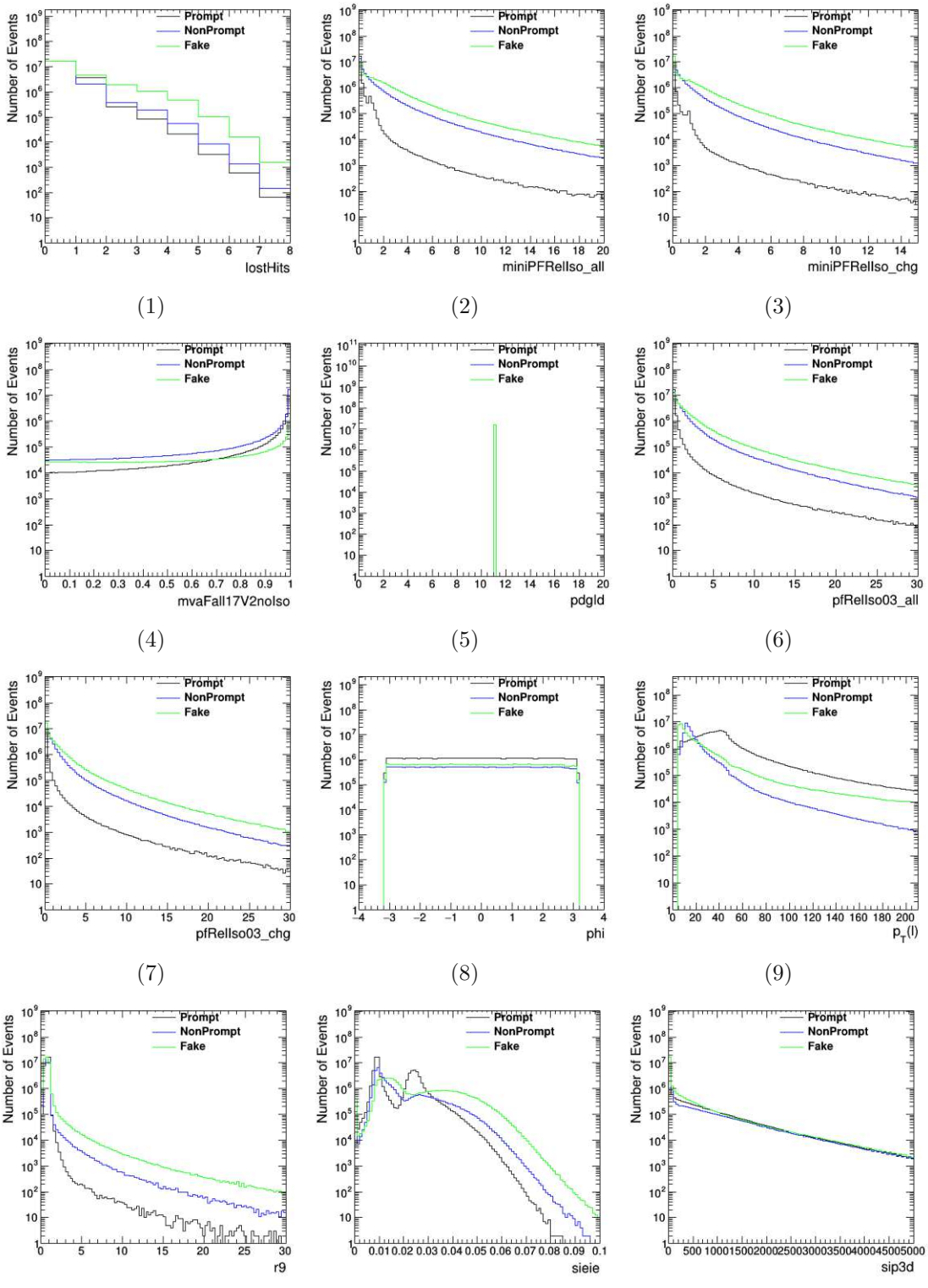


Figure B.7: Input feature plots before conversion 2017 electron DYvsQCD (3/4).

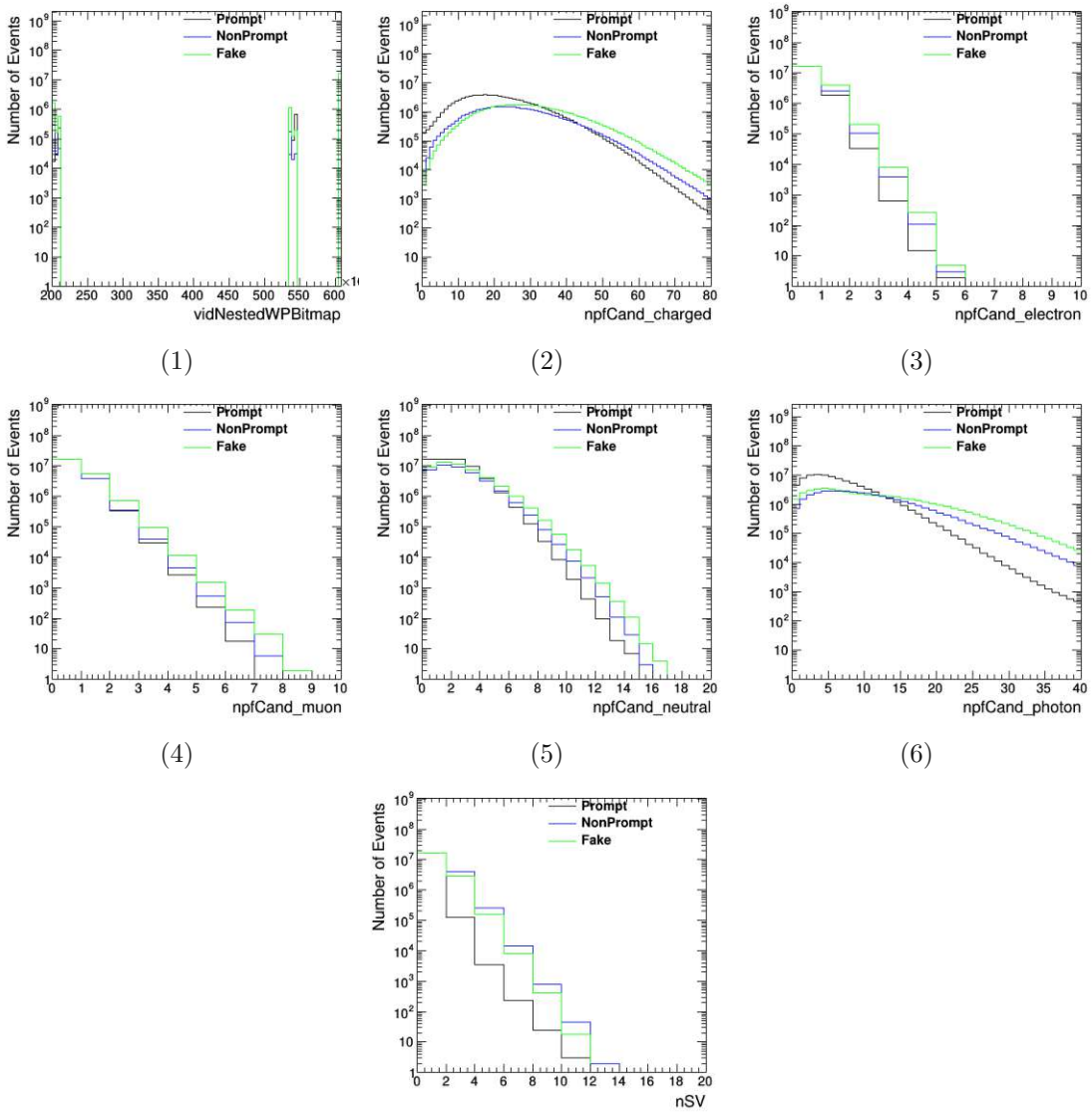


Figure B.8: Input feature plots before conversion 2017 electron DYvsQCD (4/4).

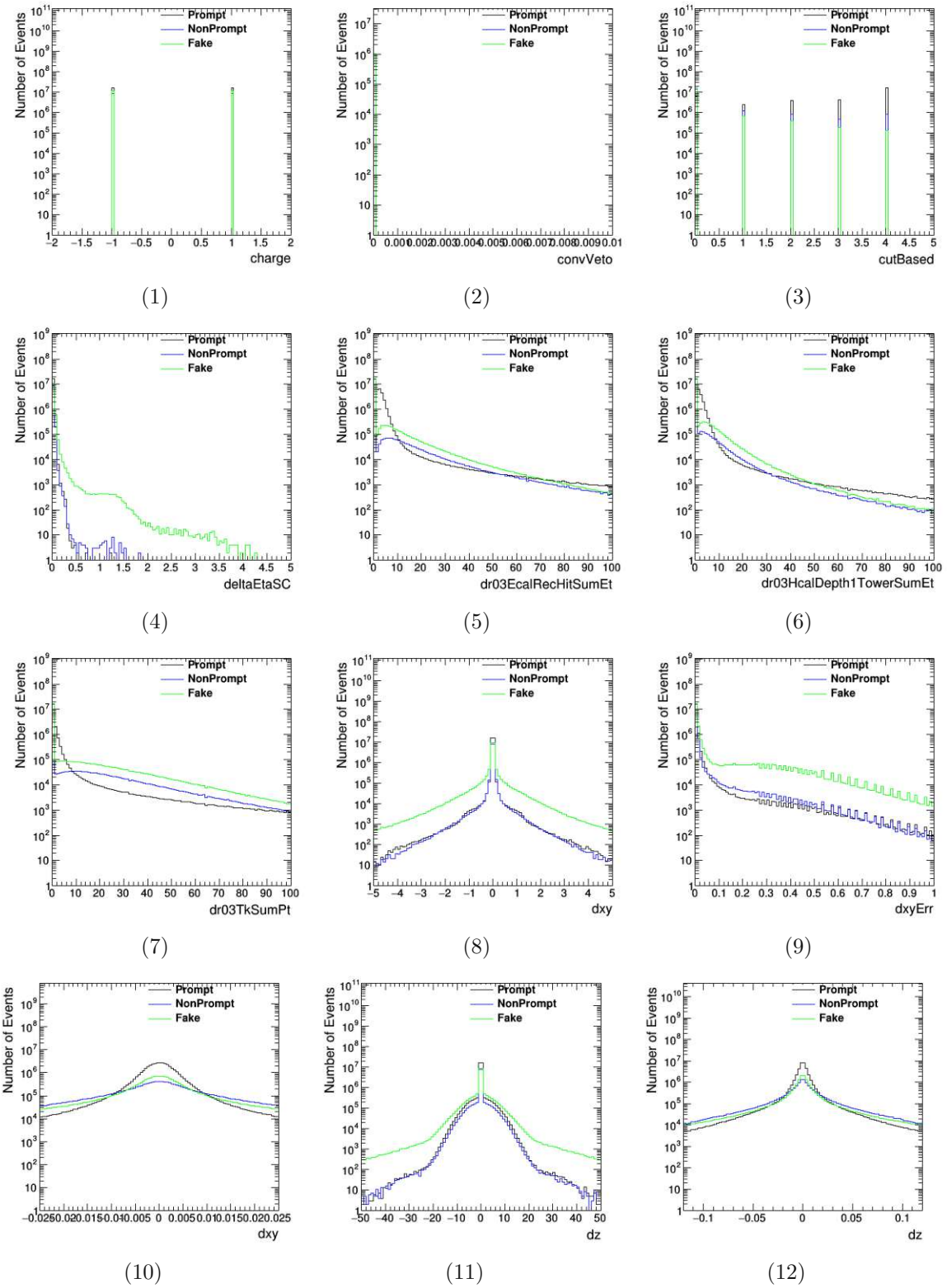


Figure B.9: Input feature plots before conversion 2018 electron DYvsQCD (1/4).

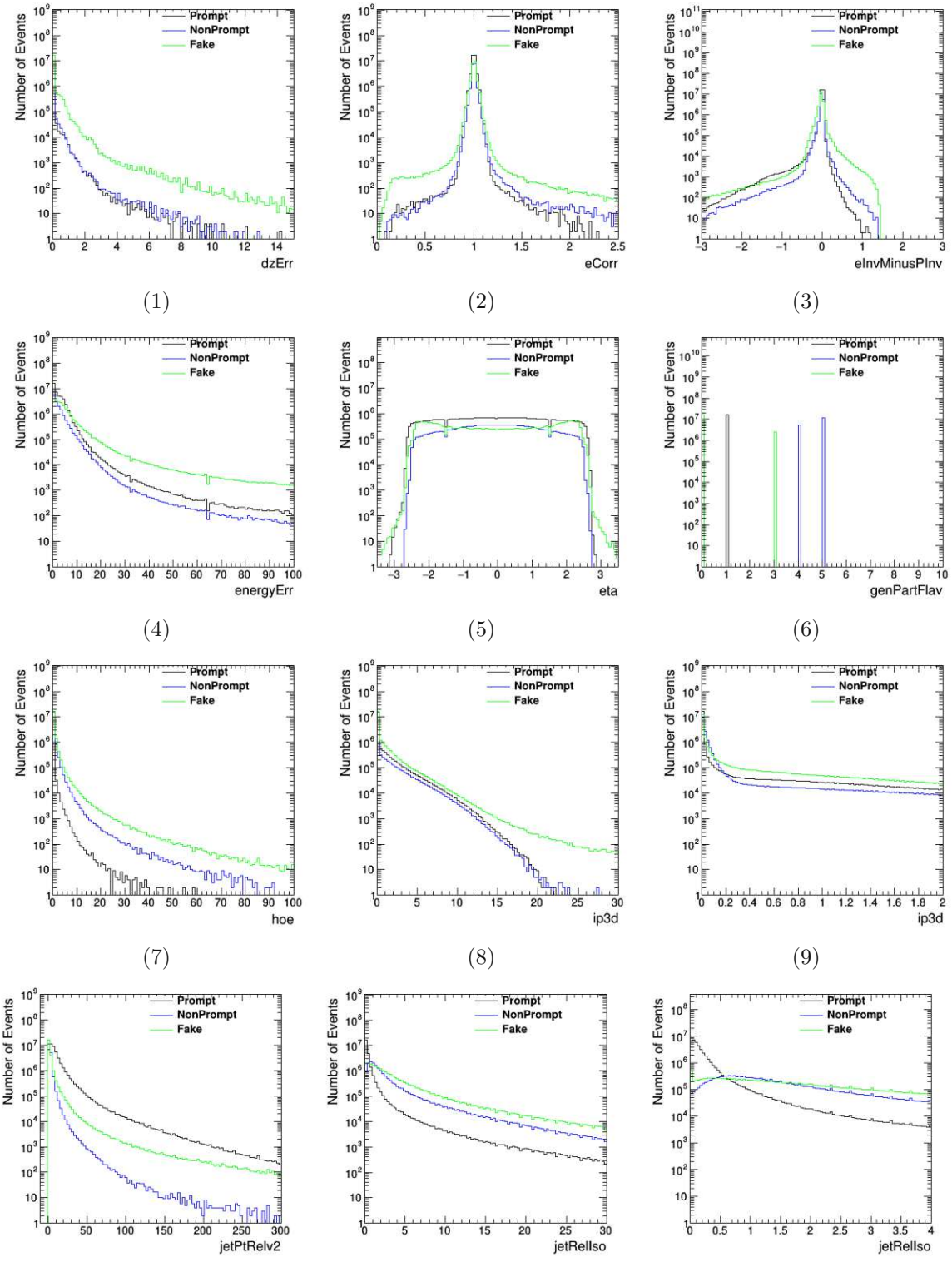


Figure B.10: Input feature plots before conversion 2018 electron DYvsQCD (2/4).

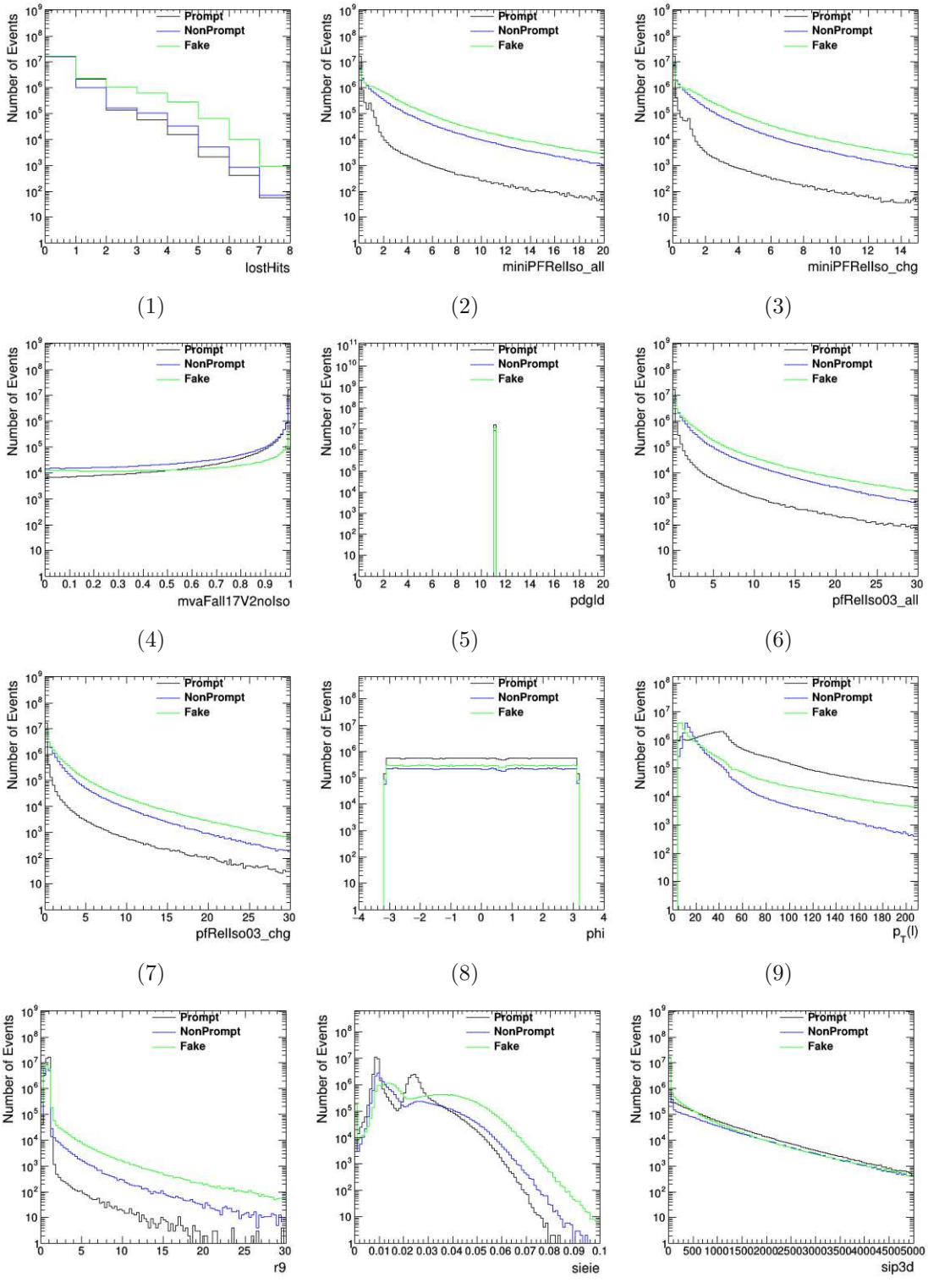


Figure B.11: Input feature plots before conversion 2018 electron DYvsQCD (3/4).

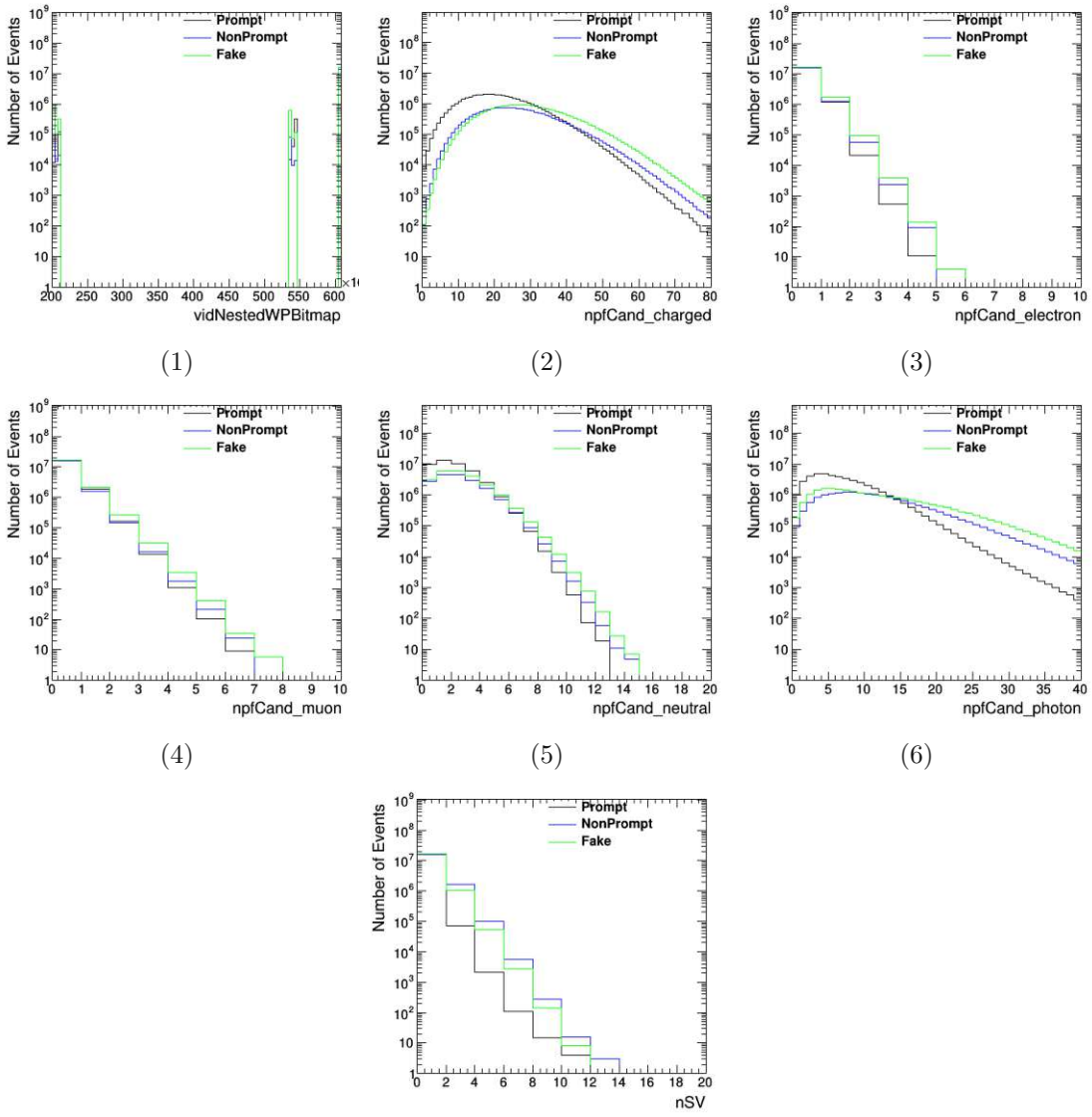
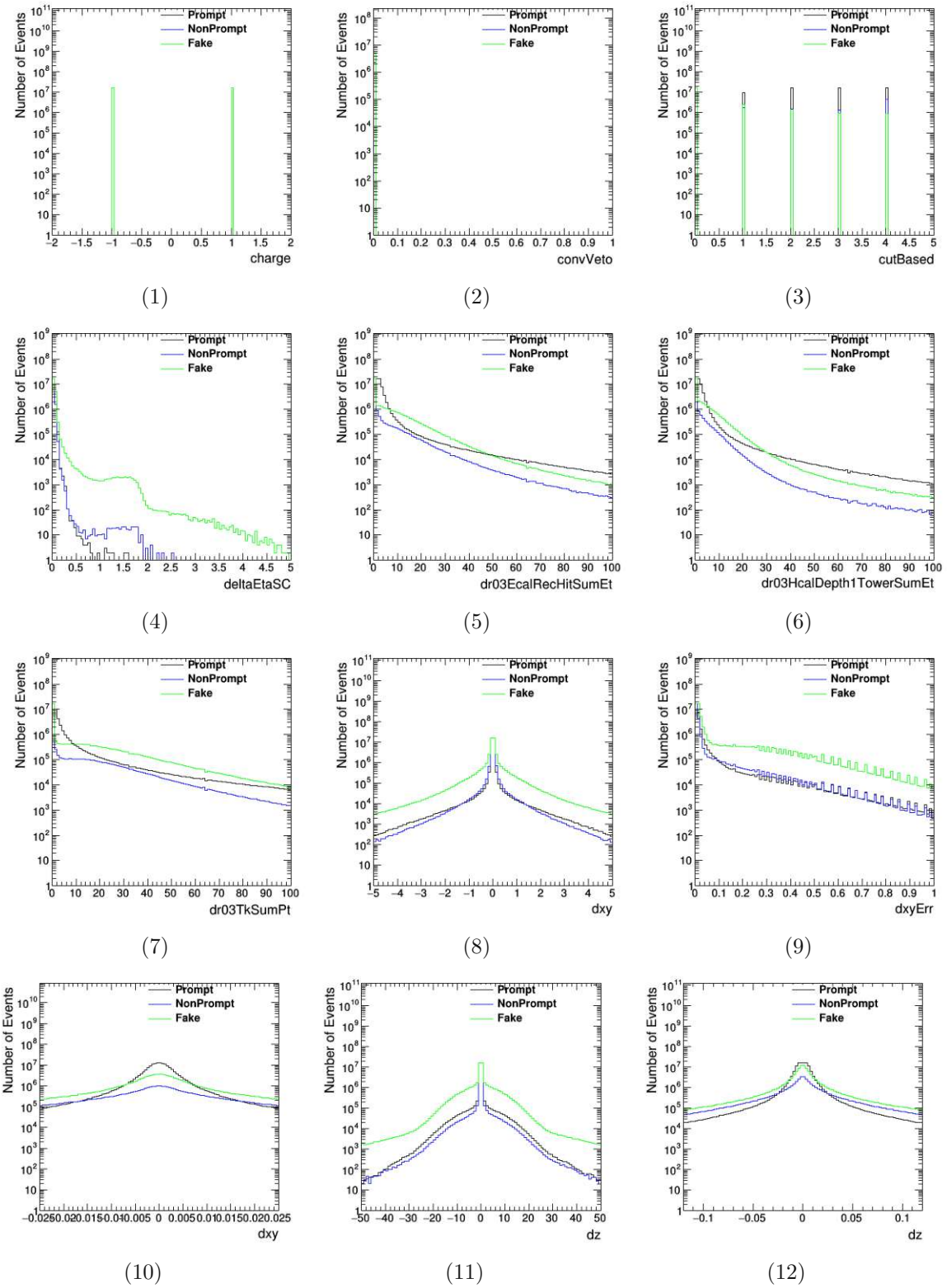


Figure B.12: Input feature plots before conversion 2018 electron DYvsQCD (4/4).



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11) (12)

Figure B.13: Input feature plots before conversion 2016 electron $t\bar{t}$ (1/4).

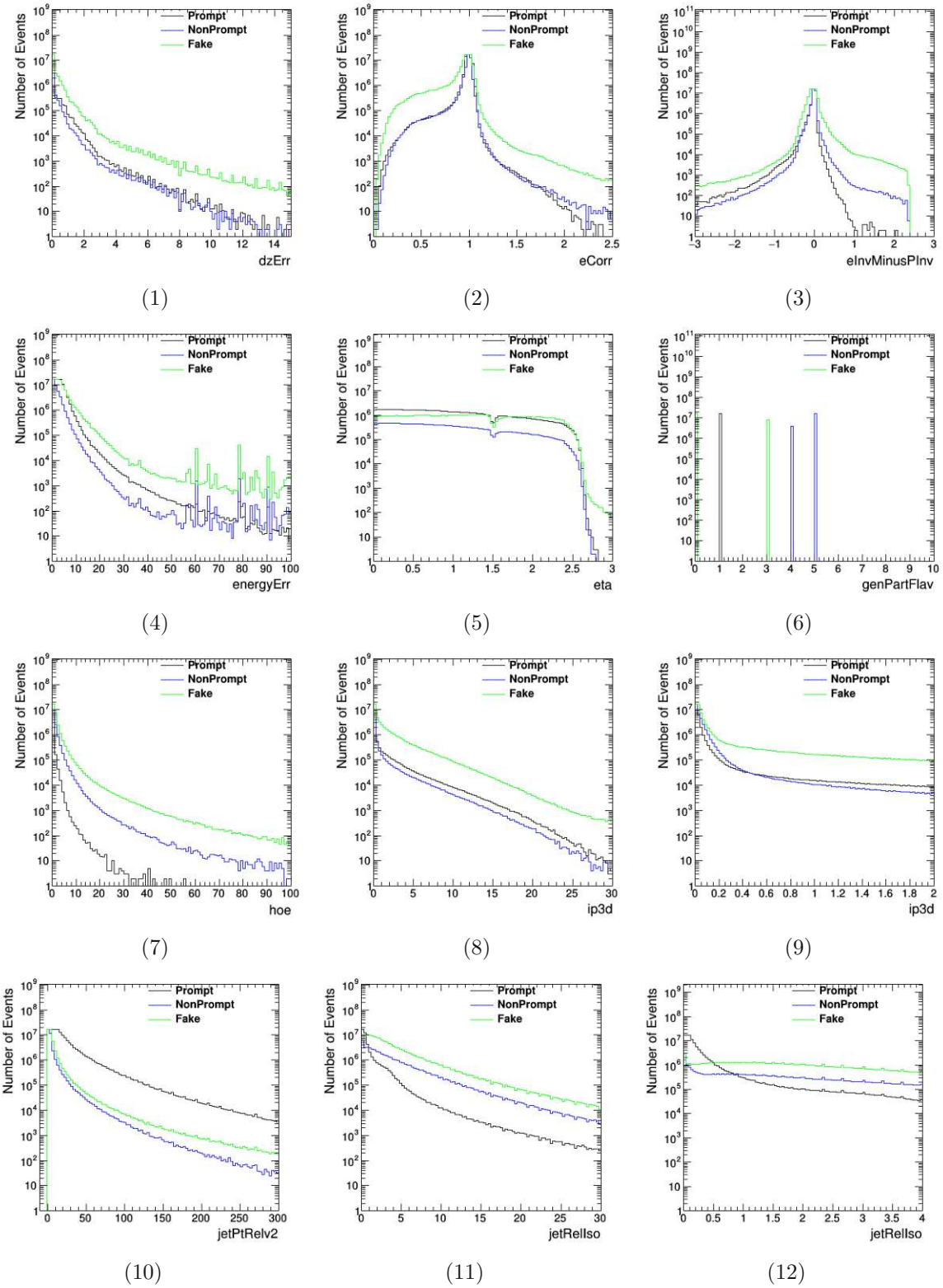


Figure B.14: Input feature plots before conversion 2016 electron $t\bar{t}$ (2/4).

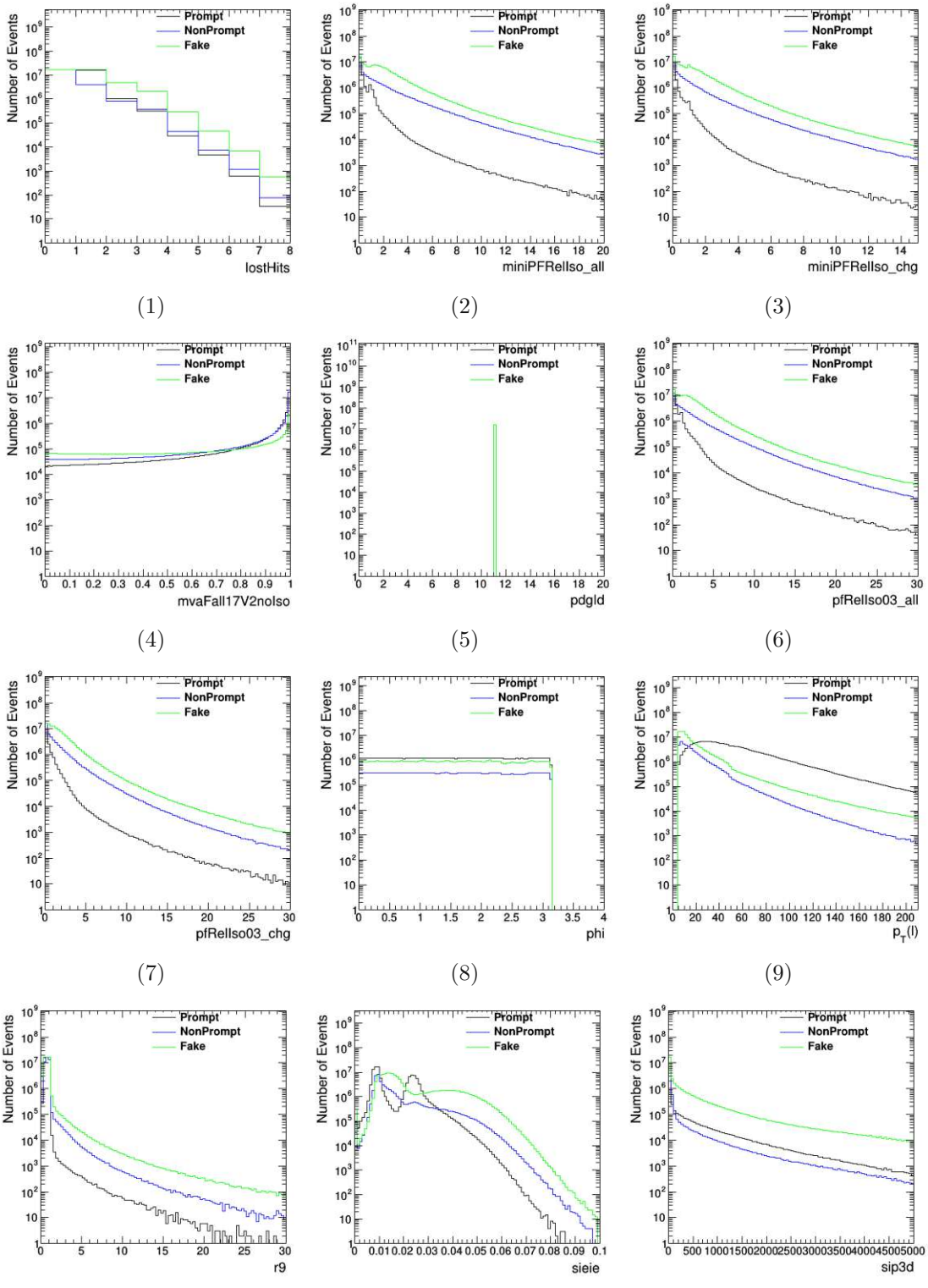


Figure B.15: Input feature plots before conversion 2016 electron $t\bar{t}$ (3/4).

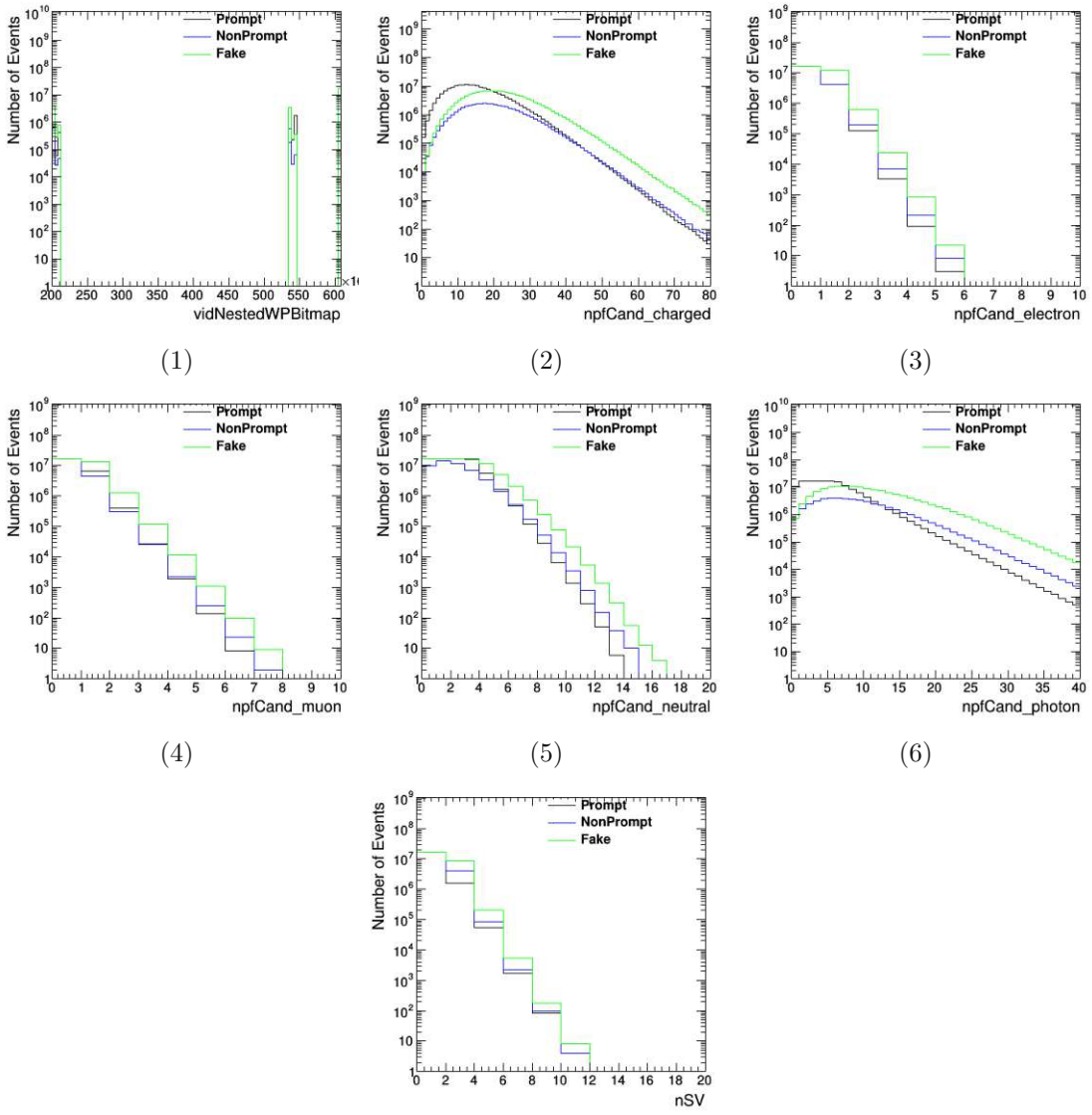


Figure B.16: Input feature plots before conversion 2016 electron $t\bar{t}$ (4/4).

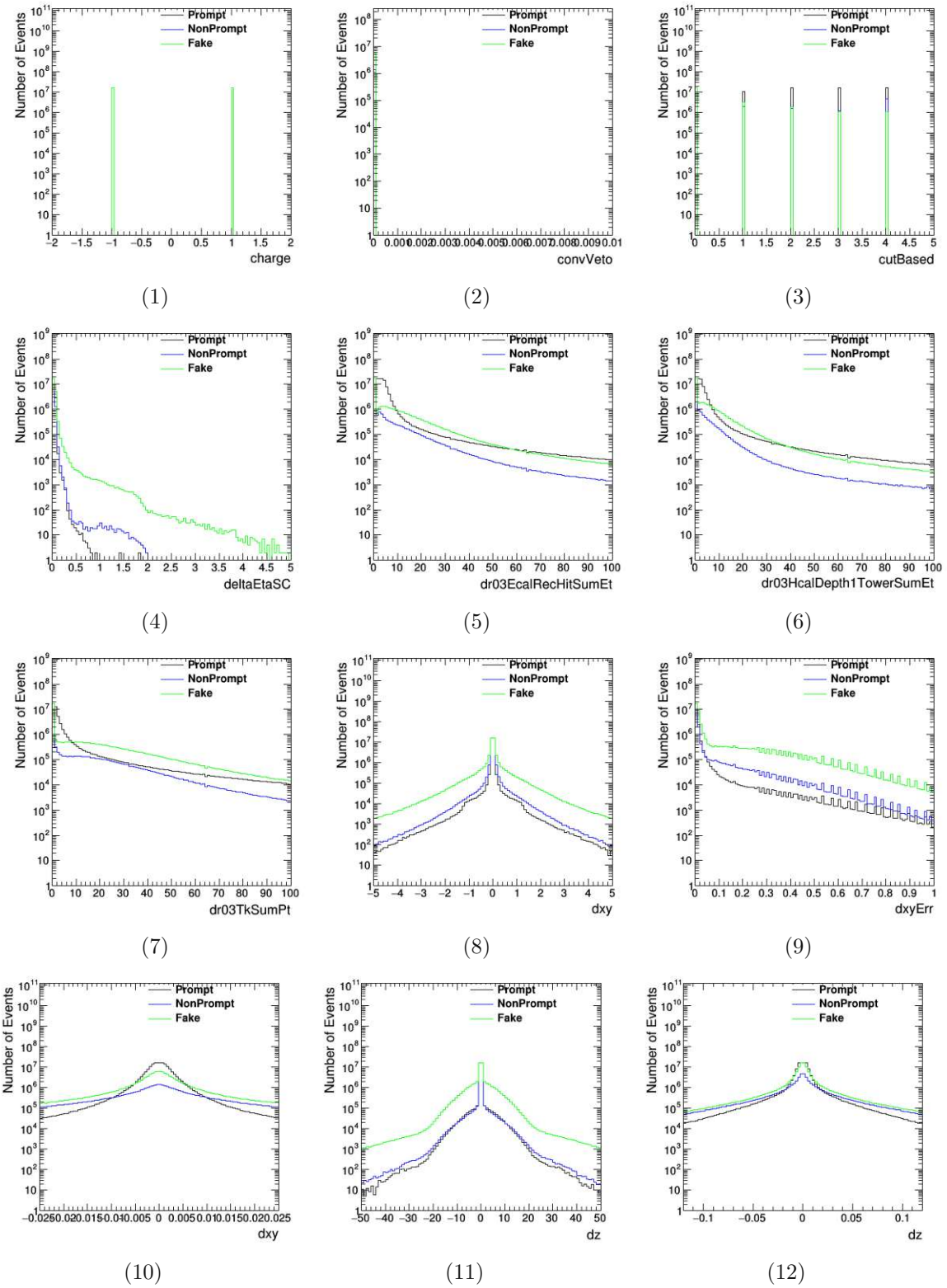


Figure B.17: Input feature plots before conversion 2017 electron $t\bar{t}$ (1/4).

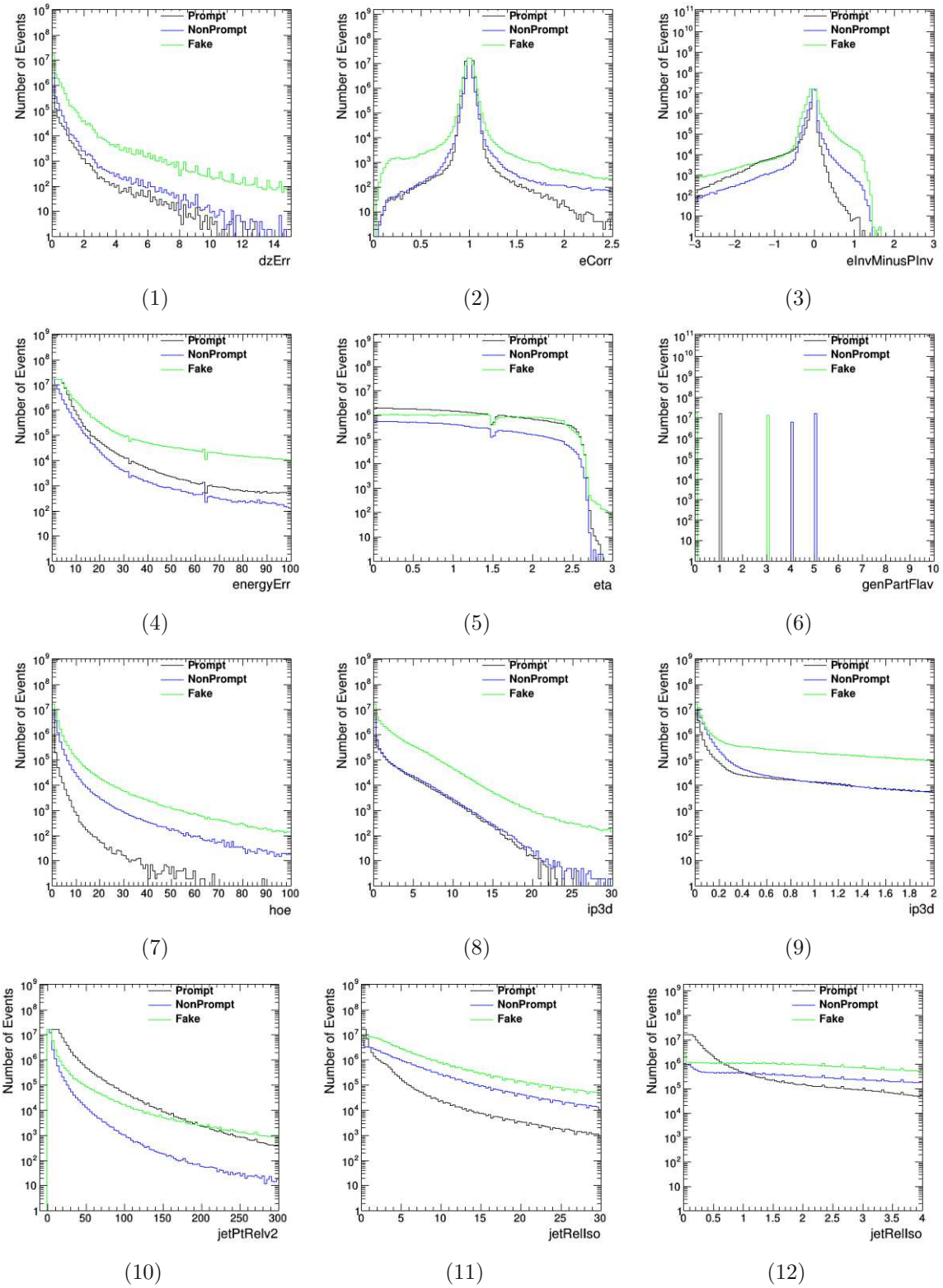


Figure B.18: Input feature plots before conversion 2017 electron $t\bar{t}$ (2/4).

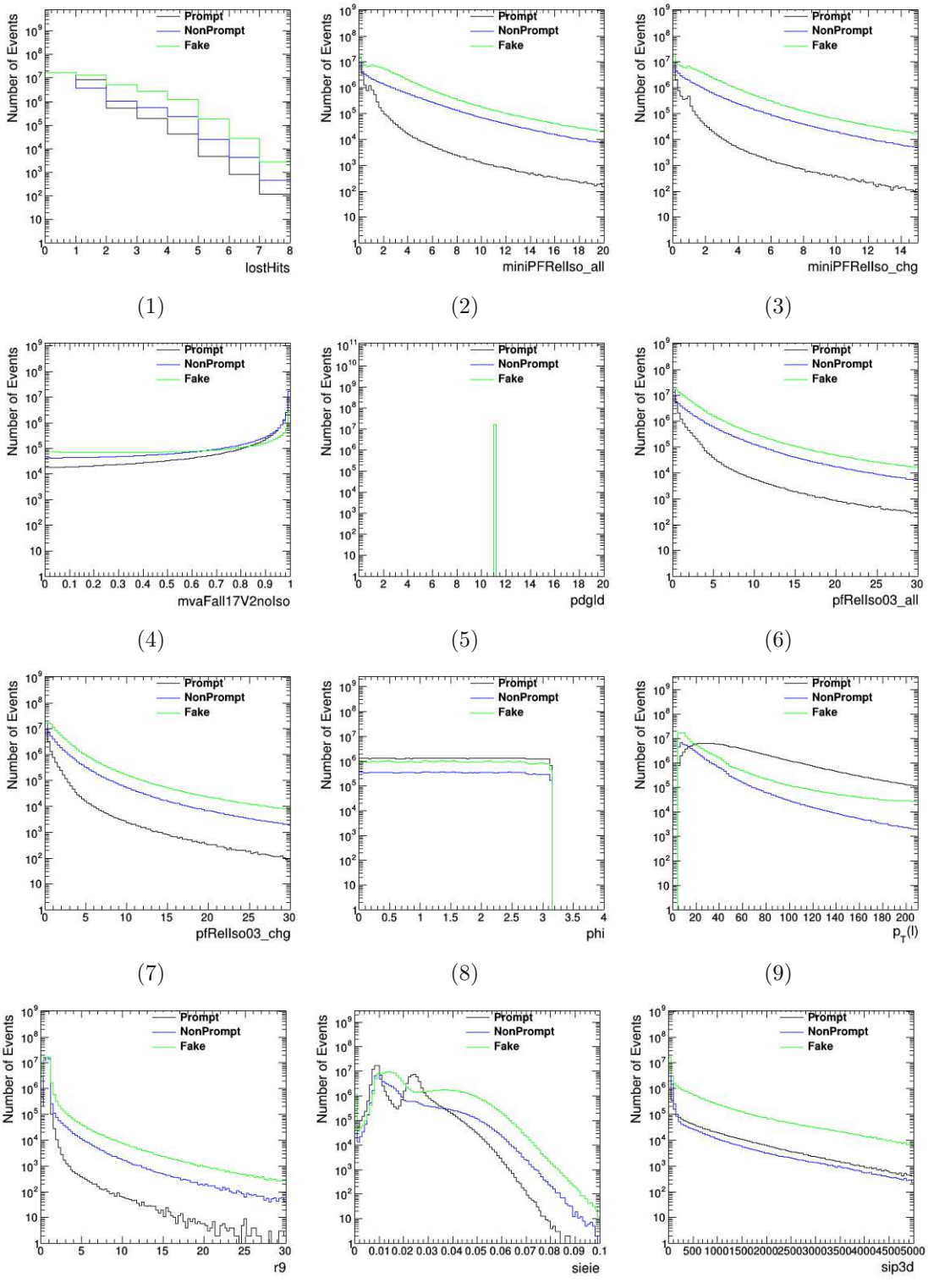


Figure B.19: Input feature plots before conversion 2017 electron $t\bar{t}$ (3/4).

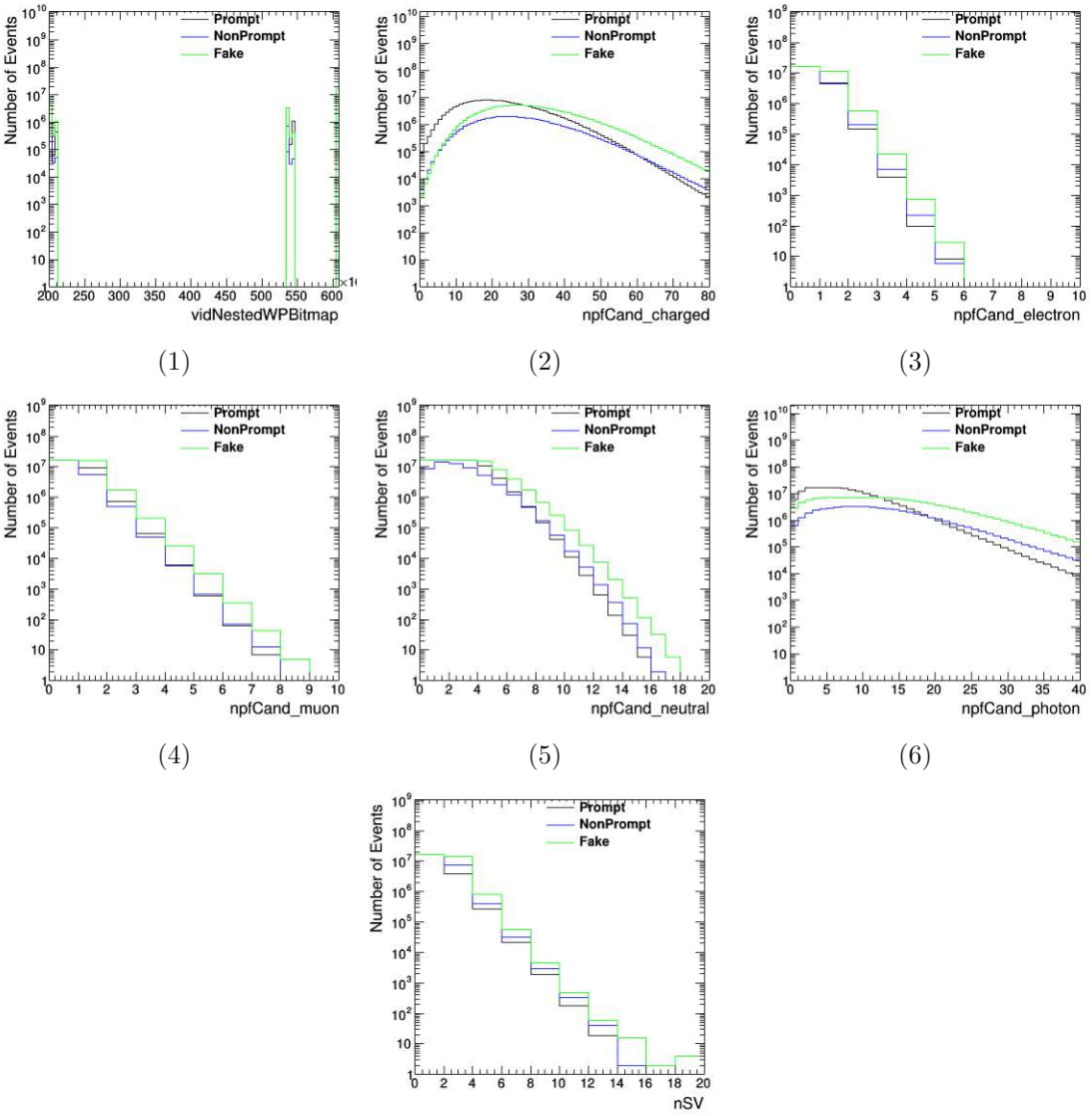
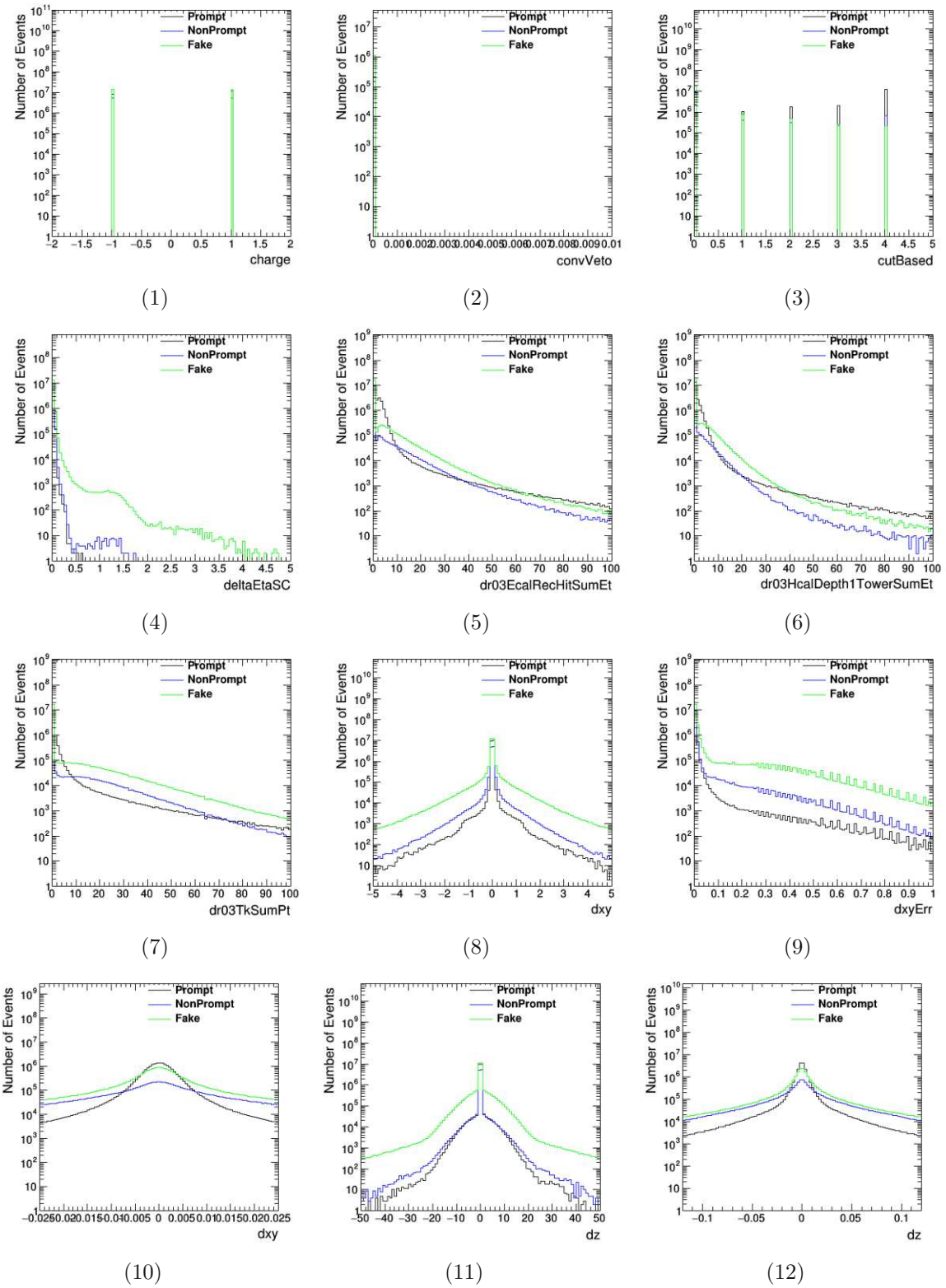


Figure B.20: Input feature plots before conversion 2017 electron $t\bar{t}$ (4/4).



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11) (12)

Figure B.21: Input feature plots before conversion 2018 electron $t\bar{t}$ (1/4).

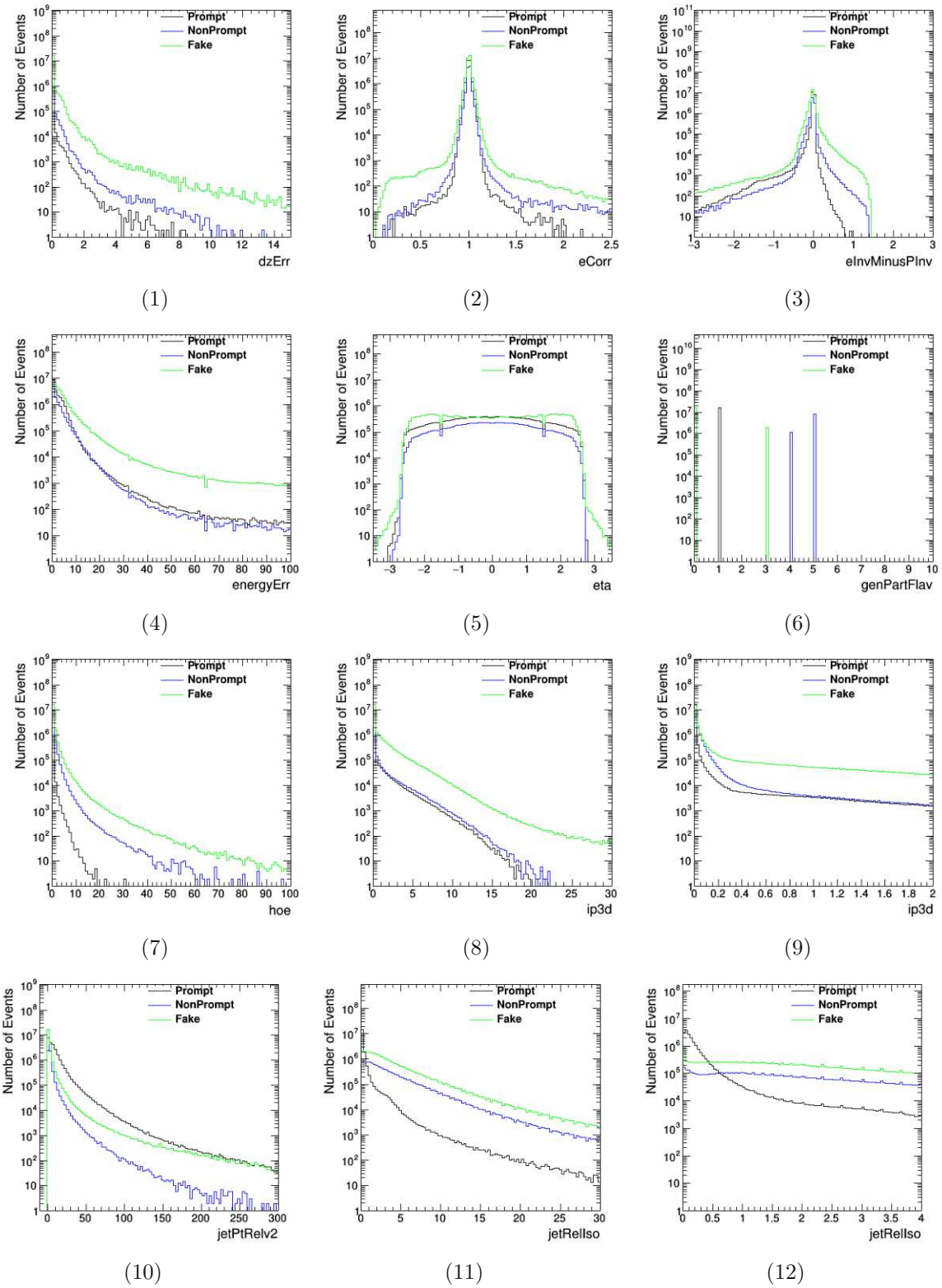


Figure B.22: Input feature plots before conversion 2018 electron $t\bar{t}$ (2/4).

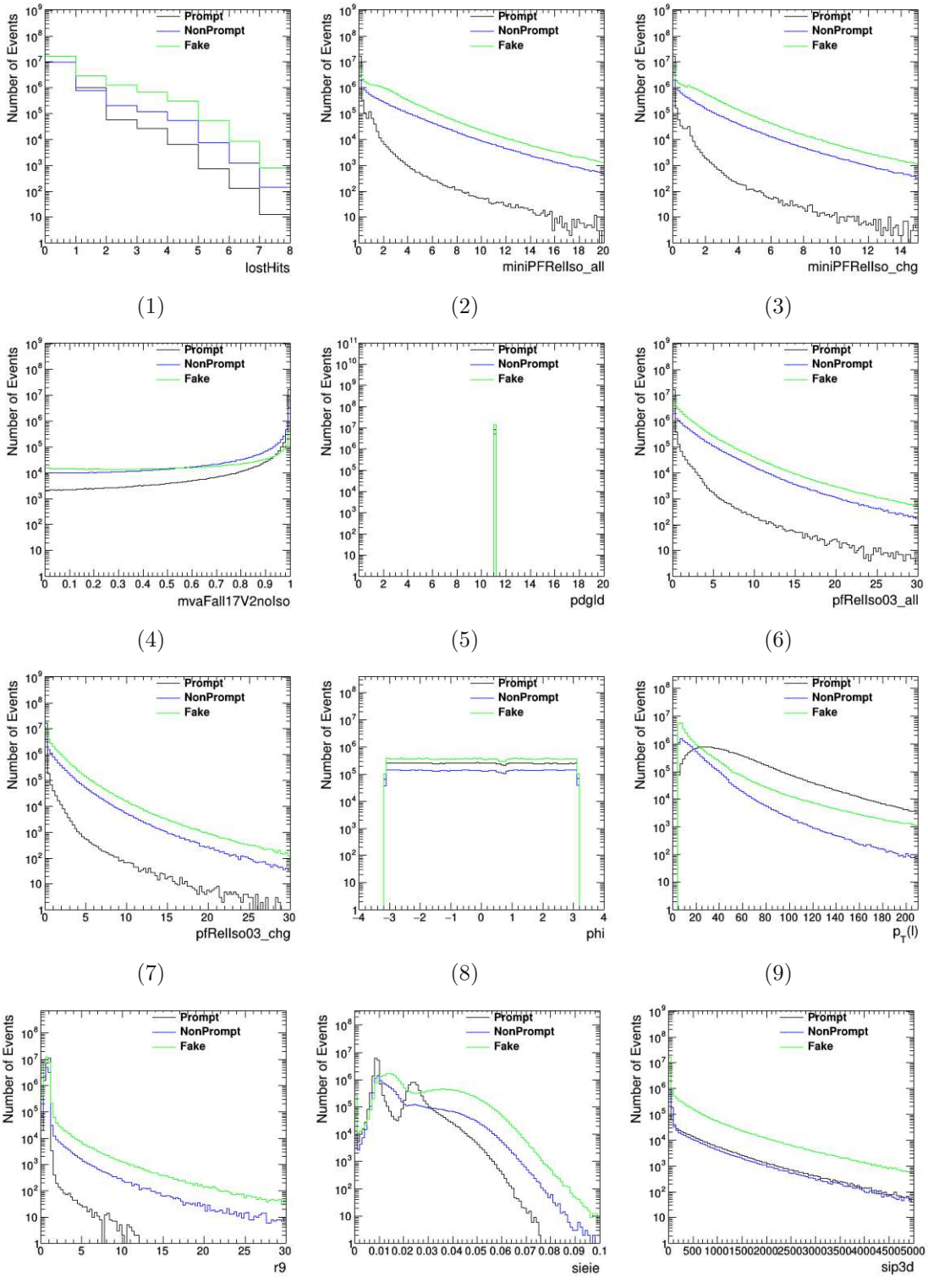


Figure B.23: Input feature plots before conversion 2018 electron $t\bar{t}$ (3/4).

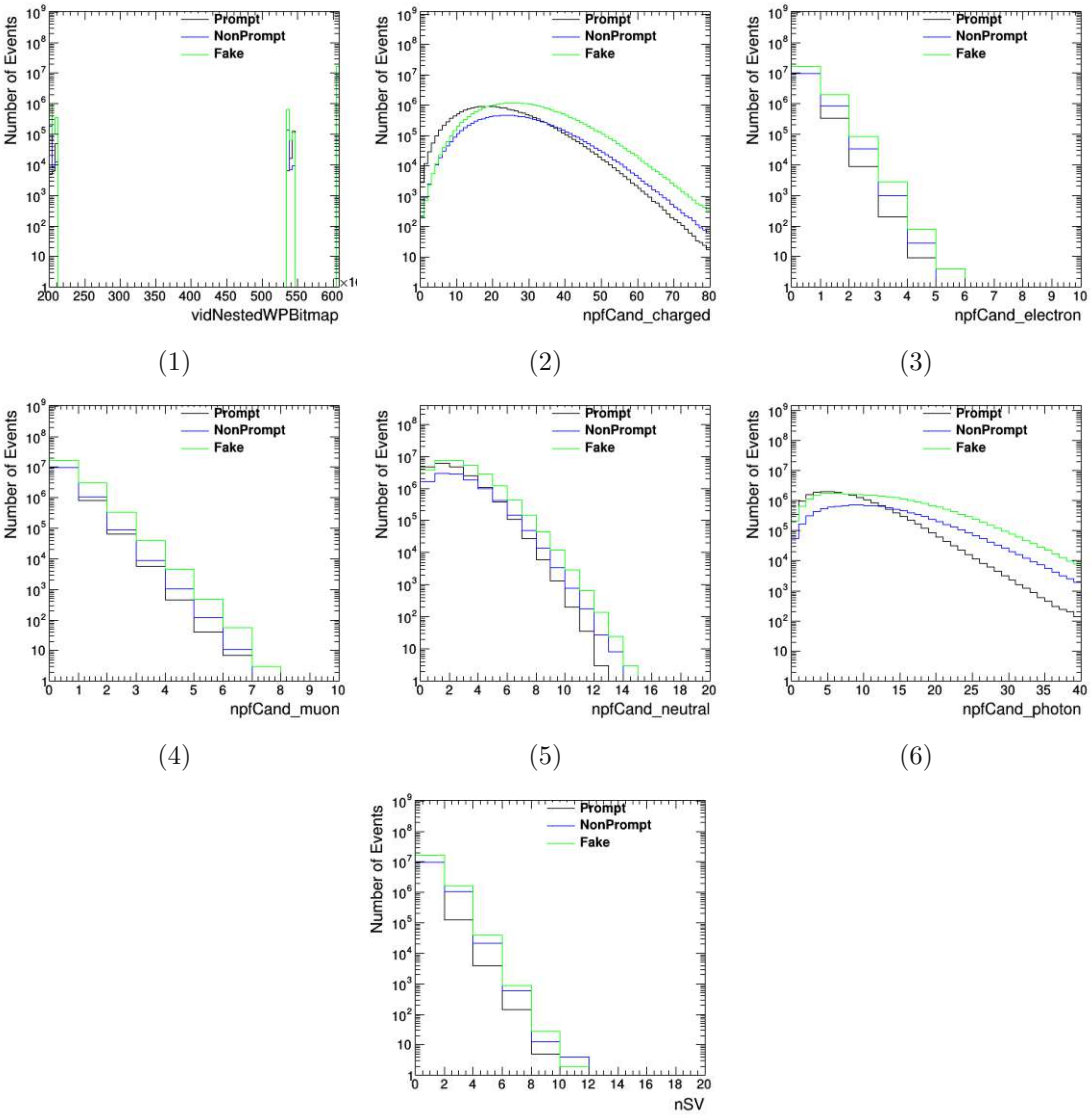


Figure B.24: Input feature plots before conversion 2018 electron $t\bar{t}$ (4/4).

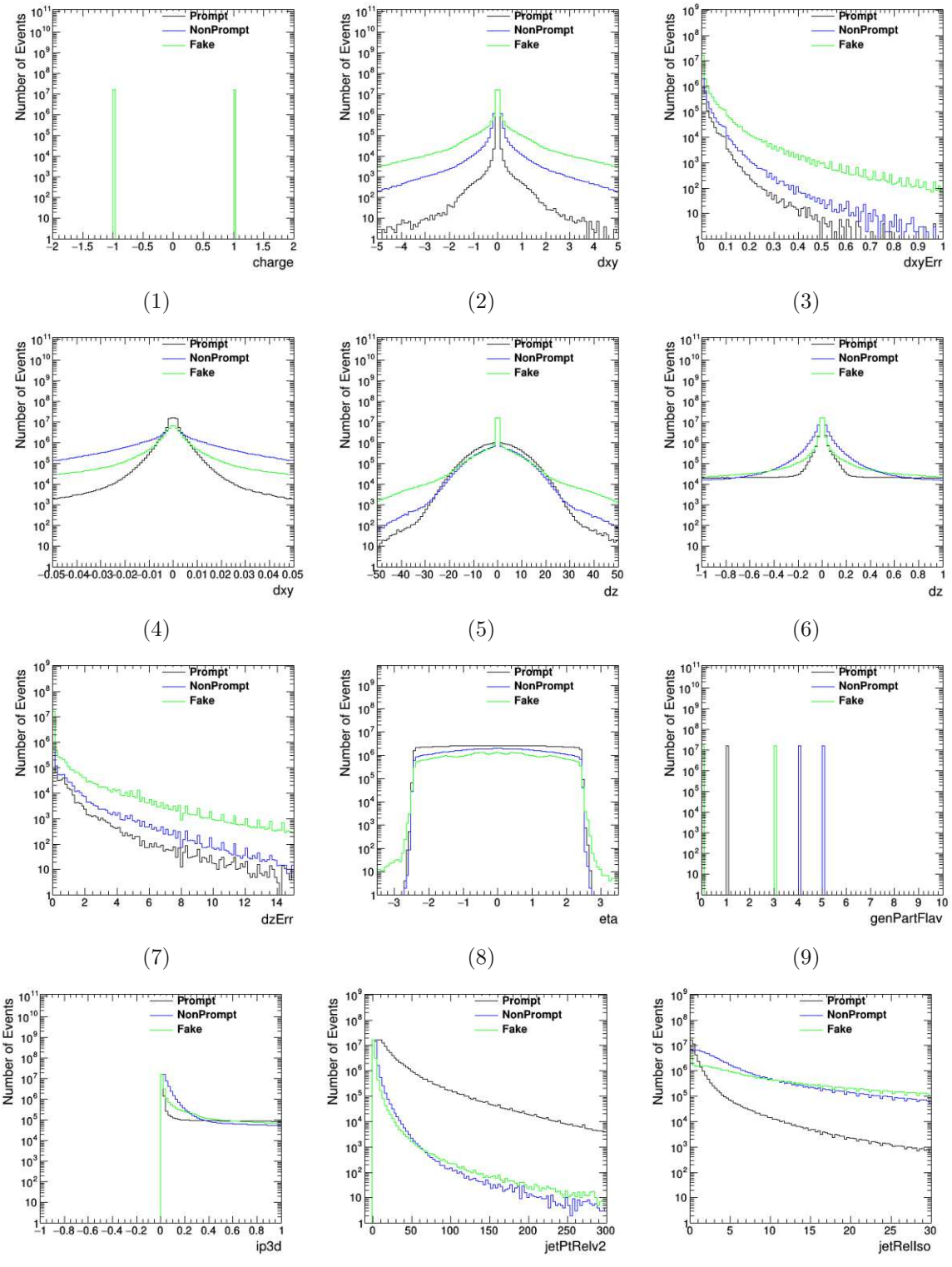


Figure B.25: Input feature plots before conversion 2016 muon DYvsQCD (1/3).

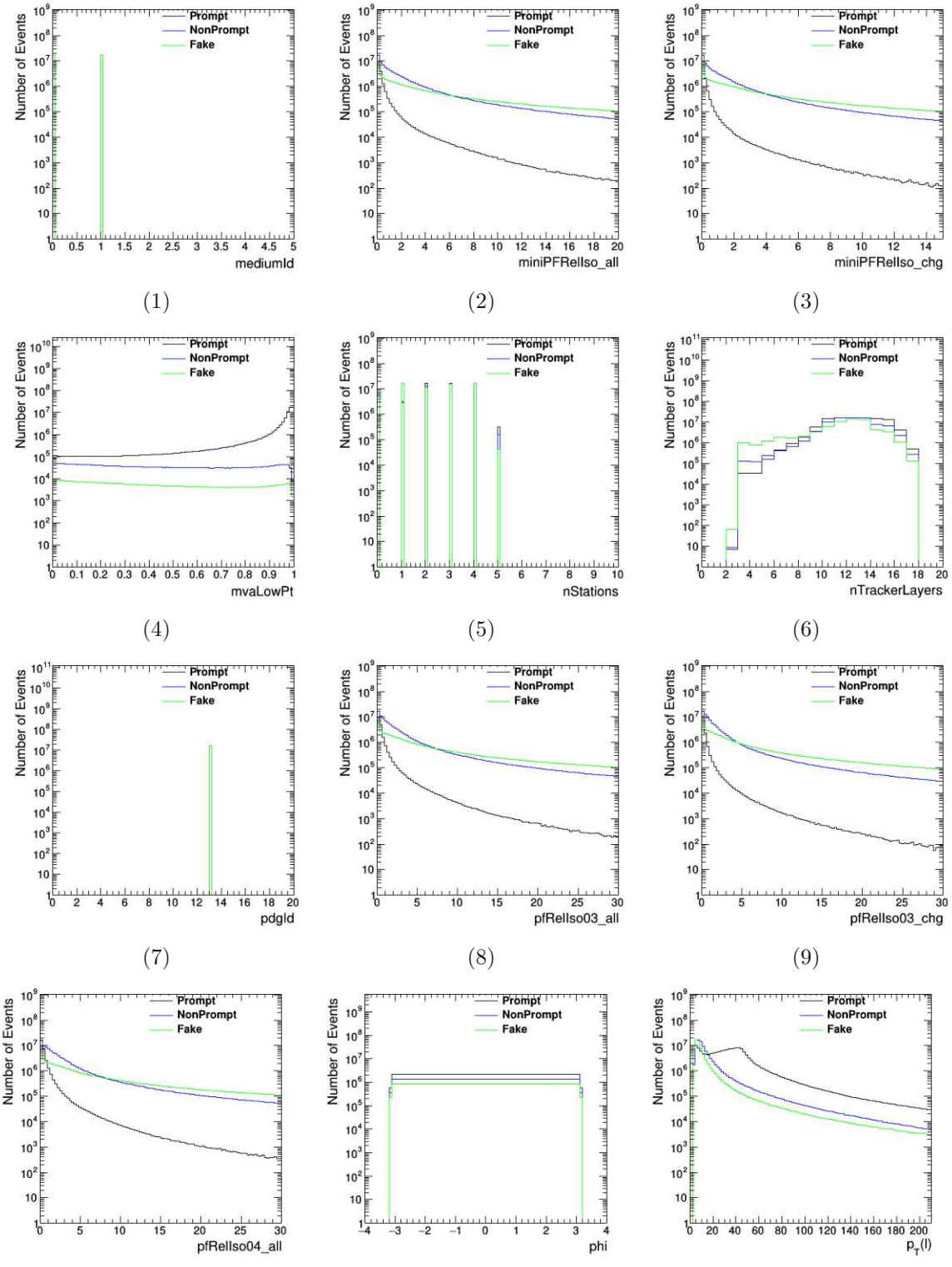


Figure B.26: Input feature plots before conversion 2016 muon DYvsQCD (2/3).

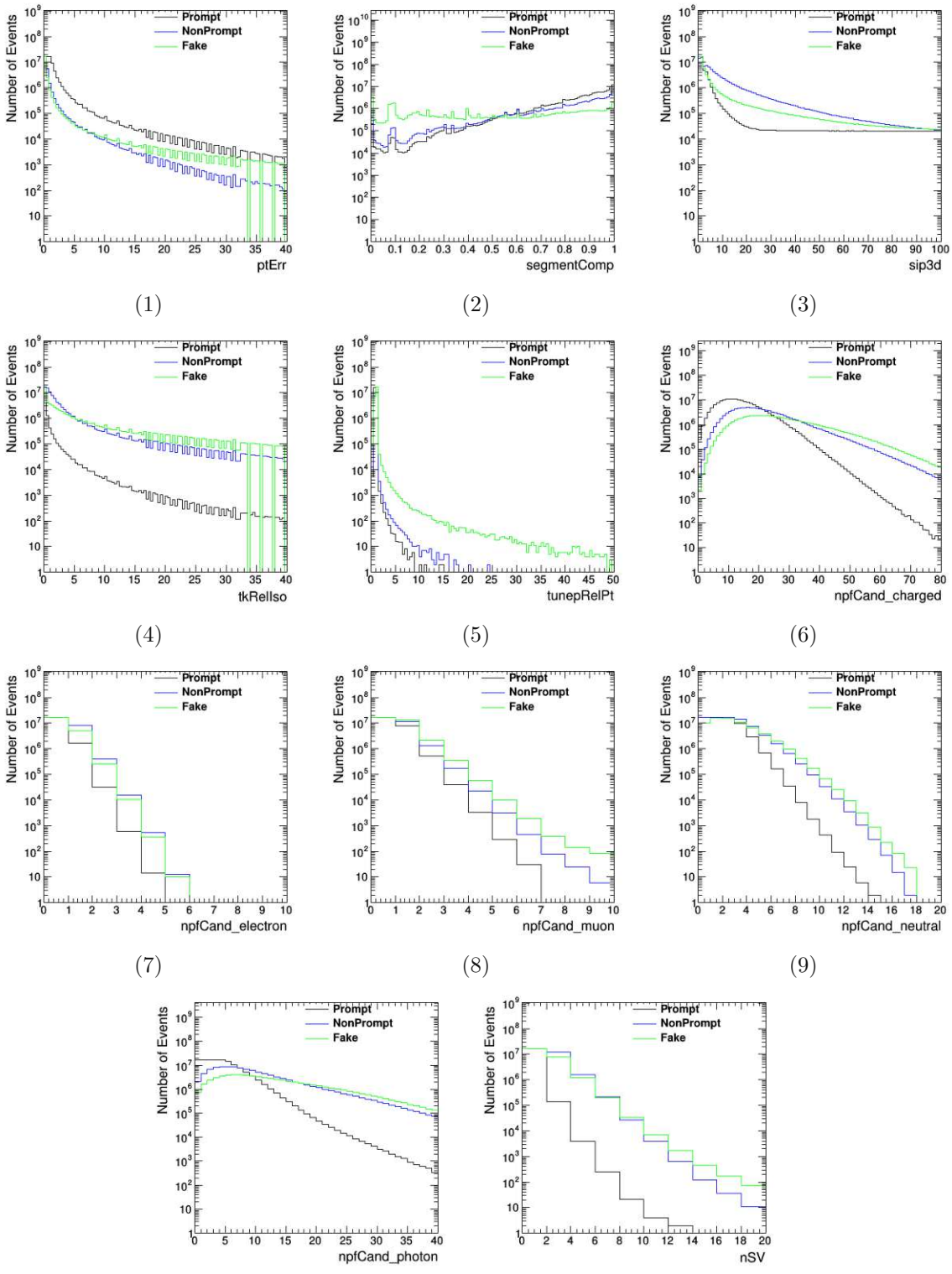


Figure B.27: Input feature plots before conversion 2016 muon DYvsQCD (3/3).

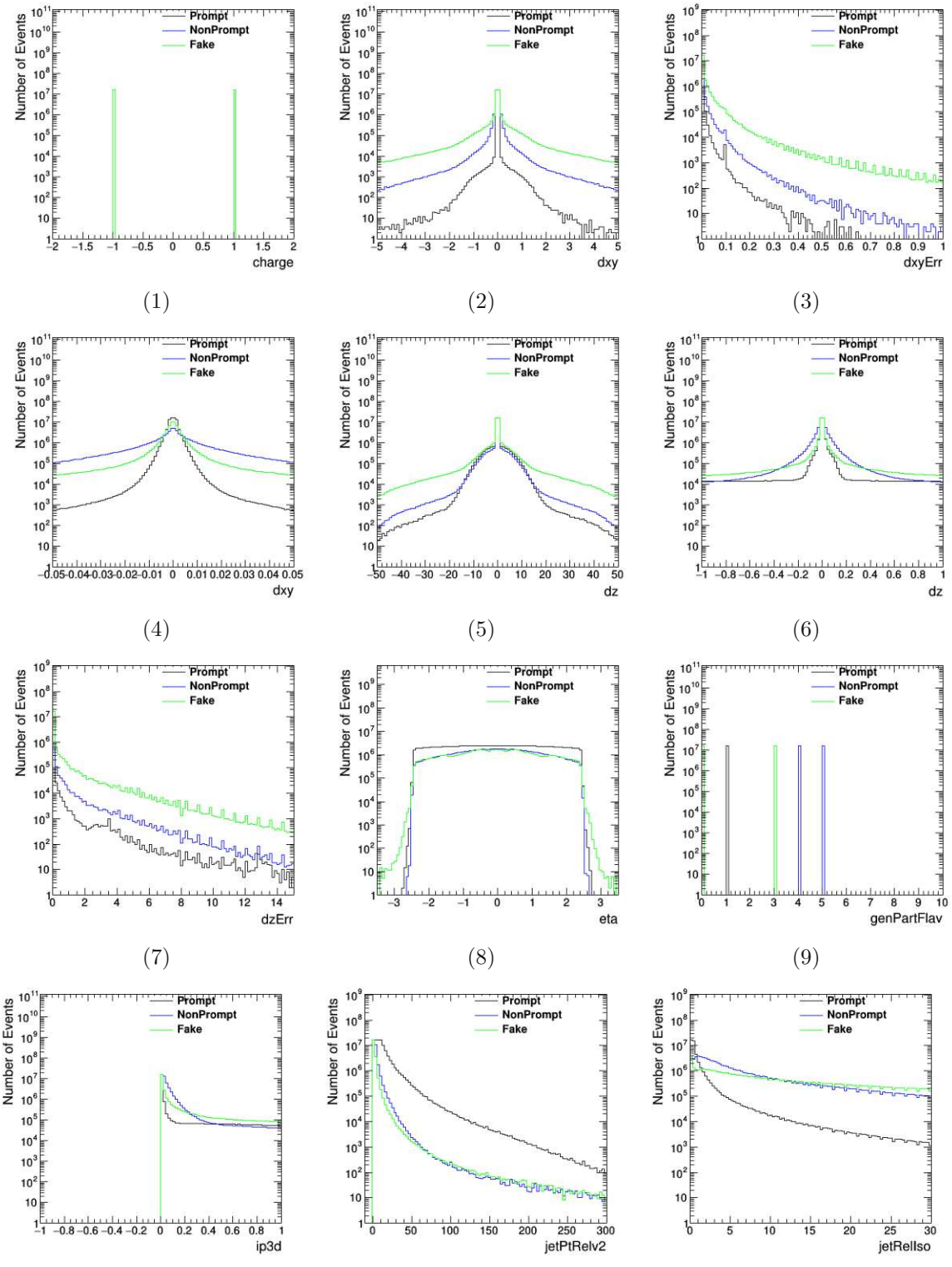


Figure B.28: Input feature plots before conversion 2017 muon DYvsQCD (1/3).

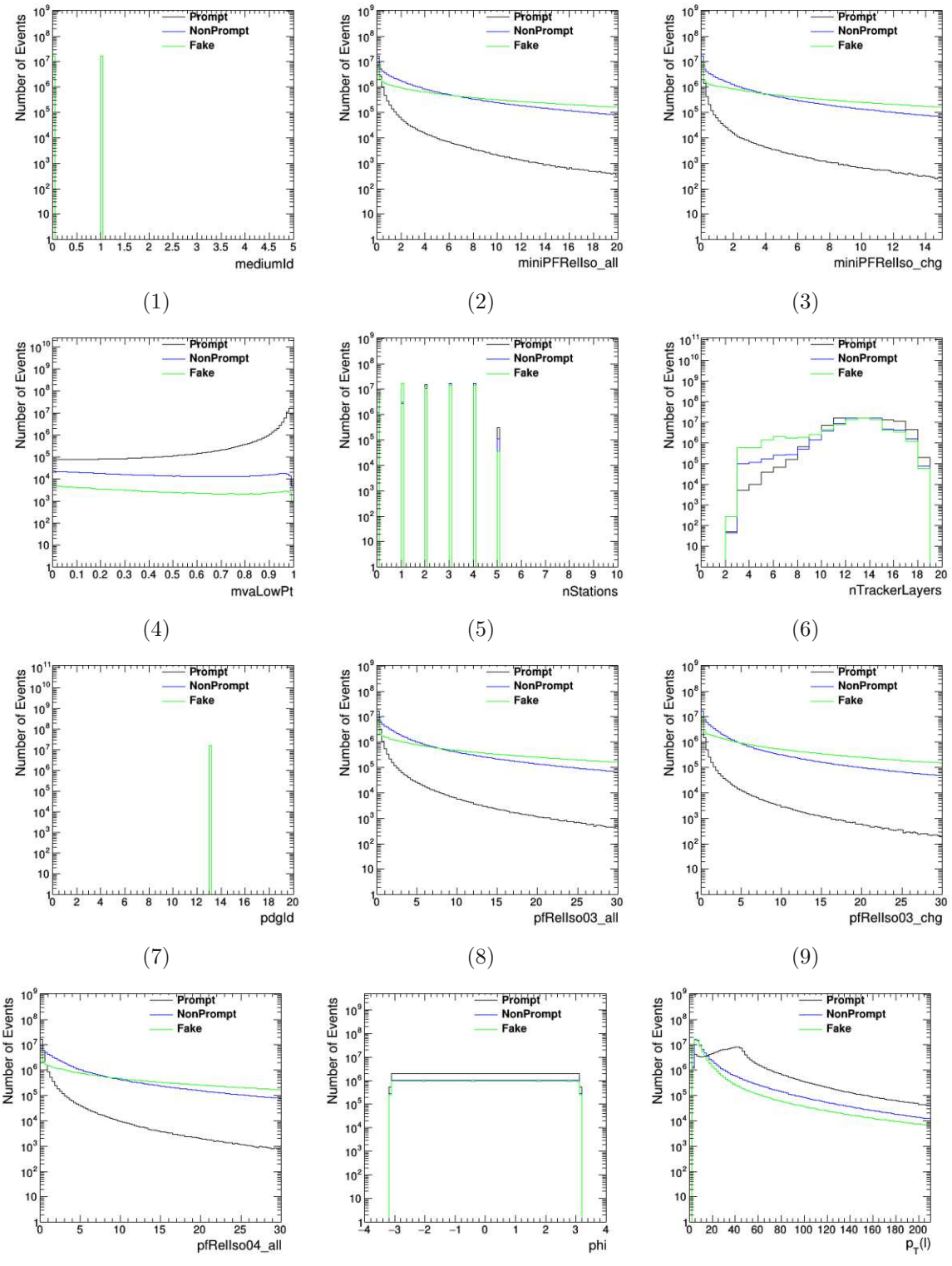
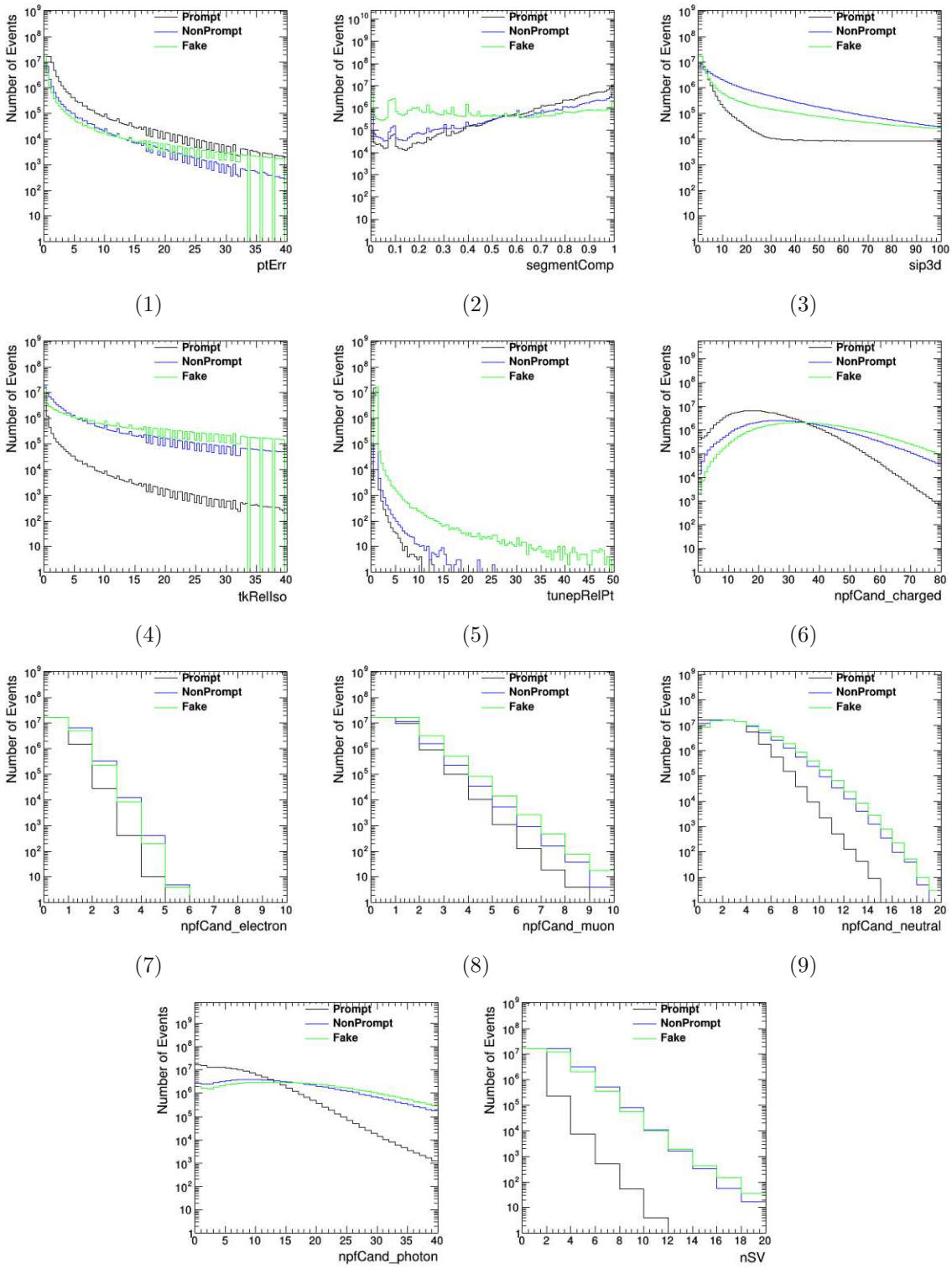


Figure B.29: Input feature plots before conversion 2017 muon DYvsQCD (2/3).



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11)

Figure B.30: Input feature plots before conversion 2017 muon DYvsQCD (3/3).

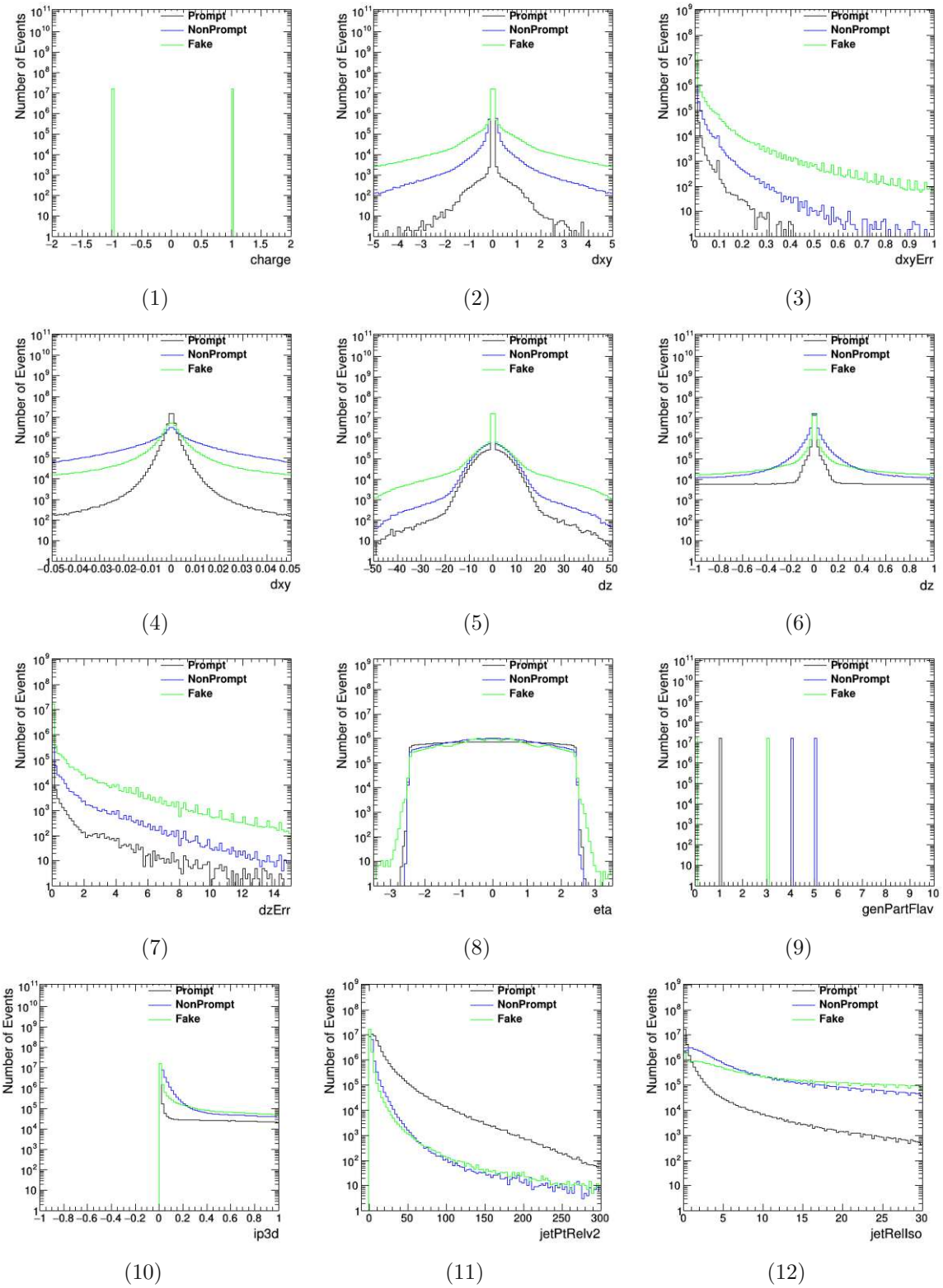


Figure B.31: Input feature plots before conversion 2018 muon DYvsQCD (1/3).

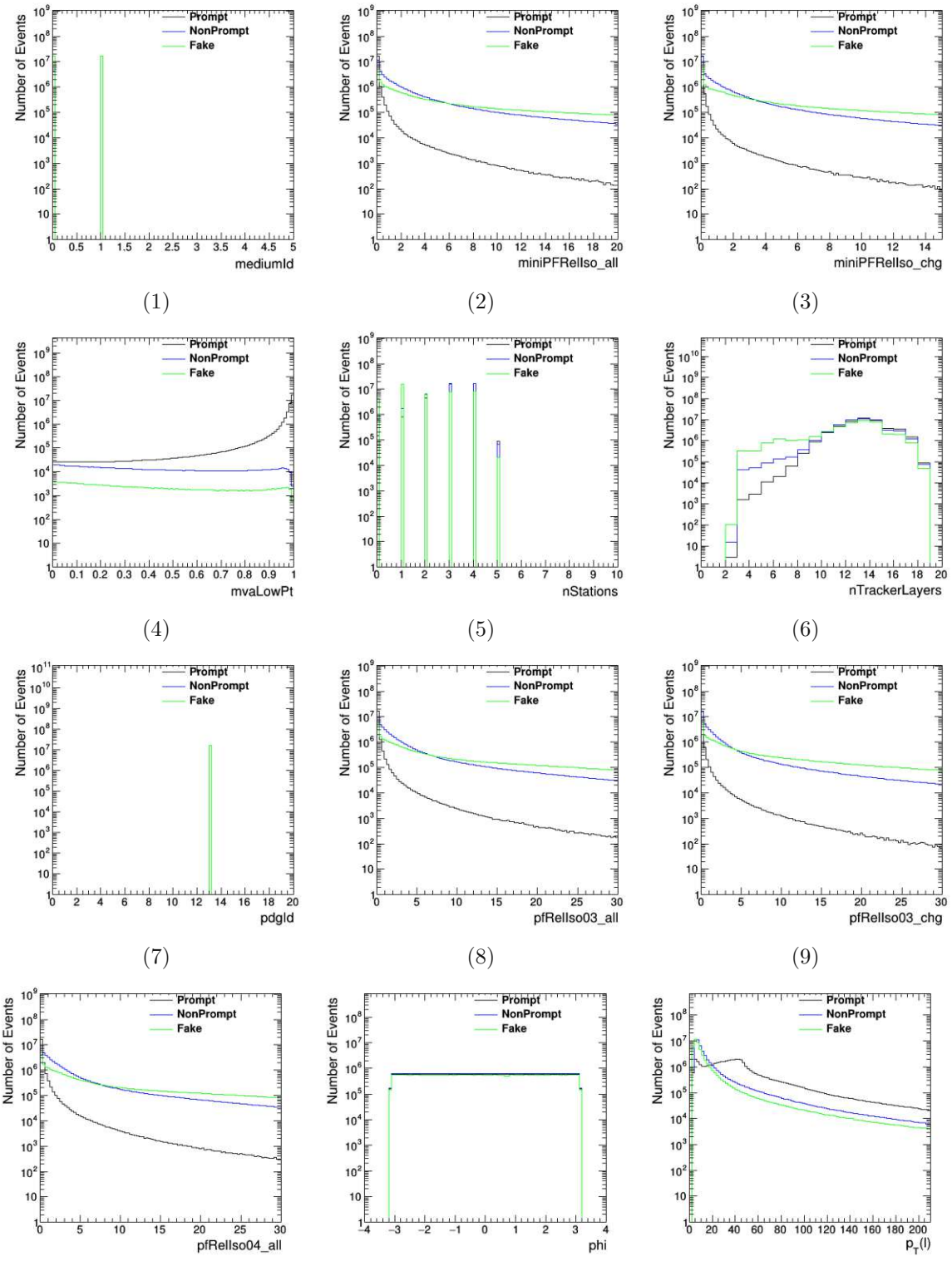
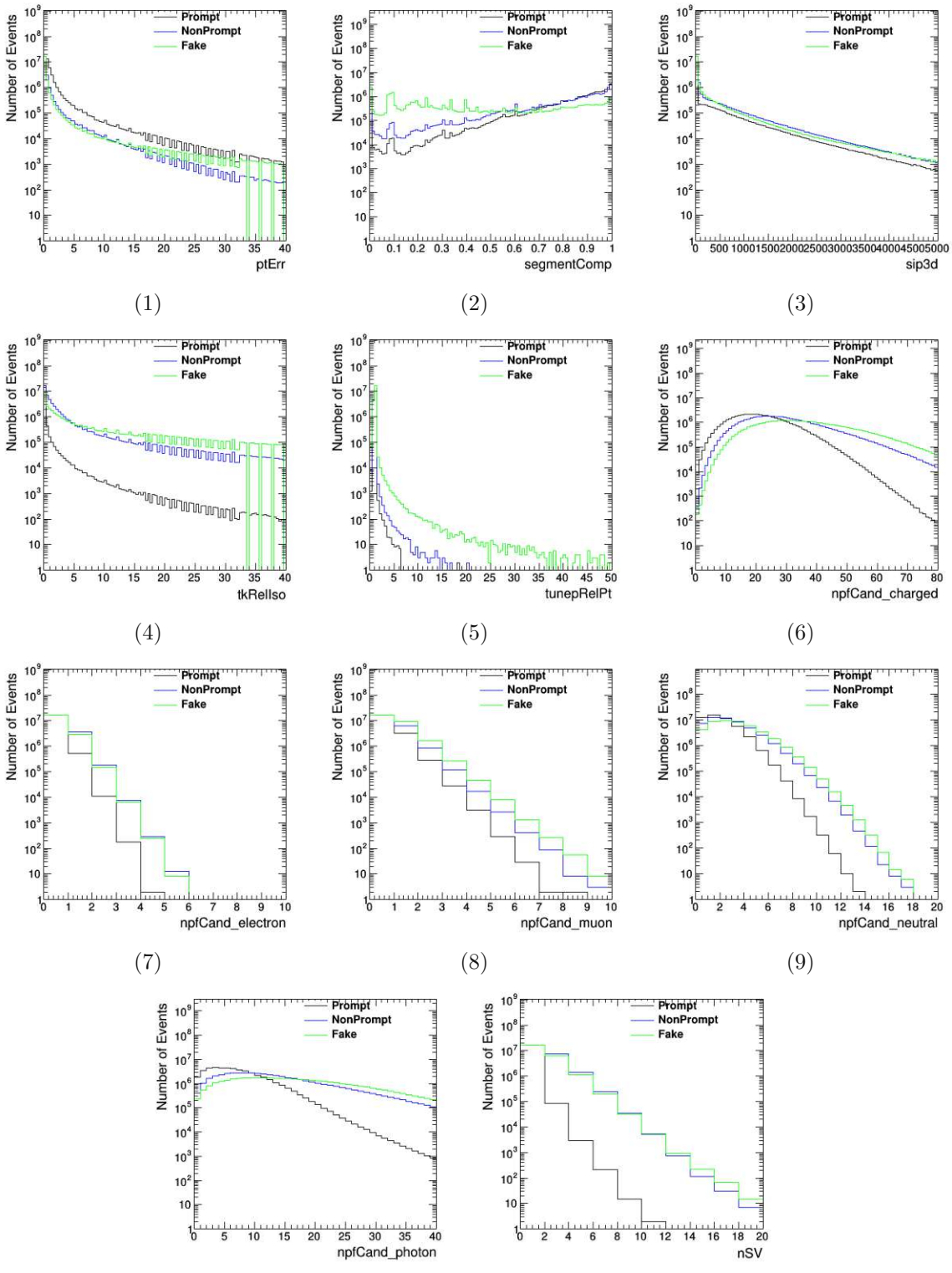


Figure B.32: Input feature plots before conversion 2018 muon DYvsQCD (2/3).



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11)

Figure B.33: Input feature plots before conversion 2018 muon DYvsQCD (3/3).

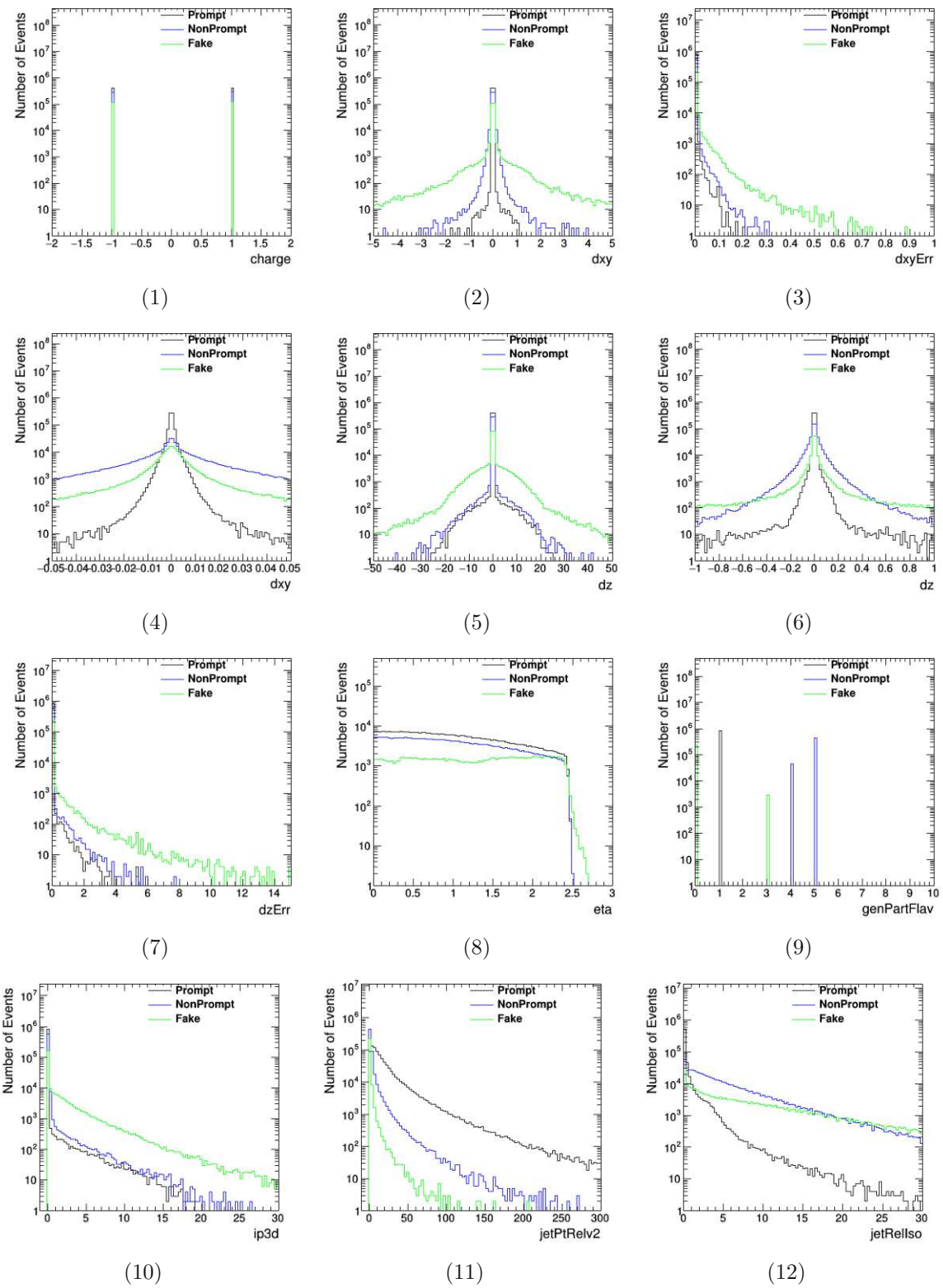


Figure B.34: Input feature plots before conversion 2016 muon $t\bar{t}$ (1/3).

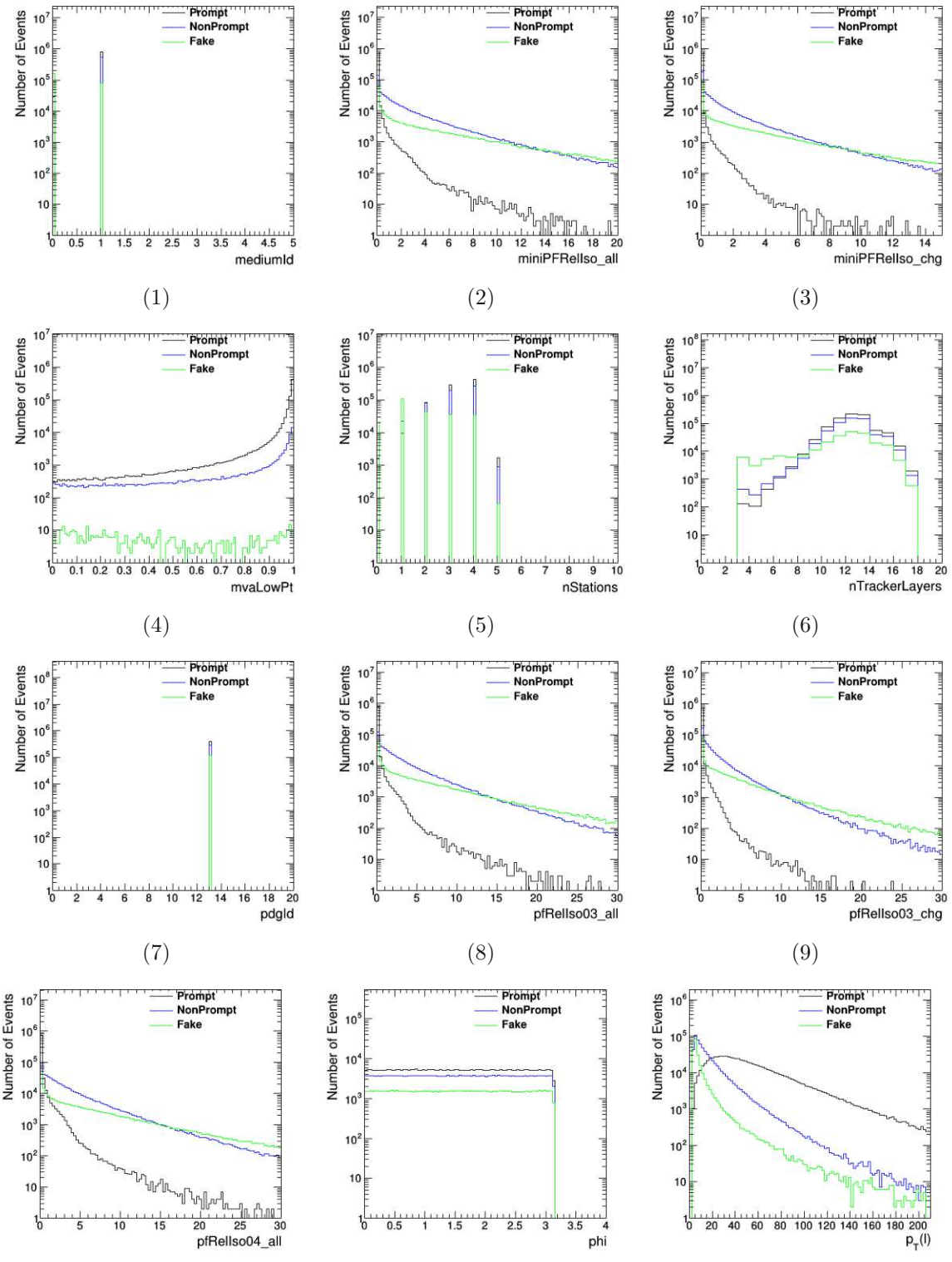
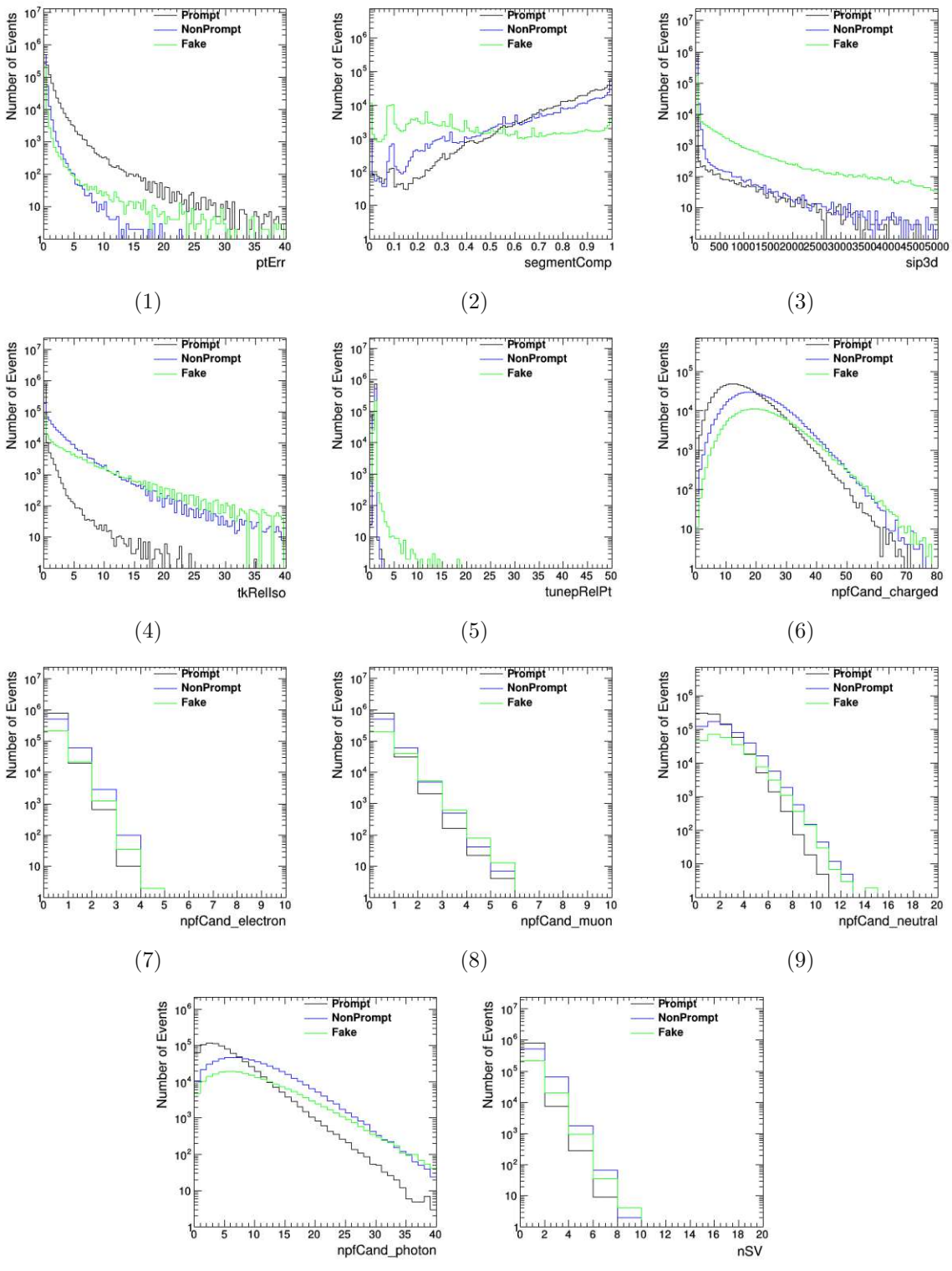


Figure B.35: Input feature plots before conversion 2016 muon $t\bar{t}$ (2/3).



(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11)

Figure B.36: Input feature plots before conversion 2016 muon $t\bar{t}$ (3/3).

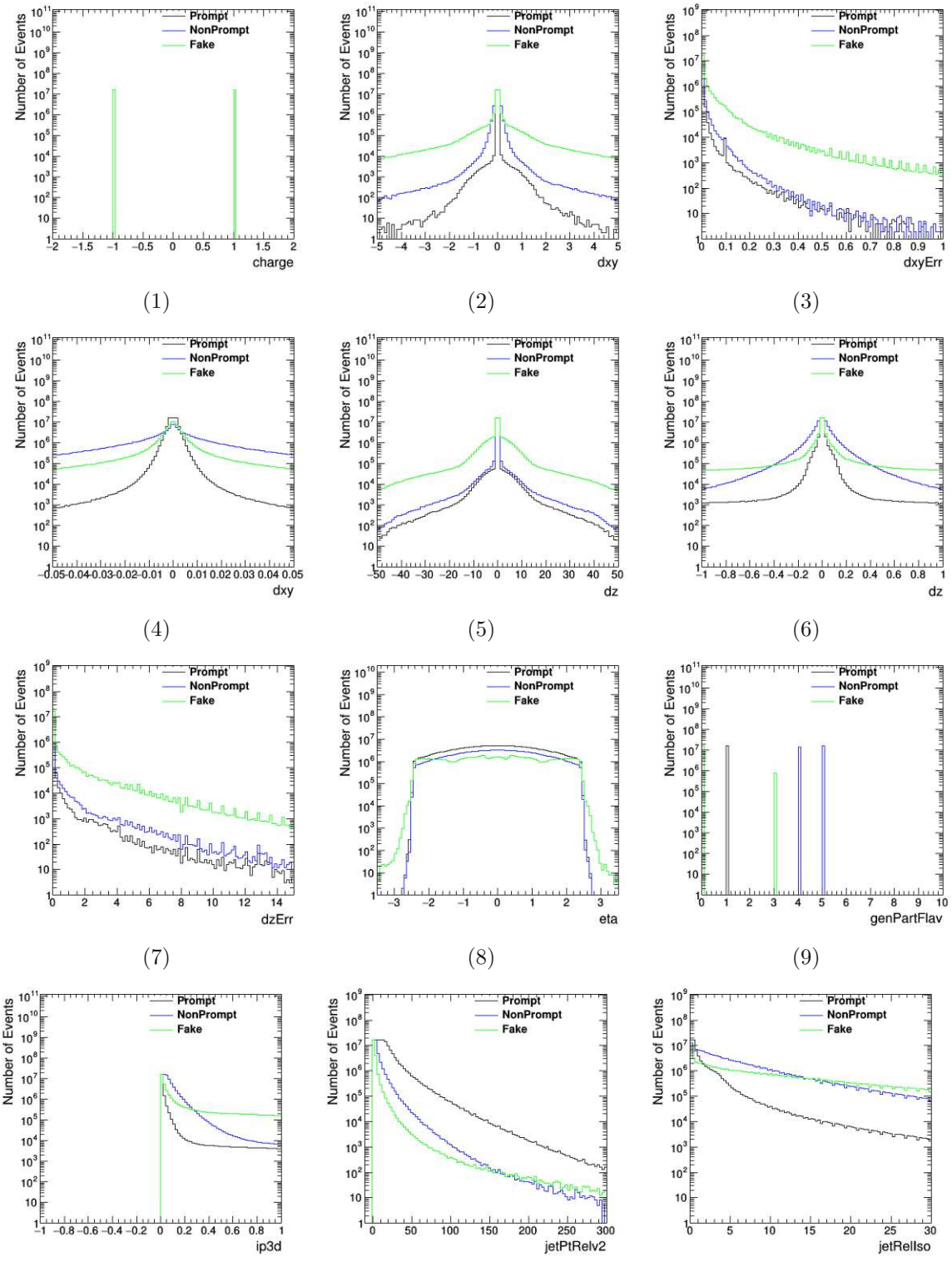


Figure B.37: Input feature plots before conversion 2017 muon $t\bar{t}$ (1/3).

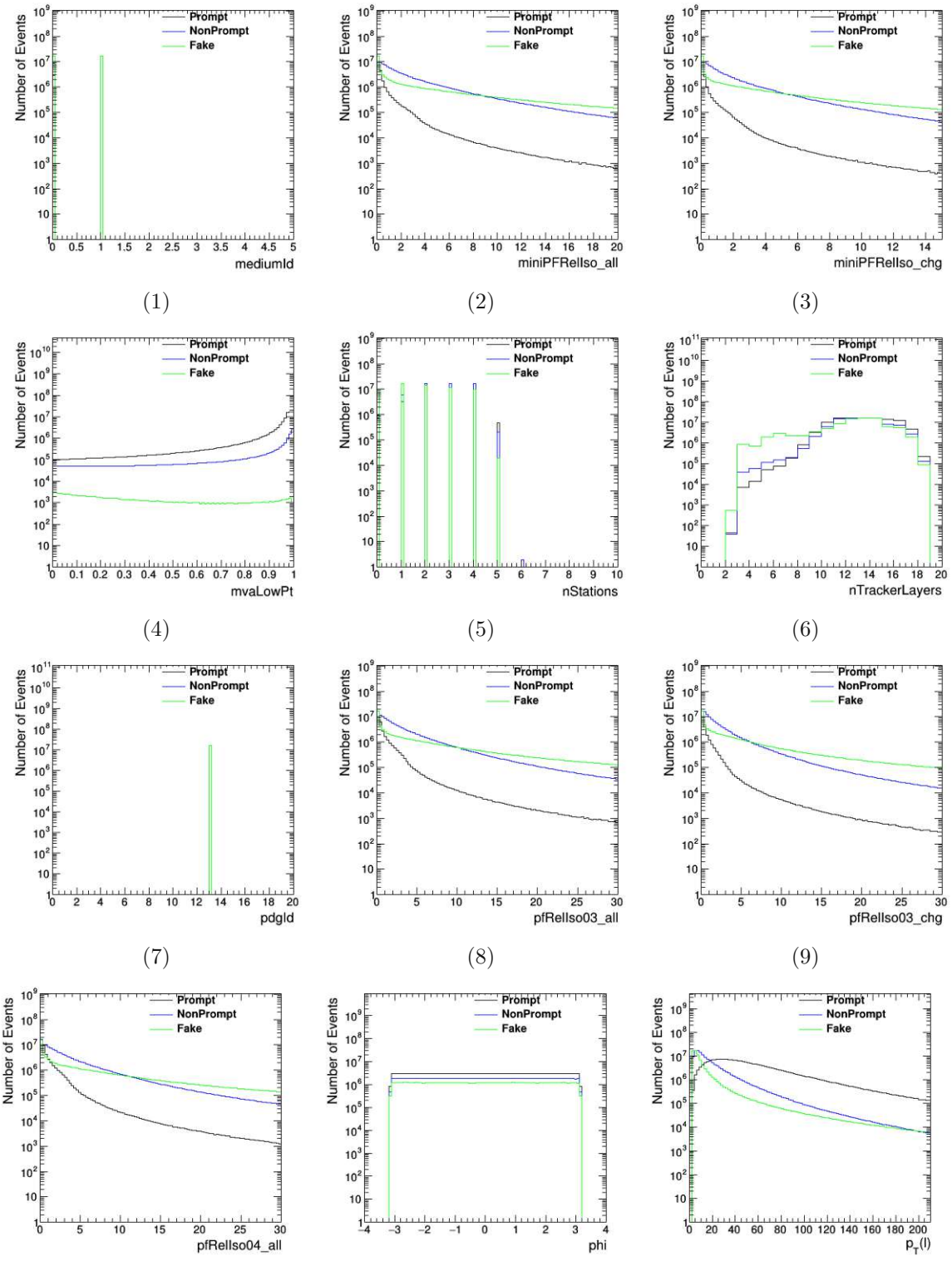


Figure B.38: Input feature plots before conversion 2017 muon $t\bar{t}$ (2/3).

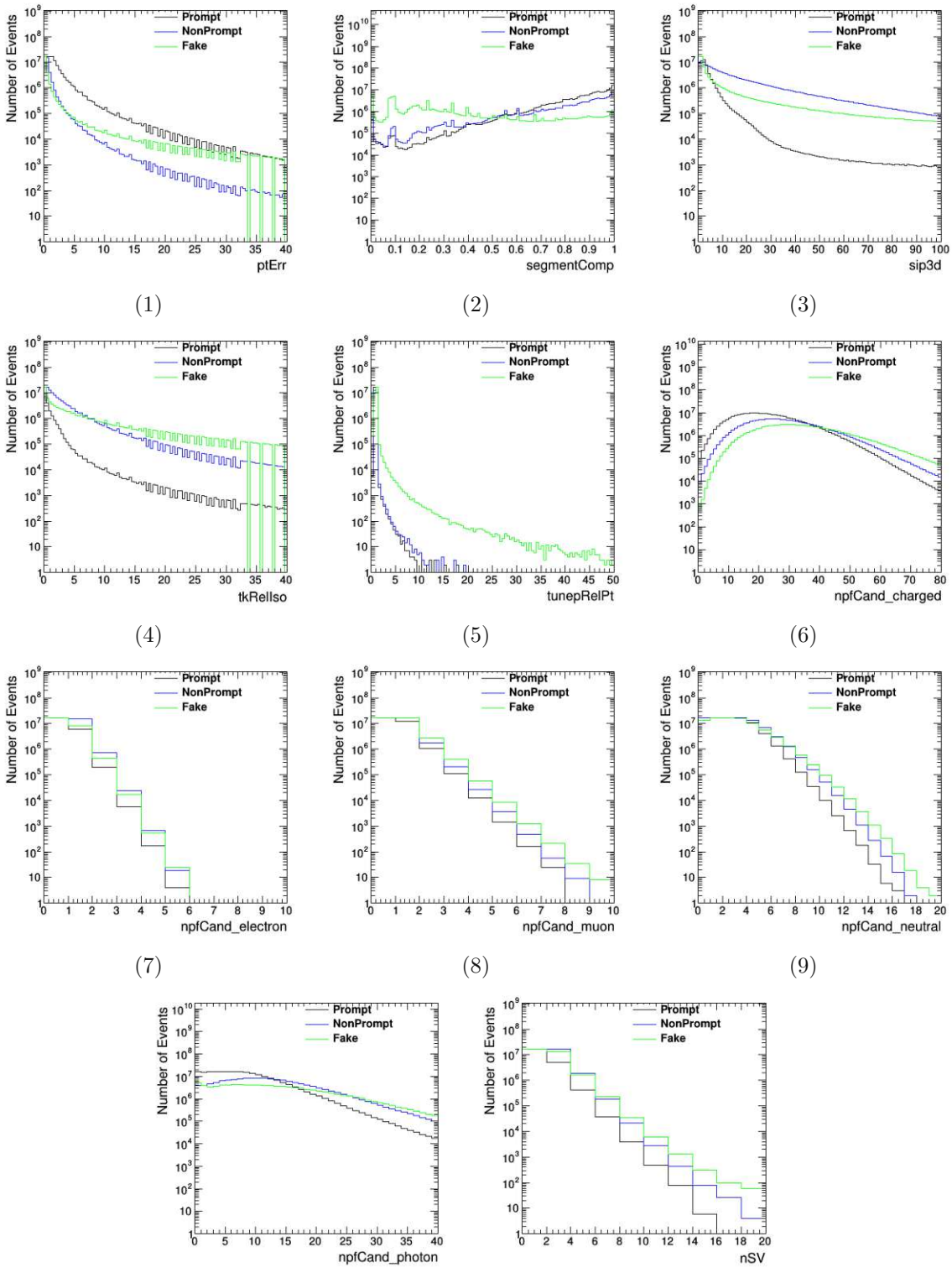


Figure B.39: Input feature plots before conversion 2017 muon $t\bar{t}$ (3/3).

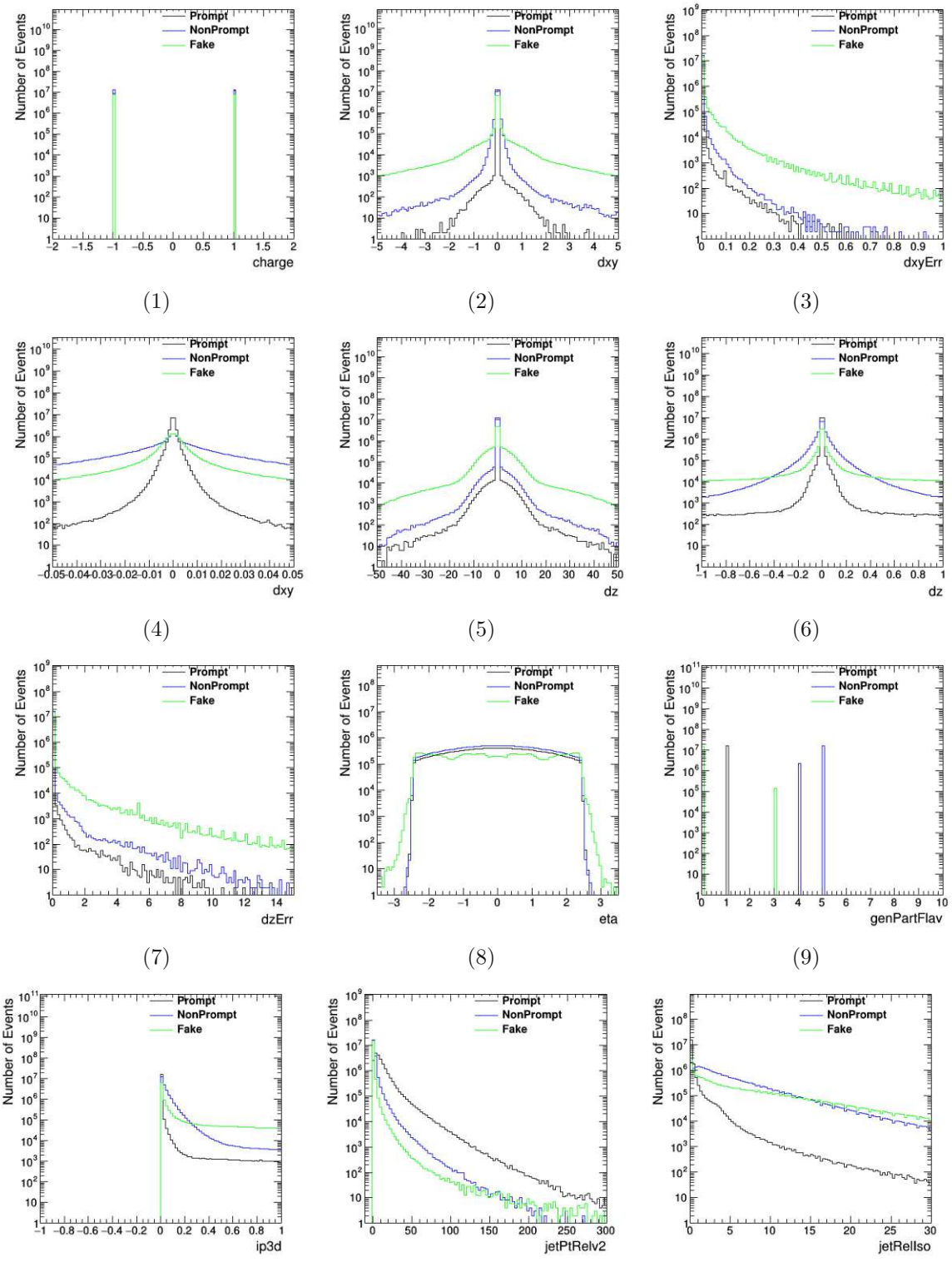


Figure B.40: Input feature plots before conversion 2018 muon $t\bar{t}$ (1/3).

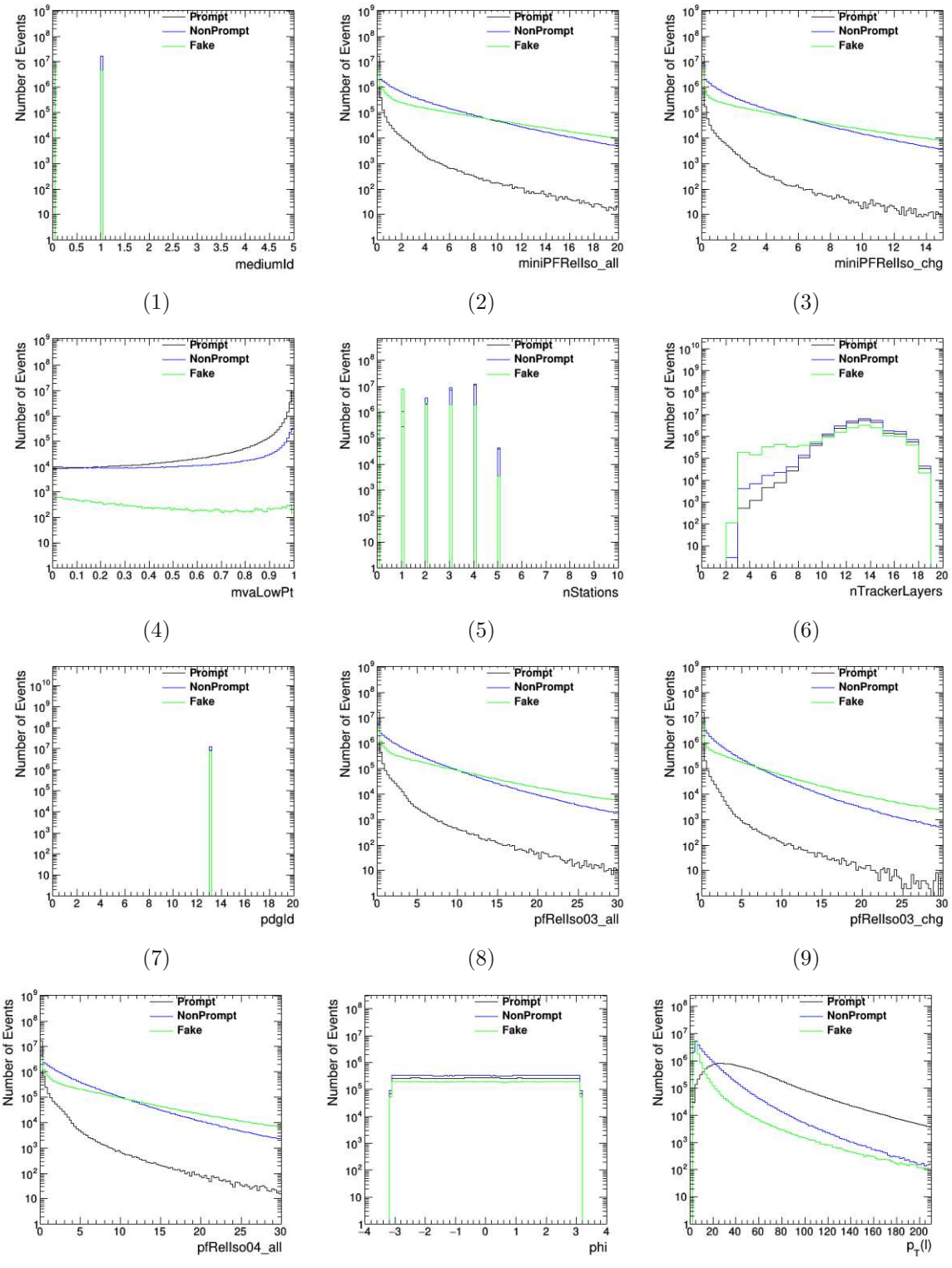
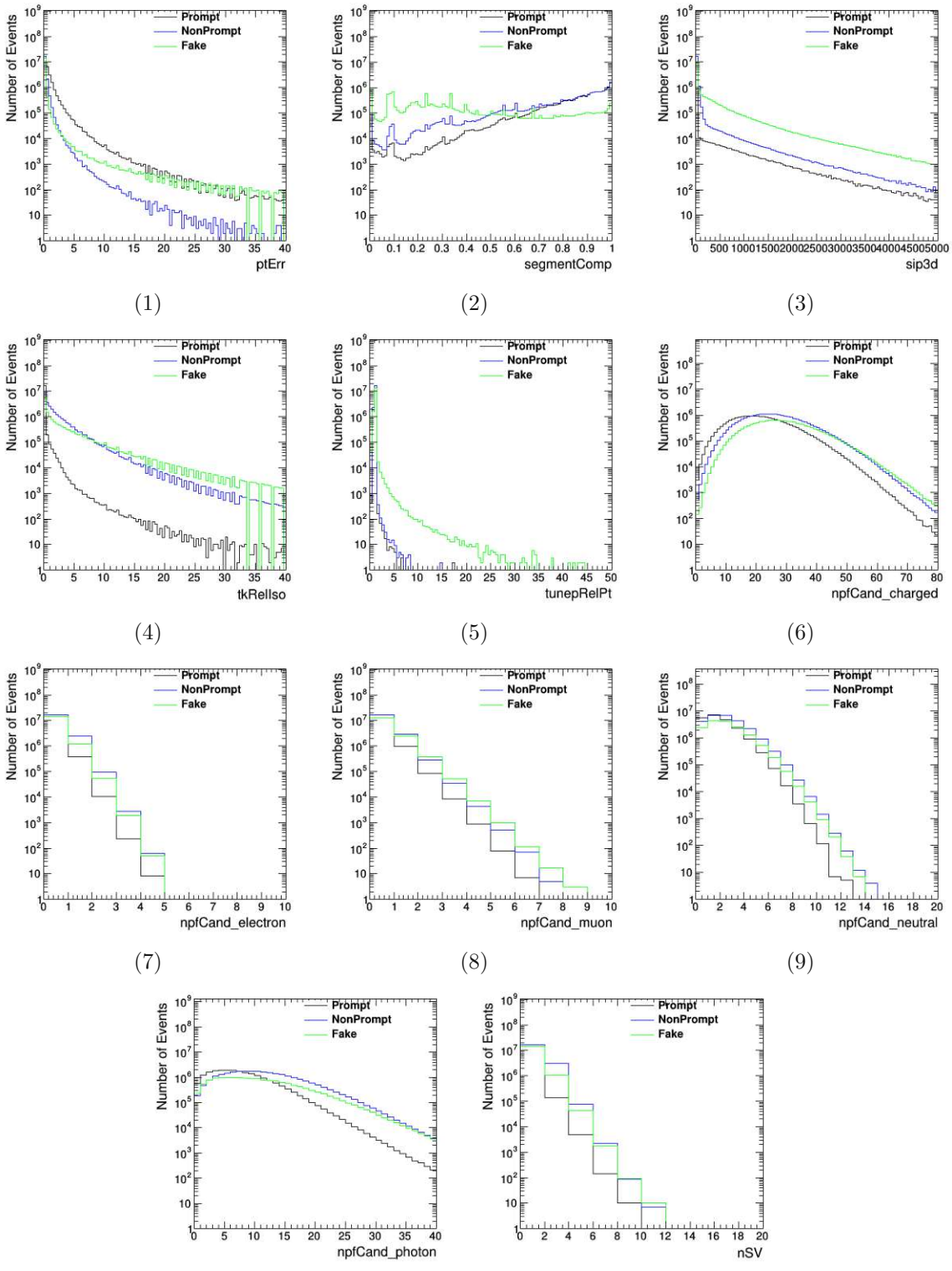


Figure B.41: Input feature plots before conversion 2018 muon $t\bar{t}$ (2/3).



(1) (2) (3)

(4) (5) (6)

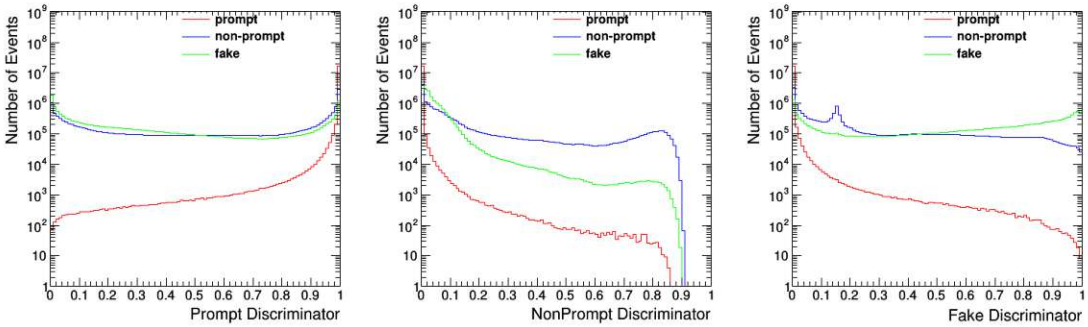
(7) (8) (9)

(10) (11)

Figure B.42: Input feature plots before conversion 2018 muon $t\bar{t}$ (3/3).

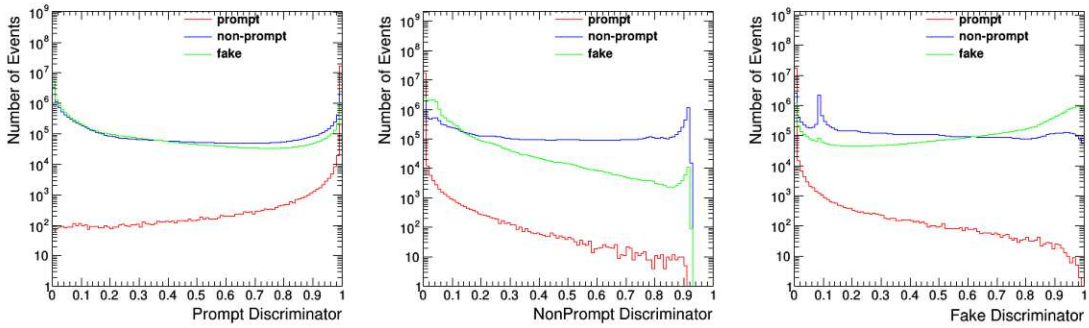
Appendix C

Discriminator histograms



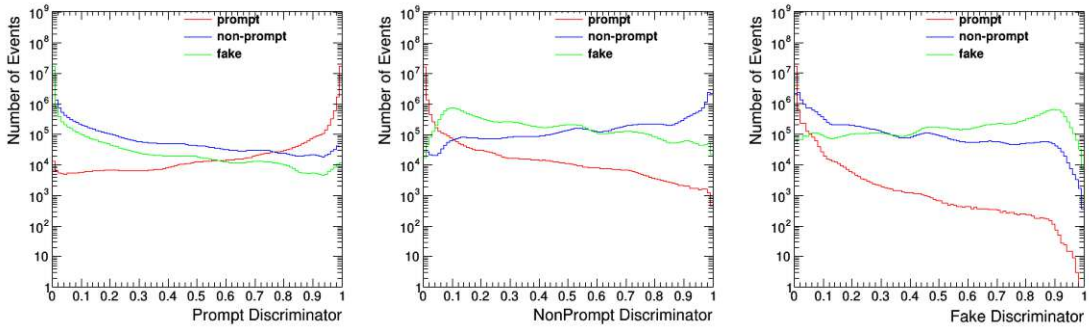
(a) (b) (c)

Figure C.1: Discriminator value histograms of the 2016 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext2* samples and samples in Table A.2. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



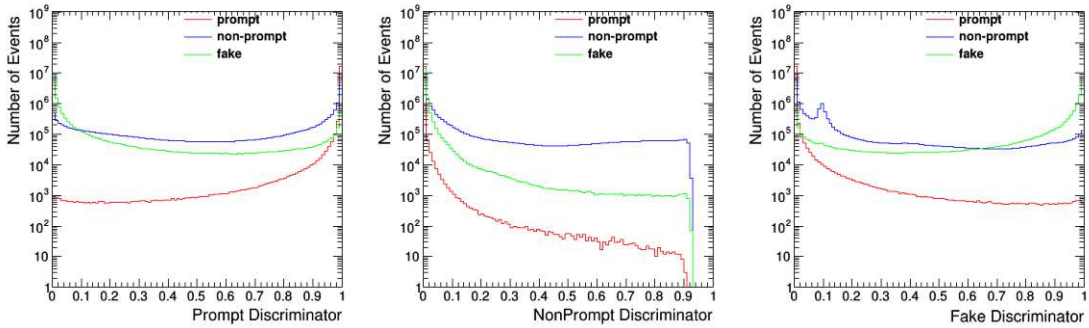
(a) (b) (c)

Figure C.2: Discriminator value histograms of the 2017 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext1* samples and samples in Table A.6. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



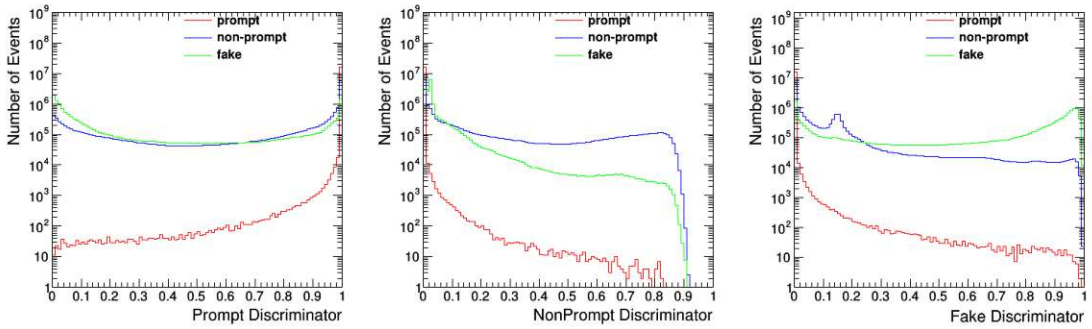
(a) (b) (c)

Figure C.3: Discriminator value histograms of the 2018 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_NLO* samples and samples in Table A.10. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



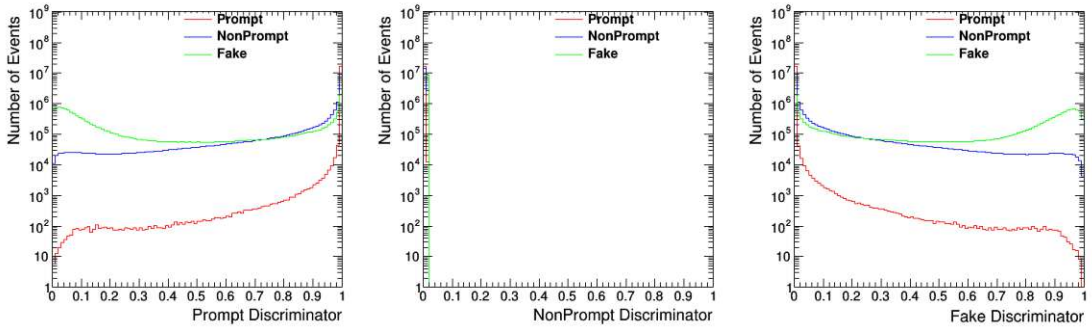
(a) (b) (c)

Figure C.4: Discriminator value histograms of the 2016 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext2* samples and samples in Table A.3. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



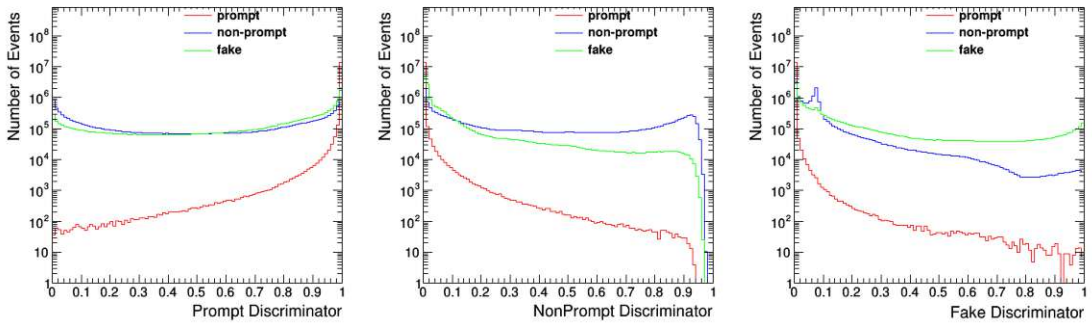
(a) (b) (c)

Figure C.5: Discriminator value histograms of the 2017 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext1* samples and samples in Table A.7. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



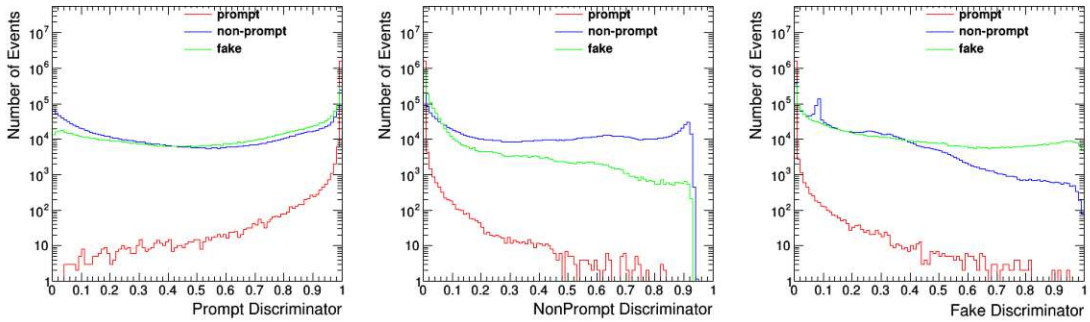
(a) (b) (c)

Figure C.6: Discriminator value histograms of the 2018 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_NLO* samples and samples in Table A.11. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



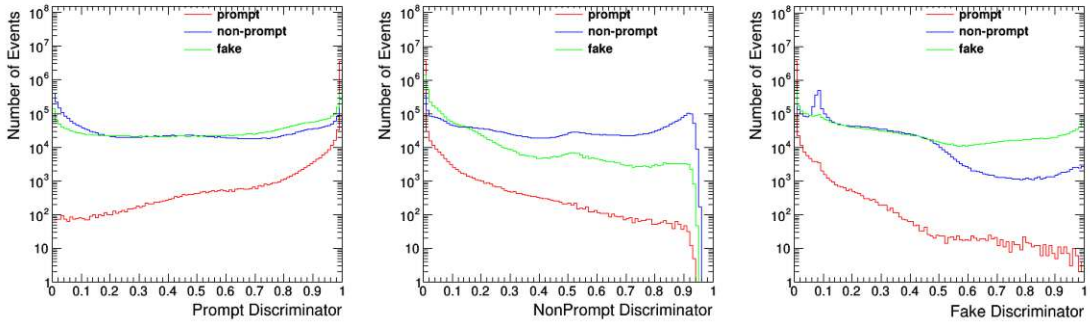
(a) (b) (c)

Figure C.7: Discriminator value histograms of the 2016 muon $t\bar{t}$ training from a prediction on the TT_pow sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



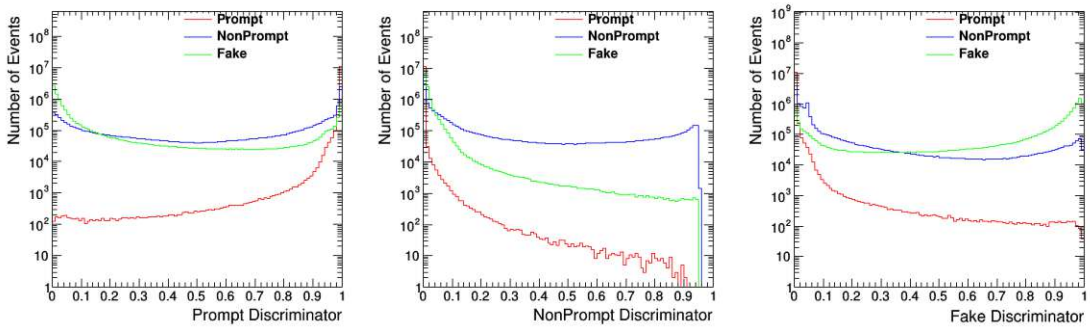
(a) (b) (c)

Figure C.8: Discriminator value histograms of the 2017 muon $t\bar{t}$ training from a prediction on the TTJets_LO sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



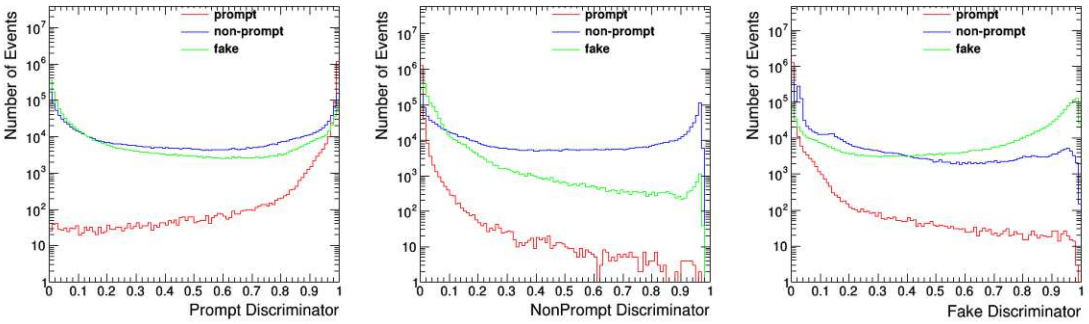
(a) (b) (c)

Figure C.9: Discriminator value histograms of the 2018 muon $t\bar{t}$ training from a prediction on the TTWJets_LO sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



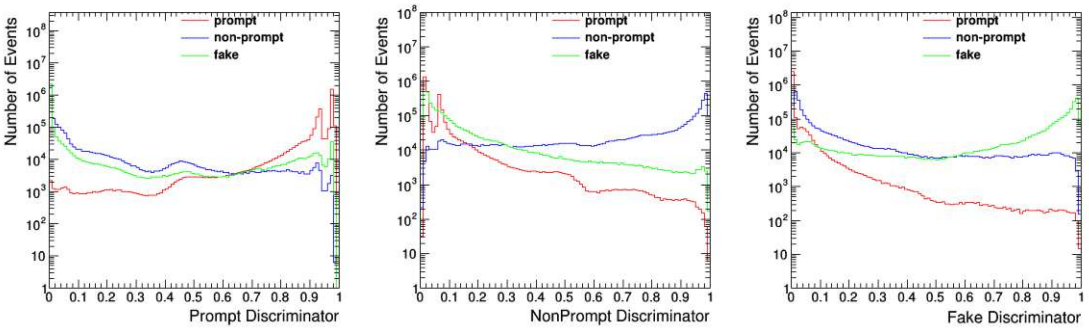
(a) (b) (c)

Figure C.10: Discriminator value histograms of the 2016 electron $t\bar{t}$ training from a prediction on the TT_pow sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



(a) (b) (c)

Figure C.11: Discriminator value histograms of the 2017 electron $t\bar{t}$ training from a prediction on the TTJets_LO sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.



(a) (b) (c)

Figure C.12: Discriminator value histograms of the 2018 electron $t\bar{t}$ training from a prediction on the TTWJets_LO sample. They are shown for the prompt (a) non-prompt (b) and fake (c) classes.

Appendix D

Efficiency

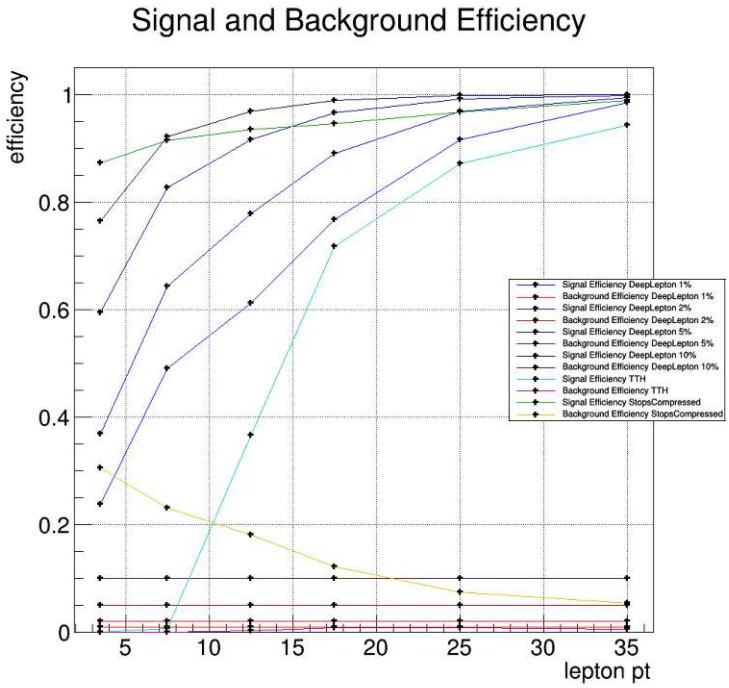


Figure D.1: Efficiency curves of the 2016 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext2* samples and samples in Table A.2.

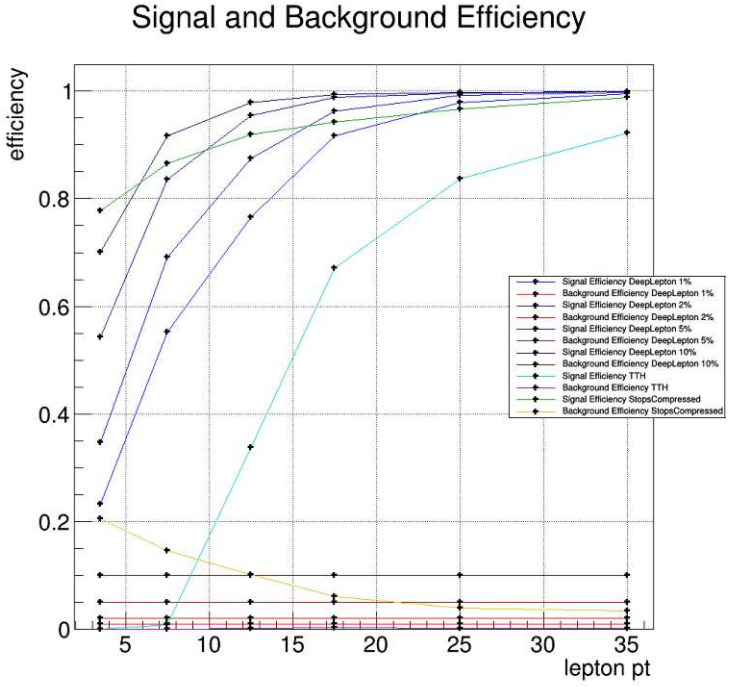


Figure D.2: Efficiency curves of the 2017 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext1* samples and samples in Table A.6.

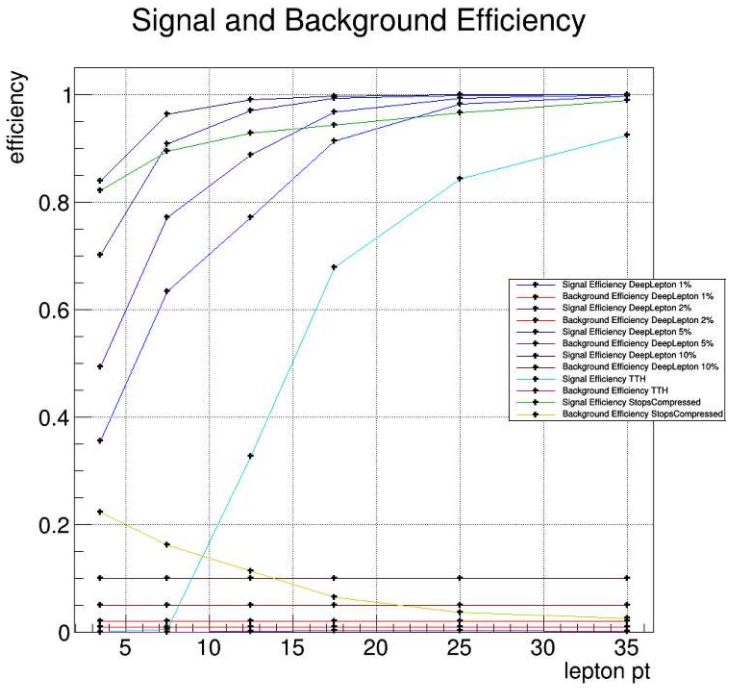


Figure D.3: Efficiency curves of the 2018 muon DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_NLO* samples and samples in Table A.10.

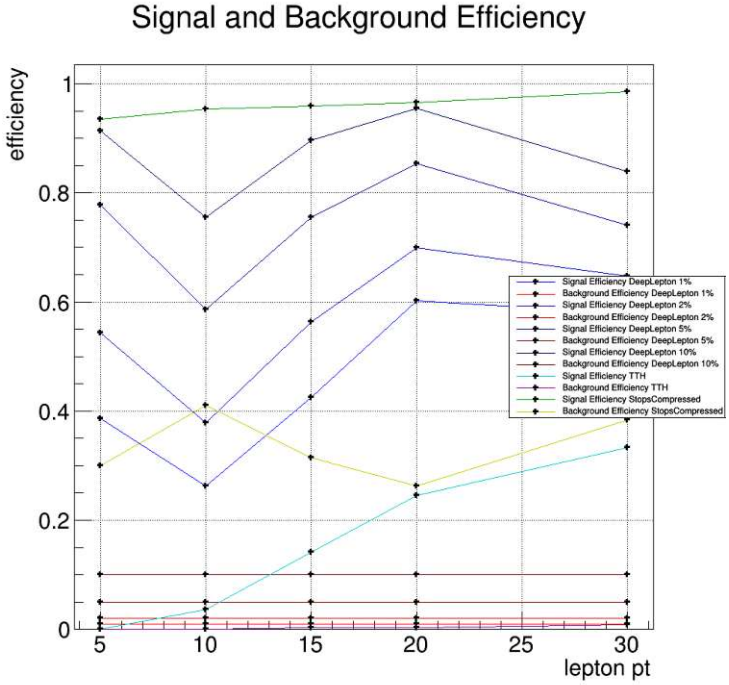


Figure D.4: Efficiency curves of the 2016 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext2* samples and samples in Table A.3.

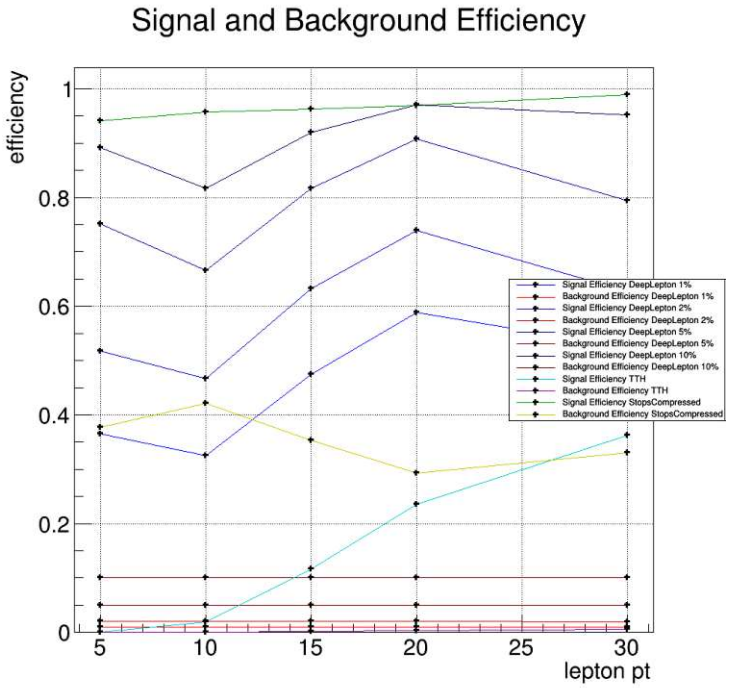


Figure D.5: Efficiency curves of the 2017 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext1* samples and samples in Table A.7.

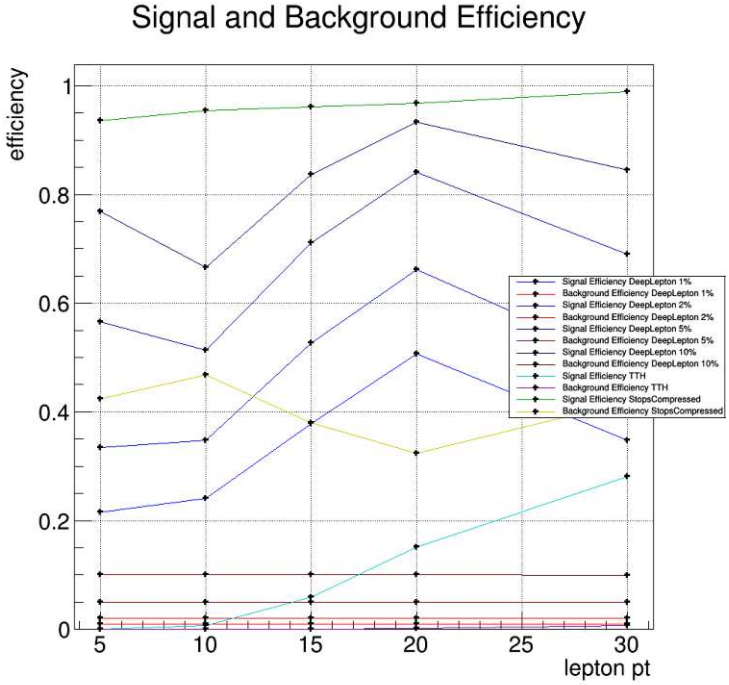


Figure D.6: Efficiency curves of the 2018 electron DYvsQCD training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_NLO* samples and samples in Table A.11.

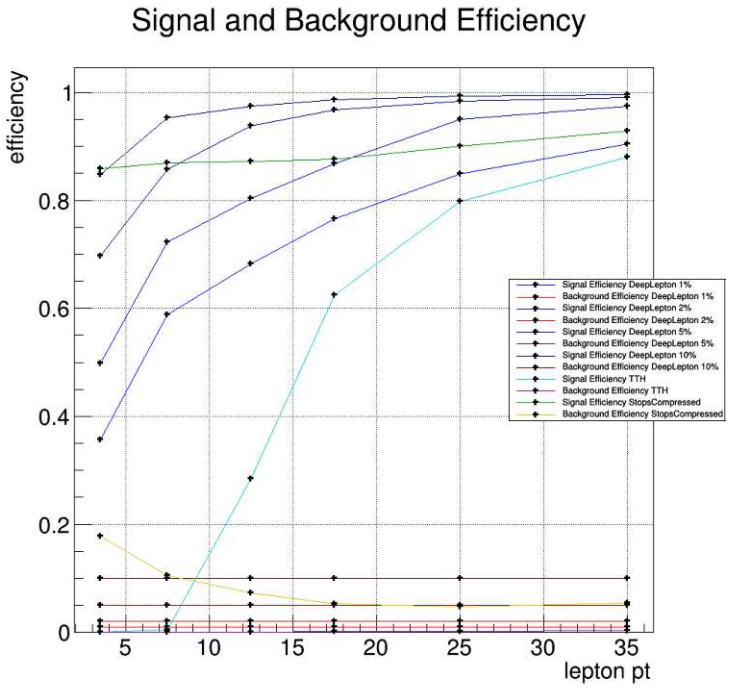


Figure D.7: Efficiency curves of the 2016 muon $t\bar{t}$ training from a prediction on the *TT_pow* sample.

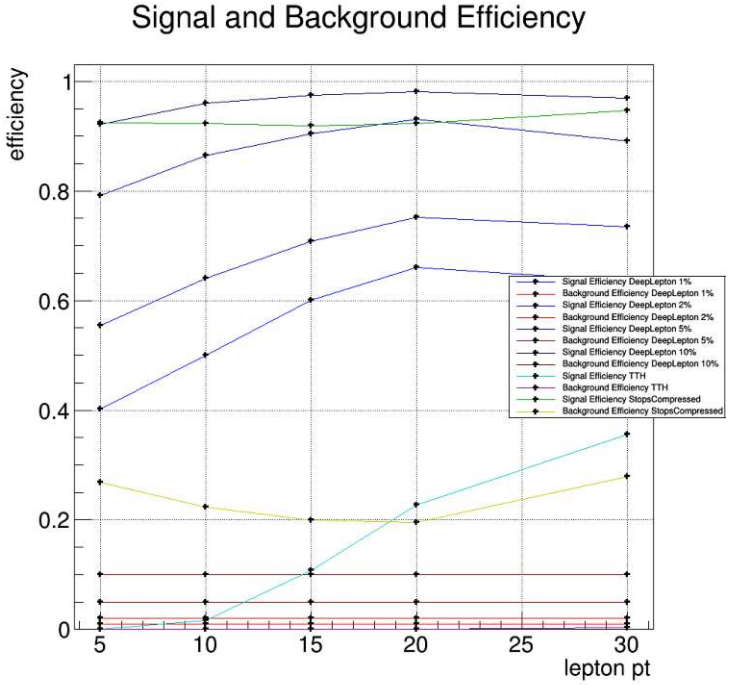


Figure D.8: Efficiency curves of the 2017 muon $t\bar{t}$ training from a prediction on the *TTJets_LO* sample.

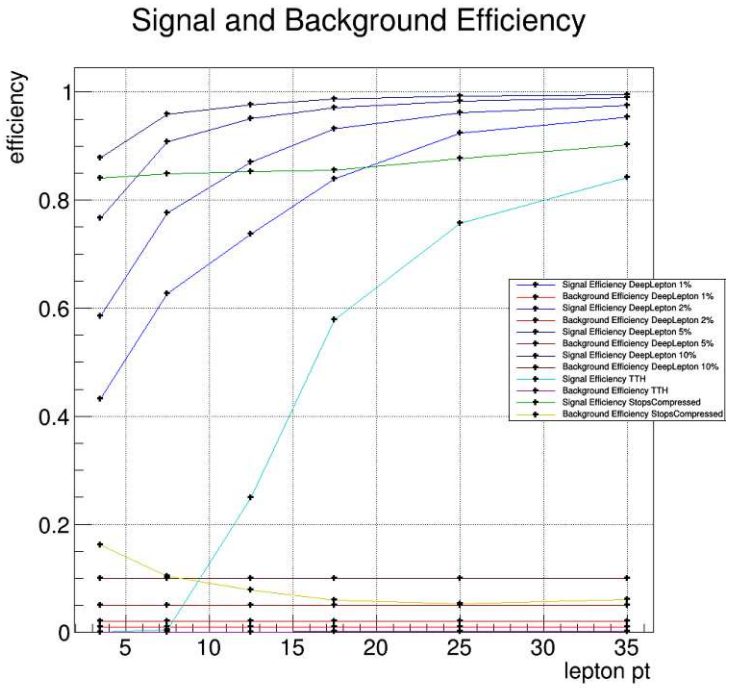


Figure D.9: Efficiency curves of the 2018 muon $t\bar{t}$ training from a prediction on the *TTWJets_LO* sample.

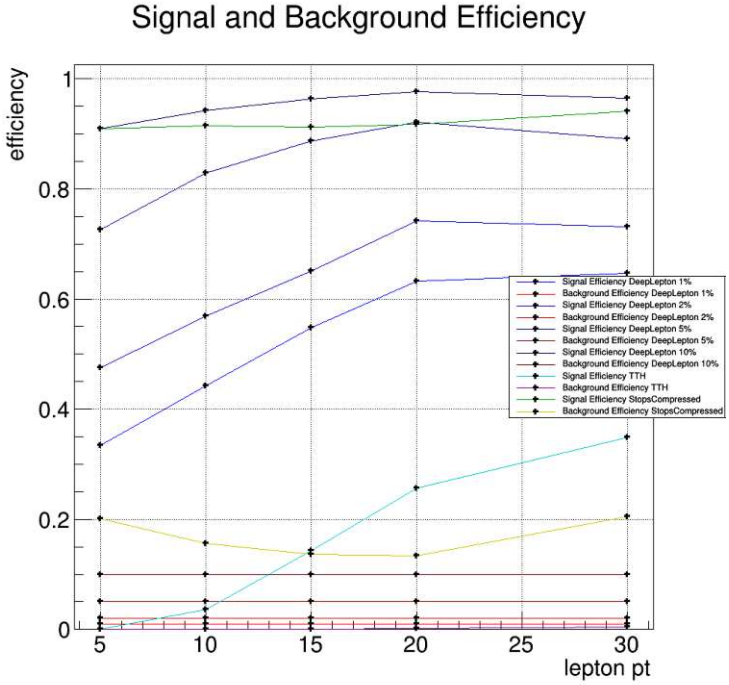


Figure D.10: Efficiency curves of the 2016 electron $t\bar{t}$ training from a prediction on the *DYJetsToLL_M50_LO*, *DYJetsToLL_M50_LO_ext2* samples and samples in Table A.3

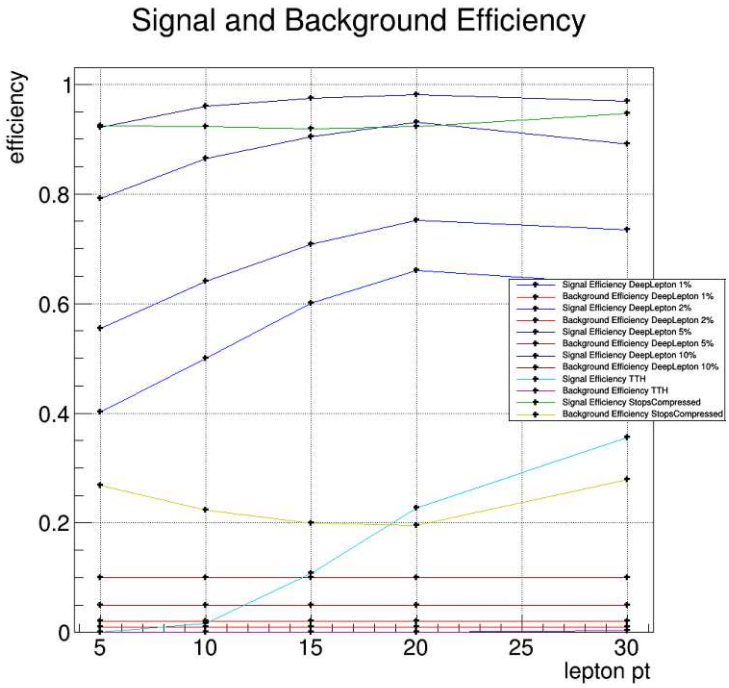


Figure D.11: Efficiency curves of the 2017 electron $t\bar{t}$ training from a prediction on the *TTJets_LO* sample.

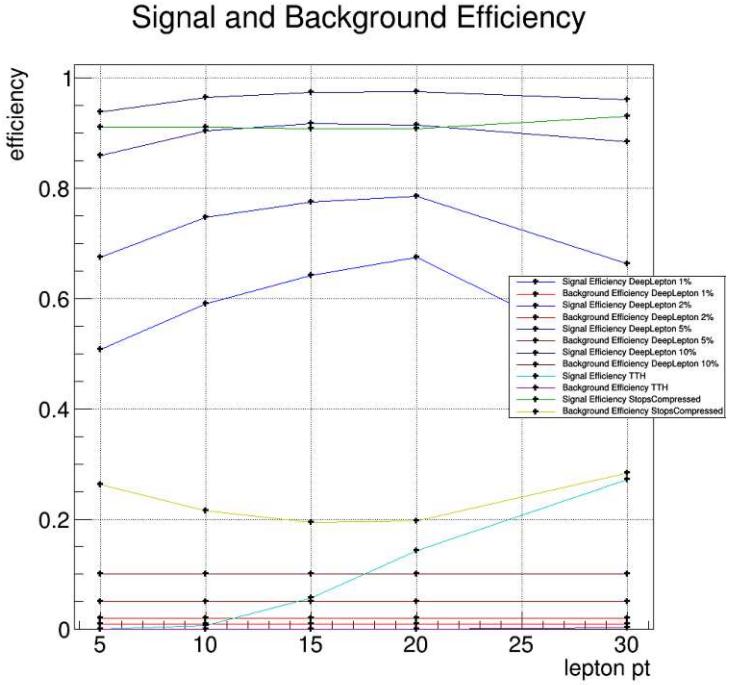


Figure D.12: Efficiency curves of the 2018 electron $t\bar{t}$ training from a prediction on the *TTWJets_LO* sample.

List of Figures

List of Tables

Bibliography

[1] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008. doi: 10.1088/1748-0221/3/08/s08004.

[2] Oliver S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC Design Report Vol.1: The LHC Main Ring. 6 2004. doi: 10.5170/CERN-2004-003-V-1.

[3] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural computation*, 9(8):1735–1780, 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735.

[4] Valerio Bertacchi. DeepCore: Convolutional Neural Network for high p_T jet tracking, 2020.

[5] Georg Mörtl. DeepLepton: muon identification in 13 TeV pp collisions at the CMS experiment using deep learning techniques. Master's thesis, 2019.

[6] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.

[7] CERN-Website. The standard model. URL `https://home.cern/science/physics/standard-model`. Visited on 2021-07-29.

[8] S. Beole et al. ALICE technical design report: Detector for high momentum PID. 8 1998.

[9] LHCbCollaboration. LHCb magnet: Technical design report. 2000.

[10] L Arnaudon, P Baudrenghien, M Baylac, G Bellodi, Y Body, J Borburgh, P Bourquin, J Broere, O Brunner, L Bruno, C Carli, Friedhelm Caspers, S M Cousineau, Y Cuvet, C De Almeida Martins, T Dobers, T Fowler, R Garoby, F Gerigk, B Goddard, K Hanke, M Hori, M Jones, K Kahle, Willi Kalbreier, T Kroyer, D Küchler, A M Lombardi, L A López-Hernandez, M Magistris, M Martini, S Maury, E Page, M Paoluzzi, M Pasini, U Raich, C Rossi, J P Royer, E Sargsyan, J Serrano, R Scrivens, M Silari, M Timmins, W Venturini-Delsolaro, M Vretenar, R Wegner, W Weterings, and T Zickler. Linac4 Technical Design Report. Technical report, CERN, Geneva, 2006. URL `https://cds.cern.ch/record/1004186`. revised version submitted on 2006-12-14 09:00:40.

[11] CERN-Website. The Proton Synchrotron Booster, . URL `https://home.cern/science/accelerators/proton-synchrotron-booster`. Visited on 2021-11-19.

[12] CERN-Website. The Proton Synchrotron, . URL `https://home.cern/science/accelerators/proton-synchrotron`. Visited on 2021-11-19.

[13] CERN-Website. The Super Proton Synchrotron, . URL `https://home.cern/science/accelerators/super-proton-synchrotron`. Visited on 2021-11-19.

[14] Esma Mobs. The CERN accelerator complex - 2019. URL `https://cds.cern.ch/images/CERN-GRAPHICS-2019-002-1`. Visited on 2021-09-25.

[15] N. Siegel. Status of the Large Hadron Collider and magnet program. *IEEE Trans. Appl. Supercond.*, 7(2):252–257, 1997. doi: 10.1109/77.614478.

[16] Linda Stoel, Michael Barnes, Wolfgang Bartmann, F Burkart, Bethany Goddard, W Herr, Thomas Kramer, A Milanese, G Rumolo, Elena Shaposhnikova, and Geneva Switzerland. HIGH ENERGY BOOSTER OPTIONS FOR A FUTURE CIRCULAR COLLIDER AT CERN. 2016.

[17] D. Boussard and T. Linnecar. THE LHC SUPERCONDUCTING RF SYSTEM.

[18] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2013.

[19] Andreas Hoecker. Physics at the LHC Run-2 and Beyond, 2016.

[20] 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. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.08.020.

[21] CERN-Website. The higgs boson, . URL `https://home.cern/science/physics/higgs-boson`. Visited on 2021-07-29.

[22] Piotr Traczyk. 59 new hadrons and counting. URL `https://home.cern/news/news/physics/59-new-hadrons-and-counting`. Visited on 2021-03-03.

[23] CERN-Website. Heavy ions and quark-gluon plasma, . URL `https://home.cern/science/physics/heavy-ions-and-quark-gluon-plasma`. Visited on 2021-07-29.

[24] Peter GO Freund and Peter George Oliver Freund. *Introduction to supersymmetry*. Cambridge University Press, 1988.

[25] Neta A. Bahcall, Lori M. Lubin, and Victoria Dorman. Where Is the Dark Matter? *The Astrophysical Journal*, 447(2), jul 1995. doi: 10.1086/309577.

[26] Daniel Abercrombie, Nural Akchurin, Ece Akilli, Juan Alcaraz Maestre, Brandon Allen, Barbara Alvarez Gonzalez, Jeremy Andrea, Alexandre Arbey, Georges Azuelos, Patrizia Azzi, and et al. Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. *Physics of the Dark Universe*, 27:100371, Jan 2020. ISSN 2212-6864. doi: 10.1016/j.dark.2019.100371.

[27] CERN-Website. Dark matter, . URL `https://home.cern/science/physics/dark-matter`. Visited on 2021-07-29.

[28] CERN-Website. The matter-antimatter asymmetry problem, . URL `https://home.cern/science/physics/matter-antimatter-asymmetry-problem`. Visited on 2021-07-30.

[29] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[30] CMS. Cms detector. URL `https://cms.cern/detector`. Visited on 2020-11-28.

[31] The CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[32] CMS-Website. Bending Particles, . URL `https://cms.cern/detector/bending-particles`. Visited on 2021-08-16.

[33] CMS-Website. Tracking, . URL `https://cms.cern/detector/identifying-tracks`. Visited on 2021-08-06.

[34] CMS-Website. Silicon pixels, . URL `https://cms.cern/detector/identifying-tracks/silicon-pixels`. Visited on 2021-08-06.

[35] CMS-Website. Silicon strips, . URL `https://cms.cern/detector/identifying-tracks/silicon-strips`. Visited on 2021-08-06.

[36] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[37] CMs-Website. Energy of Electrons and Photons (ECAL). URL `https://cms.cern/detector/measuring-energy/energy-electrons-and-photons-ecal`. Visited on 2021-08-16.

[38] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006. URL `https://cds.cern.ch/record/922757`. There is an error on cover due to a technical problem for some items.

[39] The CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[40] CMS-Website. Energy of Hadrons (HCAL), . URL `https://cms.cern/detector/measuring-energy/energy-hadrons-hcal`. Visited on 2021-08-16.

[41] CMS-Website. Detecting muons, . URL `https://cms.cern/detector/detecting-muons`. Visited on 2021-08-16.

[42] CMS-Website. Muon drift tubes, . URL `https://cms.cern/detector/detecting-muons/muon-drift-tubes`. Visited on 2021-08-16.

[43] CMS-Website. Cathode strip chambers, . URL `https://cms.cern/detector/detecting-muons/cathode-strip-chambers`. Visited on 2021-08-16.

[44] CMS-Website. Resistive plate chambers, . URL `https://cms.cern/detector/detecting-muons/resistive-plate-chambers`. Visited on 2021-08-16.

[45] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical design report. CMS. CERN, Geneva, 2002.

[46] CMS-Website. Triggering and data acquisition, . URL `https://cms.cern/detector/triggering-and-data-acquisition`. Visited on 2021-08-16.

[47] The CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, oct 2017. doi: 10.1088/1748-0221/12/10/p10003.

[48] The CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$= 8 TeV. *Journal of Instrumentation*, 10 (06):P06005–P06005, jun 2015. doi: 10.1088/1748-0221/10/06/p06005.

[49] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. ISSN 0168-9002. doi: 10.1016/0168-9002(87)90887-4.

[50] The CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, jun 2018. doi: 10.1088/1748-0221/13/06/p06015.

[51] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008. doi: 10.1088/1126-6708/2008/04/063.

[52] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Höche, Hendrik Hoeth, Frank Krauss, Leif Lönnblad, Emily Nurse, Peter Richardson, Steffen Schumann, Michael H. Seymour, Torbjörn Sjöstrand, Peter Skands, and Bryan Webber. General-purpose event generators for LHC physics. *Physics Reports*, 504(5):145–233, 2011. ISSN 0370-1573. doi: 10.1016/j.physrep.2011.03.005.

[53] Michael H. Seymour and Marilyn Marx. Monte Carlo Event Generators, 2013.

[54] PYTHIA Website. Pythia website. URL `https://pythia.org/`. Visited on 2021-08-23.

[55] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, 2015. ISSN 0010-4655. doi: 10.1016/j.cpc.2015.01.024.

[56] C. Oleari. The POWHEG BOX. *Nuclear Physics B - Proceedings Supplements*, 205-206: 36–41, 2010. ISSN 0920-5632. doi: 10.1016/j.nuclphysbps.2010.08.016.

[57] Simone Alioli, Keith Hamilton, Paolo Nason, Carlo Oleari, Emanuele Re, Giulia Zanderighi, and Tomáš Ježo. The POWHEG BOX. URL `http://powhegbox.mib.infn.it/`. Visited on 2021-08-23.

[58] The Herwig collaboration. The Herwig Event Generator. URL `https://herwig.hepforge.org/`. Visited on 2021-08-23.

[59] Stefano Moretti. HERWIG: an event generator for MSSM processes. *Arxiv preprint hep-ph/0205105*, 2002.

[60] MADGraph Website. The MadGraph5_aMC@NLO homepage. URL `http://madgraph.physics.illinois.edu/`. Visited on 2021-08-23.

[61] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014. ISSN 1029-8479. doi: 10.1007/jhep07(2014)079.

[62] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8.

[63] Md Zahangir Alom, Tarek M. Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Esesn, Abdul A S. Awwal, and Vijayan K. Asari. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches, 2018.

[64] Sébastien Meunier. The Impacts and Challenges of Artificial Intelligence in Finance. URL `https://internationalbanker.com/finance/the-impacts-and-challenges-of-artificial-intelligence-in-finance/`. Visited on 2021-07-24.

[65] Harris Drucker and Corinna Cortes. Boosting Decision Trees. *Advances in Neural Information Processing Systems*, 8:479–485, 01 1995.

[66] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, 1992. doi: 10.1145/130385.130401.

[67] Frank Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. pages 183–190, 2021. doi: 10.7551/mitpress/12274.003.0020.

[68] Geeks for Geeks. Clustering in machine learning. URL `https://www.geeksforgeeks.org/clustering-in-machine-learning/`. Visited on 2021-07-24.

[69] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[70] Douglas Reynolds. *Gaussian Mixture Models*, pages 659–663. Springer US, Boston, MA, 2009. ISBN 978-0-387-73003-5. doi: 10.1007/978-0-387-73003-5_196. URL `https://doi.org/10.1007/978-0-387-73003-5_196`.

[71] Blazej Ozinski and Konrad Budek. What is reinforcement learning? The complete guide. URL `https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/`. Visited on 2021-07-24.

[72] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[73] François Chollet et al. Keras. `https://keras.io`, 2015.

[74] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In *2014 Science and Information Conference*, pages 372–378, 2014. doi: 10.1109/SAI.2014.6918213.

[75] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *CoRR*, abs/1512.03385, 2015. doi: 10.1109/cvpr.2016.90.

[76] Aravind Pai. CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning. URL `https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/`.

[77] Arjun Kashyap. Math behind perceptrons. URL `https://medium.com/@iamask09/math-behind-perceptrons-7241d5dadbfc`. Visited on 2021-07-25.

[78] Daniel H. Mathematical representation of a perceptron layer (with example in tensorflow). URL https://medium.com/@daniel.hellwig.p/mathematical-representation-of-a-perceptron-layer-with-example-in-tensorflow-754a38833b44. Visited on 2021-07-25.

[79] Aniruddha Karajgi. How Neural Networks Solve the XOR Problem. URL https://towardsdatascience.com/how-neural-networks-solve-the-xor-problem-59763136bdd7. Visited on 2021-07-25.

[80] Matthew Stewart. Simple introduction to convolutional neural networks. URL https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac.

[81] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *CoRR*, abs/1808.03314:132306, 2018. ISSN 0167-2789. doi: 10.1016/j.physd.2019.132306.

[82] Denny Britz. Recurrent neural networks tutorial, part 1 – introduction to rnns. URL http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/. Visited on 2021-07-26.

[83] Hemanth Pedamallu. Rnn vs gru vslstm. URL https://medium.com/analytics-vidhya/rnn-vs-gru-vs-lstm-863b0b7b1573. Visited on 2021-07-23.

[84] What is the architecture behind the Keras LSTM cell? URL https://stackoverflow.com/questions/50488427/what-is-the-architecture-behind-the-keras-lstm-cell. Visited on 2021-07-26.

[85] Christopher Olah. Understanding lstm networks. URL https://colah.github.io/posts/2015-08-Understanding-LSTMs/. Visited on 2021-07-26.

[86] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR*, abs/1406.1078, 2014. doi: 10.3115/v1/d14-1179.

[87] KERAS. Keras bidirectional layer, . URL https://keras.io/api/layers/recurrent_layers/bidirectional/. Visited on 2021-07-22.

[88] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997. ISSN 1053-587X. doi: 10.1109/78.650093.

[89] Christian Szegedy Sergey Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[90] KERAS. Keras batch normalisation, . URL https://keras.io/api/layers/normalization_layers/batch_normalization/. Visited on 2020-11-26.

[91] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[92] Alex Krizhevsky Nitish Srivastava, Georey Hinton. Dropout: A Simple Way to Prevent Neural Networks from Overtting. 60:84–90, 2017. ISSN 0001-0782. doi: 10.1145/3065386.

[93] Sagar Sharma and Simone Sharma. Activation functions in neural networks. *Towards Data Science*, 6(12):310–316, 2017. ISSN 2455-2143. doi: 10.33564/ijeast.2020.v04i12.054.

[94] Thomas Wood. What is the Softmax Function? URL `https://deepai.org/machine-learning-glossary-and-terms/softmax-layer`. Visited on 2021-07-27.

[95] KERAS. Softmax Function, . URL `https://keras.io/api/layers/activations/#softmax-function`. Visited on 2021-07-27.

[96] KERAS. Regression losses, . URL `https://keras.io/api/losses/regression_losses/`. Visited on 2021-07-27.

[97] Jason Brownlee. Gradient Descent For Machine Learning. URL `https://machinelearningmastery.com/gradient-descent-for-machine-learning/`. Visited on 2021-07-27.

[98] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[99] Seppo Linnainmaa. Taylor expansion of the accumulated rounding error. *BIT Numerical Mathematics*, 16(2):146–160, 1976. ISSN 0006-3835. doi: 10.1007/bf01931367.

[100] Bohra Yash. The Challenge of Vanishing/Exploding Gradients in Deep Neural Networks. URL `https://www.analyticsvidhya.com/blog/2021/06/the-challenge-of-vanishing-exploding-gradients-in-deep-neural-networks/`. Visited on 2021-07-27.

[101] Sepp Hochreiter. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998. ISSN 0218-4885. doi: 10.1142/s0218488598000094.

[102] Hong Hui Tan and King Hann Lim. Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization. In *2019 7th International Conference on Smart Computing Communications (ICSCC)*, pages 1–4, 2019. doi: 10.1109/ICSCC.2019.8843652.

[103] Sidney D. Drell and Tung-Mow Yan. Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies. *Phys. Rev. Lett.*, 25:316–320, Aug 1970. doi: 10.1103/PhysRevLett.25.316.

[104] Particle Data Group. The Review of Particle Physics (2021). URL `https://pdg.lbl.gov/`. Visited on 2022-01-06.

[105] Krasny, M., Fayette, F., Płaczek, W. et al. Z boson as "the standard candle" for high-precision W boson physics at LHC. doi: 10.1140/epjc/s10052-007-0321-8.

[106] M. JINDAL, BOURILKOV D., and K. MAZUMDAR. Drell–Yan process at Large Hadron Collider. doi: 10.1007/s12043-011-0014-6.

[107] The CMS Collaboration. Measurement of the Drell-Yan cross section in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of High Energy Physics*, 2011(10), Oct 2011. ISSN 1029-8479. doi: 10.1007/jhep10(2011)007.

[108] Jakob Salfeld-Nebgen. Dimuon invariant mass spectrum reconstructed in the High Level Trigger system of the CMS detector for data collected in 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTDiMuon2017and2018`.

[109] Nikolaos Kidonakis. Top quark production, 2013.

[110] Fabian Kohn. Measurement of the charge asymmetry in top quark pair production in pp collision data at sqrt(s) = 7 TeV using the ATLAS detector, 2012.

[111] D0 Collaboration. Observation of the top quark. *Physical Review Letters*, 74(14):2632–2637, 1995. ISSN 1079-7114. doi: 10.1103/physrevlett.74.2632. URL http://dx.doi.org/10.1103/PhysRevLett.74.2632.

[112] Dhiman Chakraborty, Jacobo Konigsberg, and David Rainwater. TOP-QUARK PHYSICS. *Annual Review of Nuclear and Particle Science*, 53(1):301–351, 2003. doi: 10.1146/annurev.nucl.53.041002.110601.

[113] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963. doi: 10.1103/PhysRevLett.10.531.

[114] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973. ISSN 0033-068X. doi: 10.1143/PTP.49.652.

[115] Y. Sakai A. Ceccucci, Z. Ligeti. CKM Quark-Mixing Matrix. URL https://pdg.lbl.gov/2020/reviews/rpp2020-rev-ckm-matrix.pdf.

[116] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, 1970. doi: 10.1103/PhysRevD.2.1285. URL https://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[117] The ATLAS collaboration. Search for anomalous production of prompt same-sign lepton pairs and pair-produced doubly charged Higgs bosons with $\sqrt{s}$=8 TeV pp collisions using the ATLAS detector. doi: 10.1007/JHEP03(2015)041.

[118] Baffioni, S., Charlot, C., Ferri, F. et al. Electron reconstruction in CMS. doi: 10.1140/epjc/s10052-006-0175-5.

[119] Thomas P. S. Gillam. *Identifying fake leptons in ATLAS while hunting SUSY in 8 TeV proton-proton collisions*. PhD thesis.

[120] The CMS Collaboration. Search for top squarks decaying via four-body or chargino-mediated modes in single-lepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Journal of High Energy Physics*, 2018(9), Sep 2018. ISSN 1029-8479. doi: 10.1007/jhep09(2018)065.

[121] The CMS Collaboration. Observation of $t\bar{t}H$ Production. *Physical Review Letters*, 120 (23), Jun 2018. ISSN 1079-7114. doi: 10.1103/physrevlett.120.231801.

[122] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable Parallel Programming with CUDA: Is CUDA the Parallel Programming Model That Application Developers Have Been Waiting For? *Queue*, 6(2):40–53, March 2008. ISSN 1542-7730. doi: 10.1145/1365490.1365500.

[123] S. Harper and C. Shepherd-Themistocleous. Improving sigma eta eta. URL https://indico.cern.ch/event/27560/contributions/1618477/attachments/499353/689807/sigmaEtaEta_sharper.pdf.pdf. Visited on 2021-12-03.

[124] Chris Seez. Shower shape variables. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEgammaShowerShape. Visited on 2022-01-07.

[125] Giovanni Petrucciani Cristina Botta. Proposalofanew “Medium” Muon Id. URL https://indico.cern.ch/event/357213/contribution/2/material/slides/0.pdf. Visited on 2021-12-04.

[126] NanoAOD variables. URL `https://cms-nanoaod-integration.web.cern.ch/integration/master-102X/mc102X_doc.html`. Visited on 2021-09-06.

[127] SWGuideCrab. URL `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCrab`. Visited on 2021-09-01.

[128] Numpy. URL `https://numpy.org/`. Visited on 2021-09-01.

[129] scientifically. ROOT: analyzing petabytes of data. Root. URL `https://root.cern.ch/`. Visited on 2021-09-01.

[130] Tom Fawcett. ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31:882–891, 2004. ISSN 0167-8655. doi: 10.1016/j.patrec.2005.10.012.

[131] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 04 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq134.

[132] Apeksha Shewalkar, Deepika Nyavanandi, and Simone A. Ludwig. Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU. *Journal of Artificial Intelligence and Soft Computing Research*, 9(4):235–245, aug 2019. doi: 10.2478/jaiscr-2019-0006.

[133] Shudong Yang, Xueying Yu, and Ying Zhou. LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example. In *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, pages 98–101. IEEE, jun 2020. doi: 10.1109/IWECAI50956.2020.00027.