Universiteit Utrecht

Experimental Physics

Using machine learning techniques for non-prompt D^0 analysis

Master Thesis

Colin Ronan Bolle

First Reviewer:

Dr. A. Grelli
GRASP

Second Reviewer:

Prof. Dr. R. Snellings
GRASP

June 26, 2023

Abstract

In the last decade the p_T differential cross-sections of heavy mesons such as D^0 mesons have been measured extensively at the LHC for a variety of rapidity and energy ranges in proton-proton (pp) collisions. These measurements provide tests for standard model theories such as quantum chromo dynamics (QCD) and a baseline for measurements in heavy-ion collisions in which a plasma-like state of matter consisting of deconfined quarks and gluons (QGP) forms. Since the bottom (b) quark is the heaviest quark apart from the top quark it is produced very early in the hard scattering process making it the excellent probe. The properties of the b quark can be indirectly accessed by studying non-prompt D^0 mesons. However a large fraction of promptly hadronized D^0 mesons is present after the hadronization processes. These prompt D^0 mesons are identical to non-prompt D^0 mesons making it extremely challenging to separate the two. In this thesis we study the possibility to maximise the non-prompt over prompt D^0 ratio using two types of machine learning algorithms. We discuss the training results of boost decision trees using adaptive boosting and convolutional neural networks and compare the performance of both algorithms to choose the model which suits the scope of this thesis. We report an increase of the non-prompt fraction between 2.268 ± 0.08 and 69.76 ± 20.1 when the boost decision tree is used replace the standard selection cuts made in the invariant mass reconstruction. The invariant mass is reconstructed in the interval $5 < p_T < 24$ GeV/c with a fraction of about 18% of the data sample available with significances between 2.4 ± 0.8 and 6.3 ± 1.1. Using these significances we show that within the boundaries of the available minimum bias data it is possible to obtain significances greater than 5.0 for $5 < p_T < 24$ GeV/c. Future studies can be performed to improve the algorithms and other classification algorithms, such as transformers, can be used to increase the non-prompt D^0 fraction.

Contents

1 Introduction

In the last decade the p_T differential cross-sections of prompt and non-prompt D^0 mesons have been measured extensively for a variety of rapidity and energy ranges in proton-proton (pp) collisions. These measurements provide tests for standard model theories such as quantum chromo dynamics (QCD) and in particular models base on perturbative QCD such as FONLL. Furthermore these measurements provide a baseline for heavy-ion collision in which a plasma-like state of matter consisting of deconfined quarks and gluons forms. Ordinary matter is bound by asymptotic freedom and confinement as described in QCD. The strong coupling constant becomes asymptotically small for large momentum exchange i.e. small distances such that quarks act as free particles within the distance of the bound states. Furthermore QCD only allows for colourless states such as mesons and baryons, in which the quarks are confined. However, QCD predicts that in extreme conditions of temperature or pressure ordinairy matter may undergo a phase transition after which quarks and gluons are deconfined. This state of matter is predicted by solving the QCD equations for lattice space-time dimensions. The temperature at which this hot and dense matter forms, the critical temperature, is measured to be about 156 MeV at vanishing baryochemical potential which corresponds to $\approx 2 \times 10^{12}$ K. Once the plasma is formed temperatures can rise up to 10^{14} K before the plasma cools down as a consequence of its fast expansion. It is believed that the phase of matter of our early universe, only a few μs after the Big Bang, consisted of this QGP. This phase of hot and dense matter is researched extensively in the Large Hadron Collider (LHC) at CERN. In the collider, particles such as protons or lead ions are collided with energies per nucleon in the TeV range. After a lead-lead (Pb-Pb) collision the energy density is high enough to form the plasma. QGP characterization is performed by means of probes. Among these probes there are fundamental particles like quarks and gluons. In particular charm and bottom quarks are considered excellent tools for QGP tomography. These heavy quarks are considered such excellent probes because they are formed early during the collision, during the hard-scattering phase, and travel trough the entire QGP since their lifetimes are longer than that of the QGP. They interact with the QGP by losing energy and momentum and this can be measured indirectly by comparing Pb-Pb collisions with pp collisions. Moreover, due to their large masses, thermal production of the quarks in the plasma can be neglected. After the hadronic phase and the hadron freezeout these heavy quarks form into heavy mesons such as the D^0 meson. These heavy mesons decay before they can reach any detector system of ALICE but their decay products can be reconstructed efficiently since their lifetimes are much larger. Once the information of the decay products is available the research proceeds via invariant mass reconstruction in order to access the properties of the heavy hadron and consequently the heavy quark constituent. Finally the results of the Pb-Pb collisions are compared with results from pp collisions, in which the QGP is not produced.

The b quark is an extremely important probe of the QGP since it is created before all the lighter quark flavours and lives trough the total evolution of the QGP [1]. The b quark hadronizes in, among other things, $B^{\pm}$ and B^0 mesons. These mesons have a very short lifetime and will decay before they can pass trough the ALICE detector. However these mesons decay partly via $B \rightarrow D^0 + X$ where B can be any B meson and X any decay product(s). This so-called non-prompt D^0 meson can be efficiently reconstructed in ALICE

via $D^0 \rightarrow K^-\pi^+$. Therefore the non-prompt D^0 meson can act as a way to access the b quark properties and therefore to investigate its interaction with QGP. However, during the hadronization after the hard scattering of pp or Pb-Pb particles prompt D^0 from hadronizing charm quarks are also produced. These prompt D^0 mesons have very similar properties compared to non-prompt D^0 mesons. Therefore it is particularly challenging to distinguish between these prompt and non-prompt D^0 mesons. Since the prompt component derives from direct charm hadronization and the fact that the charm quark is much lighter than beauty quark, the result is that the expected non-prompt component is about a fraction 1/40 of the total number of D^0 mesons detected.

In this thesis we will use boost decision trees and convolutional neural networks with the goal to efficiently select non-prompt D^0 in real LHC data. The analysis will be performed in 11 separate D^0 meson p_T intervals between $0 < p_T < 24$ GeV/c. In section 2 we will discuss the relevant theory for this thesis, in section 3 we will discuss the ALICE experiment which is used for our analysis and in section 4 we will discuss D^0 event reconstruction, the framework we developed and the simulated and real data samples used in this manuscript. Section 5 shows the results of the baseline study which uses a standard set of selection cuts to reconstruct the invariant mass of the D^0 mesons. In section 6 we discuss the simulated samples used for algorithm training, the algorithms themselves and the training of the algorithms. Then we compare their performances in order to select the algorithm which suits the scope of this thesis. After the algorithm selection we discuss the validation of the best performing one and we finalise section 6 with the results from the implementation of the algorithm in the invariant mass reconstruction on a data sample. In section 7 we compare the results from the accepted fractions, validation and implementation of the algorithm with the results from the baseline analysis. We finalise this thesis with the conclusions and a discussion in section 8 followed by an outlook in section 9.

2 Theory

2.1 The Standard Model

The Standard Model describes the fundamental building blocks of matter and three of the four fundamental forces of nature. The model was developed in the second half of the 20th century. It is based on a set of fundamental particles including 3 generations of quarks and leptons, 4 gauge bosons and 1 scalar boson. Quarks and Leptons have half-integer spin while gauge bosons have integer spin. The Higgs boson is spinless. Each generation of quarks has one quark with a charge of $\frac{2}{3}e$ and one quark with a charge of $-\frac{1}{3}e$. The first generation of quarks is the lightest while the third generation of quarks is the heaviest with the top quark being by far the heaviest. Each generation of leptons consists of a massive particle with a charge of $-e$ and the corresponding neutrino which mass is expected to be in the eV range. A summary of the standard model particles can be seen in figure 1.

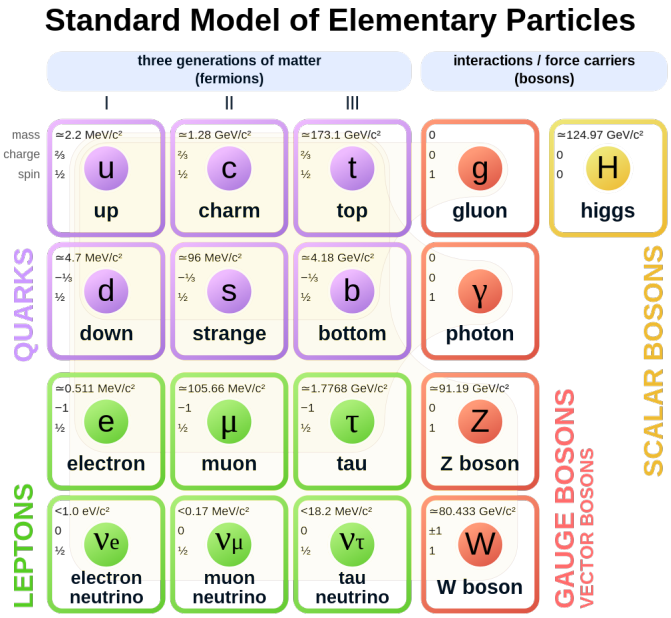


Figure 1: The Standard Model contains 3 generations of quarks and leptons, 4 gauge bosons and 1 scalar boson [2].

The photon (γ) is the carrier or the electromagnetic interaction, the $W\pm$ and Z^0 bosons are the carriers of the weak interaction and gluon is the force carrier of the strong interaction. Mathematically speaking the standard model is described by a $SU(3)\times SU(2)\times U(1)$ symmetry group. Here $U(1)$ is the group referring to the electromagnetic theory, $SU(2)$ the group which describes weak interactions and $SU(3)$ the the theory describing the strong interactions. The latter theory is called quantum chromo dynamics (QCD). The electromagnetic and weak interactions are combined in the $SU(2) \times U(1)$ electroweak theory by Gerard 't Hooft and Marinus Veltman for which they were awarded the Nobel Prize in 1999. A property of the $SU(n)$ groups is that they have in $n^2 - 1$ propagators and $U(n)$ groups have n^2

propagators, which describes the observations of 8 gluons, 3 weak bosons and 1 electromagnetic boson. For this thesis we will focus on quantum chromo dynamics (QCD) and discuss this theory in further detail.

2.2 Quantum Chromo Dynamics

Quantum chromo dynamics is the theory which describes the interactions of partons via gluon exchange. The interactions obey the mathematical principles of a non-abelian $SU(3)$ group. Hence the propagators of the group do not commute. The Lagrangian density is shown in equation 1.

$$\mathcal{L} = \bar{\psi}_q^i (i\gamma^\mu)(D_\mu)_{ij}\psi_q^j - m_q \bar{\psi}_q^i \psi_{qi} - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} \tag{1}$$

Here ψ_q^i is the i-th component of the quark (q) field with $i = \{R, G, B\}$, γ^μ is the Dirac-matrix, D_μ the covariant derivative of QCD, m_q the quark mass allowed by the Higgs field and $F_{\mu\nu}^a$ the gluon strength tensor of sort $a = \{1, .., 8\}$ [3]. The covariant derivative is $(D_\mu)_{ij} = \delta_{ij}\partial_\mu - ig_s t_{ij}^a A_\mu^a$ with $g_s^2 = 4\pi\alpha_s$ and $t_{ij}^a = \frac{1}{2}\lambda_{ij}^a$. Here A_μ^a is the gluon field, α_s the strong nuclear constant and λ_{ij}^a the element of the Gell-Mann matrices. These Gell-Mann matrices are the generators of the $SU(3)$ group and are gluon-specific. The gluon strength tensor is defined as $F_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu - g_s f^{abc} A_\mu^b A_\nu^c$ with f^{abc} being the structure constants of QCD. Requiring phase transformations of ψ to be gauge invariant results in the allowance of triple and quadruple gluon vertexes. Other properties of QCD are confinement and asymptotic freedom. Confinement means that quarks cannot exist separately because particles have to be colour neutral. Mesons consist of a colour anti-colour pair while a baryons are colourless because they contains all 3 colours. Asymptotic freedom states that quarks act as free particles within these confined states. This is due to the asymptotic behaviour of the running strong coupling constant, which grows to very small values for higher momentum exchanges i.e. small distances. Measurements of the running strong coupling constant as function of momentum exchange Q can be seen in figure 2.

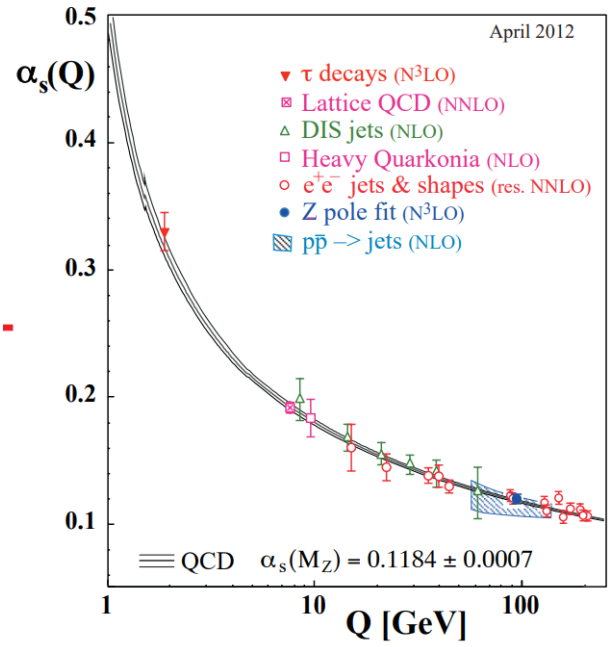


Figure 2: The running strong coupling constant as a function of momentum exchange Q [3].

2.2.1 Perturbative Quantum Chromo Dynamics

Perturbative QCD (pQCD) describes QCD at high momentum exchange. This means it can only be used if $\alpha_s << 1$. Heavy quark production is described in the frame of pQCD via the so-called factorisation theorem [4]. Heavy quark production in pp collisions should not just be considered a reference for lead-lead studies but on top of that it is a very important tool to test pQCD theories. A widely used pQCD model to calculate differential cross-sections of charmed and beauty hadrons is first order next-to-leading log (FONLL). This is a combination of next-to-leading-logarithm (NLL) and next-to-leading order (NLO) calculations for massive quarks [1]. FONLL relies on hard-scattering cross-sections at the partonic level, parton distribution functions (PDFs) and fragmentation functions (FFs). Figure 3 shows an ALICE preliminary p_T-differential cross section of prompt D^0 mesons for pp collisions at $\sqrt{s} = 13$ TeV for rapidities $|y| < 0.5$ compared to FONLL calculations.

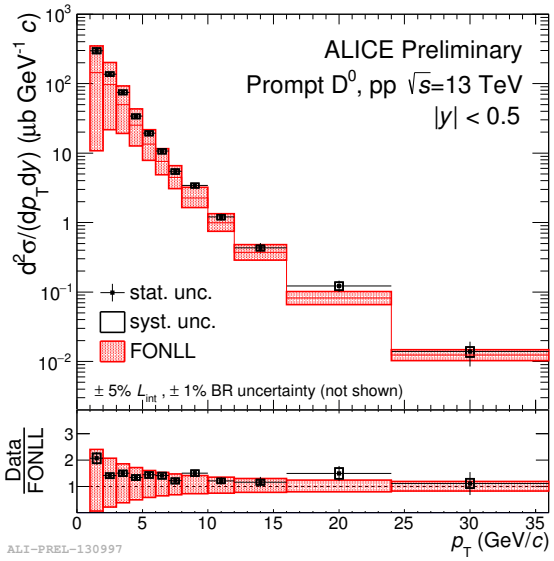


Figure 3: The p_T-differential production cross section of prompt D^0 mesons with $|y| < 0.5$ in the interval $1 < p_T < 36$ GeV/c for proton-proton collisions at $\sqrt{s} = 13$ TeV. The bottom part of the figure shows the ratio between the measured cross-section and FONLL calculations [5].

2.3 Quark Gluon Plasma

To make predictions for values other than $\alpha_s << 1$ lattice QCD was developed. Here equations are solved numerically for lattice space-time points. Lattice QCD predicts the existence of a phase transition to a plasma which has more degrees of freedoms than a system in lower temperatures. The system that is created after reaching the critical temperature for hadrons is the quark gluon plasma (QGP) in which quarks and gluons are deconfined. As can be seen in figure 4, which shows a phase diagram as function of baryonic density and temperature, the critical temperature of the plasma depends on the net baryonic density. For baryonic densities more than 5 times that of normal matter and temperatures a quark gluon plasma forms for much lower temperatures compared to ordinairy baryonic densities. At the LHC very low baryonic densities (almost 0) are studied an therefore the critical temperature of the QGP ath the LHC is approximately 156 MeV.

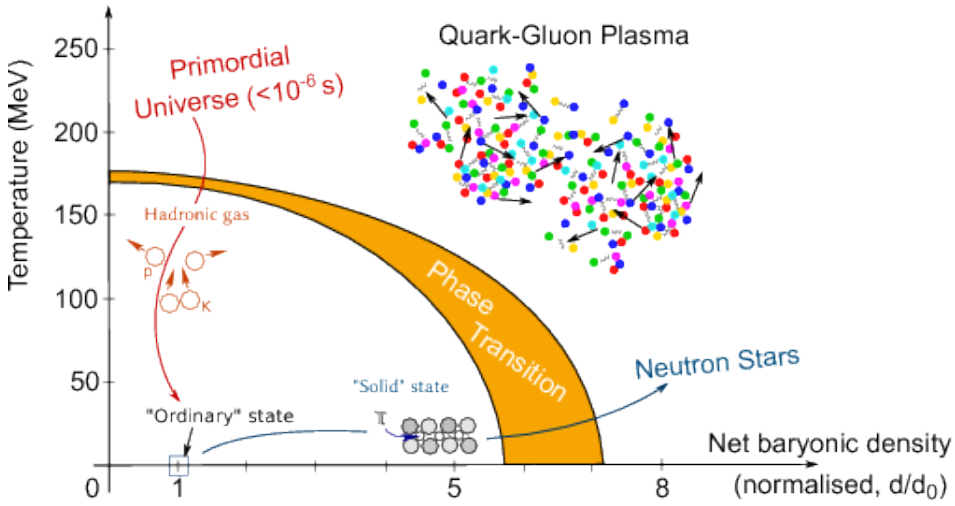


Figure 4: Phase diagram as function of baryonic density and temperature. The QGP forms for temperatures above 156 MeV for low baryonic densities. For increasing baryonic density the critical temperature becomes lower [6].

It is believed that a few μs after the big bang the universe consisted of a hot and dense plasma of deconfined quarks and gluons [1]. As the universe expanded it cooled down until it reached below the critical temperature. The plasma underwent hadronization similar to what happens in the LHC after a heavy-ion collision. In the LHC the QGP is expected to have a lifetime of 10 fm/c for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [1]. After a collisions the heavy quarks form before the lighter quarks since $\Delta t \approx 1/2m_q$. The corresponding times for charm and beauty quarks are well below the expected lifetime of the QGP making them the excellent probe of the entire lifetime of the QGP.
During the hadronization phase the charm and beauty quarks start binding into hadrons. Relevant for this thesis are B and $D^0(c\bar{u})$ meson production. The B mesons containing a (anti-)b quark can decay via $B \rightarrow D^0 + X$ where B can be any of the three B mesons and X any decay product(s). The D^0 meson coming from this decay is a non-prompt D^0 and will acts as a beauty quark proxy for QGP investigations. D^0 forming directly from a charm quark are prompt D^0 mesons and can act as a proxy as well. The aim for this thesis is to investigate the possibility to efficiently select non-prompt D^0 mesons using machine learning methods and thus allowing for the investigation of beauty quark interaction with the QGP.

2.4 Machine learning algorithms

In recent years machine learning algorithms are becoming more and more popular to solve complex problems such as classification problems. Due to the high processing capabilities from modern computers they simply outperform humans when looking at high dimensional problems. Machine learning can be supervised, where it is constantly evaluated if the algorithm makes the right decision, or proceed via unsupervised learning, where the algorithm looks for patterns in untagged data. Supervised machine learning algorithms are trained by feeding them a set of training data. The algorithms adjusts its weights such that the output value of the algorithm corresponds to the correct answer. One has to be careful that the algorithm doesn't overtrain. Overtraining means that an algorithm performs very well on a training set but very poorly on a test or evaluation dataset. The network is simply too adjusted to the training set. Even though this supervised learning procedure is quite general there is a large variety of available models, configurations and model sizes. Examples of machine learning algorithms that are commonly used for classification problems are Neural Networks (NNs) or boost decision trees (BDTs). For this thesis we will focus on BDTs and convolutional neural network (CNN), which is a neural network with extra convolutional features.

2.4.1 Boost Decision Trees

A boost decision tree (BDT) is an type of algorithms commonly used for classification problems. It takes a set of input features and selects data based on those features. A simple illustration of a decision tree can be seen in figure 5. In this example an event is only classified as signal if it has 100 or more TPC hits, more than 0.2 GeV energy and a normalized decay length larger than 8. In this case the depth of the tree is 3, because there are 3 decision steps.

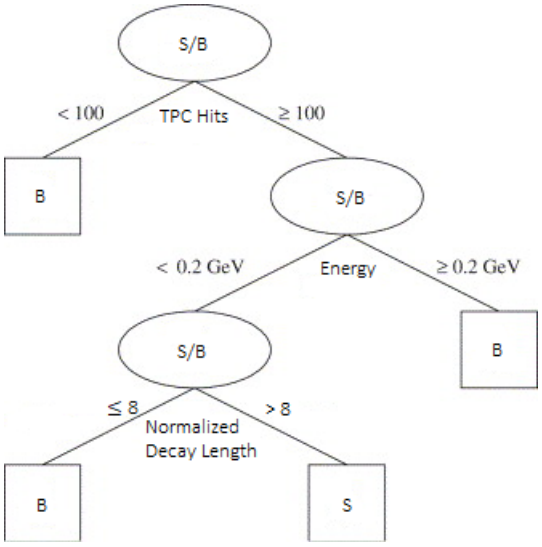


Figure 5: Example of a decision tree. Events do or do not pass depending on the value of their specific features. Here S/B are events that might be signal or background, B are background events and S are signal events.

A single decision tree is often not precise enough for complex classification problems. Therefore decision trees are often boosted. This means that multiple weak classifiers are combined to create a better performing one. Mathematically this can be written as $F(\overrightarrow{x_i}) = \sum_{k=1}^{N} \alpha_k T_k(\overrightarrow{x_i})$, where $F(\overrightarrow{x_i})$ is the output of the boosted classifier, T_k the output of a single classifier trained on training set $\mathbb{T}_k$ and α_k the weight of classifier T_k [7].

There are different types of boosting. A very commonly used type of boosting is adaptive boosting, or in short AdaBoost. For this type of boosting the misclassified events by classifier T_k are evaluated. Using signal and background label $y_i = \pm 1$ with -1 being the label for the latter and that $\mathbb{I}(X) = 1$ if statement X is satisfied we can define isMisclassified$_k = M_k = \mathbb{I}(y_i \times T(\overrightarrow{x_i}) < 0)$ [7]. Since $T_k(\overrightarrow{x_i})$ returns 1 for signal and -1 for background the product of a false prediction is always negative and hence M_k returns 1 for a false prediction. From this we can define the misclassification rate $R(T_k) = \epsilon_k$.

$$\epsilon_k = \frac{\sum_{i=1}^{N_k} w_i^k M_k(\overrightarrow{x_i})}{\sum_{i=1}^{N_k} w_i^k} \tag{2}$$

where N_k is the number of events used to train classifier T_k and w_i^k is the weight for individual event i in set $\mathbb{T}_k$. Now α_k can be expressed in terms of ϵ_k, such that classifiers that perform poorly are weighted less. $\alpha_k = \beta \ln(\frac{1-\epsilon_k}{\epsilon_k})$ where β is the strength of the boosting.

The fundamental power AdaBoost however lies in the training of T_{k+1} on training set $\mathbb{T}_{k+1}$. The weights of events i in set $\mathbb{T}_k$, w_i^k, are transformed as $w_i^{k+1} = w_i^k \times e^{\alpha_k M_k(\overrightarrow{x_i})}$ such that properly classified events remain unchanged. Because previously misclassified events now weigh more the new classifier T_{k+1} will focus more on those events. This will result in an increase of overall performance of the BDTs.

2.4.2 Neural Networks

Neural networks are artificial networks with a structure which is deduced from the neurological structure in animal brains. A neural network consists of an input and output layer and has hidden layers in between. In a neural network a node from a certain layer is connected to all nodes in the next layer similar to a brain where a neuron is connect to all other neurons. Every neuron receives input values from the previous layers and will assign weights to them. An illustration of an example neural network can be seen in figure 6. The neuron will multiply the input values by their weights and sum those values. The final step is the addition of a bias which can be either positive or negative. Then this final value is passed to an activation function. The mathematical expression for the value passed to the activation function is $v = \sum_{i=1}^{n} w_i x_i + b$, where x_i the vector of input values, w_i is the weight vector for the input variables and b is the bias value.

There are different sorts of activation functions. Often a combination of activation functions is used for different layers in a network. Commonly used activation functions are the rectified linear activation function (ReLU), exponential linear unit (ELU) and softmax. ReLU returns 0 for all negative input values and returns the input value for positive input values. ELU is very simlar to ReLU because it also returns the input value for positive values but it returns

$\alpha(e^x - 1)$ for negative input values of x. The value returned by the activation function is passed on to the nodes in the next layer. This process repeats for every node in every layer till the final output is produced.

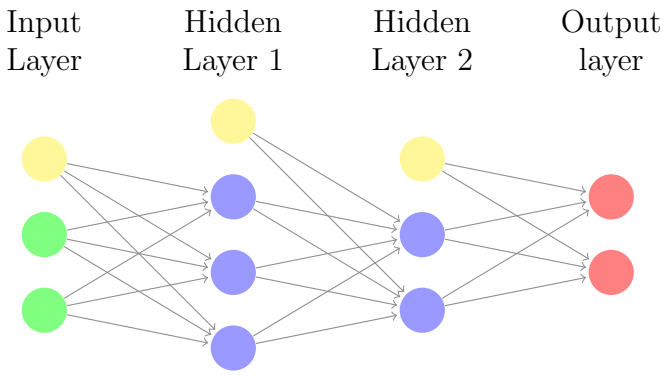


Figure 6: An example of a neural network. The green dots are the input values, the blue dots are the hidden nodes and the red dots are the output values. To each node in a layer the same bias is added (yellow dots).

In the training phase the weights of the model are adjusted such that the accuracy improves. Weights are adjusted by minimizing a so-called loss function. A commonly used loss function is the Mean Squared Error (MSE) function. It is defined as $MSE = \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2$, where y_i is the observed value, $\hat{y}_i$ the expected value and n the number of values evaluated. The weights are only adjusted if the loss function decreases. Weights are adjusting after each batch (a subset of the full training set) has been fed to the model. Furthermore the full training set is fed multiple times to the model, this is called an epoch. Varying the batchsize and the number of epochs can change the outcome of the training.

The weights are not randomly adjusted. The adjustment is done by an optimizer. This can be a function or an algorithm. Depending on the optimizer more parameters can be modified such as the learning rate of the model. The learning rate decides how much weights can be shifted when they are updated. One of the best optimizers is the adaptive moment (Adam) optimizer. The next part of this chapter follows the line of reasoning from [8]. The Adam optimizer is a gradient descent optimizer which means it aims to find the minimum of a function. It uses two ways to obtain new weights. The first term that is used to calculate the new weights can be seen in equation 3. This is called momentum.

$$w_{t+1} = w_t + \alpha_t m_t \tag{3}$$

Here w_{t+1} are the new weights, w_t the current weights, α_t the learning rate at the current time step and m_t the aggregate of the gradient. Here m_t is defined in equation 4.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\left[\frac{\delta L}{\delta w_t}\right] \tag{4}$$

Where β_1 is a fixed parameter value and L the loss function. Taking the average value of the derivative decreases the time it takes for the algorithm to converge. The second term that

is used to obtain the new weights can be seen in equation 5. This is the root mean square (RMSprop) algorithm.

$$w_{t+1} = w_t - \frac{\alpha_t}{(v_t + \varepsilon)^{1/2}} * \left[\frac{\delta L}{\delta w_t}\right] \tag{5}$$

Where ϵ is a small constant and v_t the square of the previous gradient. The definition of v_t can be seen in equation 6.

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) * \left[\frac{\delta L}{\delta w_t}\right]^2 \tag{6}$$

Where β_2 is another constant and v_{t-1} the square of the previous timestep. Note that $v_{t=0} = 0$ and $m_{t=0} = 0$ by definition. Combining these formulas results in equation 7.

$$m_{t-1} + (1 - \beta_1)\left[\frac{\delta L}{\delta w_t}\right] v_t = \beta_2 v_{t-1} + (1 - \beta_2)\left[\frac{\delta L}{\delta w_t}\right]^2 \tag{7}$$

Since both m_t and v_t are initialized as 0 they tend to be biased around 0. To account for this a bias correction $\widehat{m_t} = \frac{m_t}{1-\beta_1^t}\widehat{v_t} = \frac{v_t}{1-\beta_2^t}$ is defined. The final expression using bias terms results in equation 8.

$$w_{t+1} = w_t - \widehat{m_t}\left(\frac{\alpha}{\sqrt{\widehat{v_t}} + \varepsilon}\right) \tag{8}$$

By varying network size, the amount of layers, the activation function, the optimizer and/or the amount of nodes per layer the neural network can be optimised.

2.4.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a special type of neural networks consisting of fully connected layers, like in a standard neural network, and convolutional layers in which the dimensionality of the input is decreased. These layers reduce the dimensionality of the input but nevertheless the important information remains intact. The reduction of the dimensionality can increase the computation speed or decrease the demand of computational power. It is convenient to explain the working of a CNN using 2D images as input where each pixel has a number corresponding to a specific color but please note that this also works in an identical way for 1-dimensional arrays of input values which are used for this thesis. A convolutional layer decreases the dimensionality, among other things, by converting the input to a feature map using a kernel. An example can be seen in figure 7.

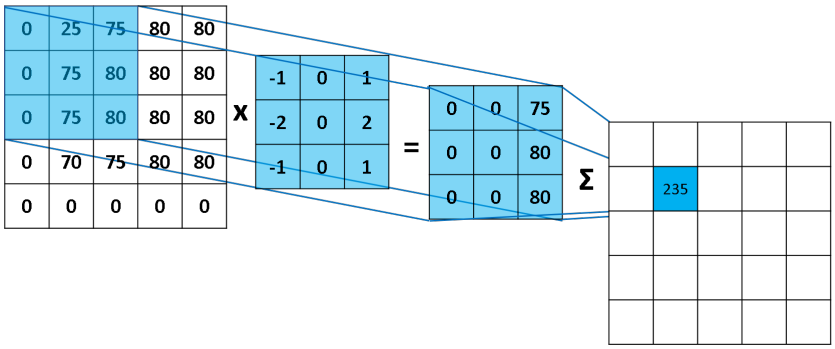


Figure 7: Illustration of the working of a kernel inside a convolutional layer. The input values are multiplied by the kernel and the sum of the kernel multiplication results in the value of the feature map [9].

In the case of this example the dimensionality is decreased from 5 by 5 to 3 by 3 provided that the shifting of the kernel after each iteration (stride) is 1. In that case the kernel shifts to the right with 1 column at the time. After the first row is completed the kernel shifts back to the left and drops 1 row. It is possible to increase the stride or kernel size to reduce the dimensionality of the feature map even further. The values of the feature map are passed trough the chosen activation function before they are pooled. Pooling is another method used in a convolutional layer to reduce the dimensionality. Forms of pooling are min-pooling, max-pooling and average-pooling. In pooling a group of values if replaced by the minimum, maximum or average value. An example of the pooling of a feature map with pool size 2 by 2 and stride 2 can be seen in figure 8.

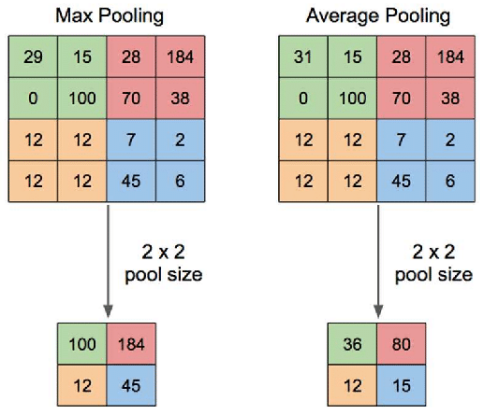


Figure 8: Illustration of max and average pooling of a feature map with stride 2 [10].

After the pooling the values are passed to the next layer. This can be another convolutional layer or a fully connected layer, which is the standard neural network layer discussed in section 2.4.2.

3 Experimental Setup

3.1 The ALICE experiment

The ALICE experiment is one of the four experiments at the large hadron collider (LHC). The LHC is the largest particle accelerator in the world with a circumference of 27 kilometer where protons or heavy ions are collided with energies in the TeV range. The ALICE experiment is designed to study strongly interacting matter at extreme energy densities [5]. The collaboration consists of more than 2000 scientist coming from 40 countries.

The ALICE detector weighs over 10.000 tonnes and is 16 meters tall and wide [11]. Surprisingly this is the smallest detector at the LHC. It consist of 17 sub-detector systems each playing their own role in nuclei collision event reconstruction. A schematic of the ALICE detector is shown in figure 9.

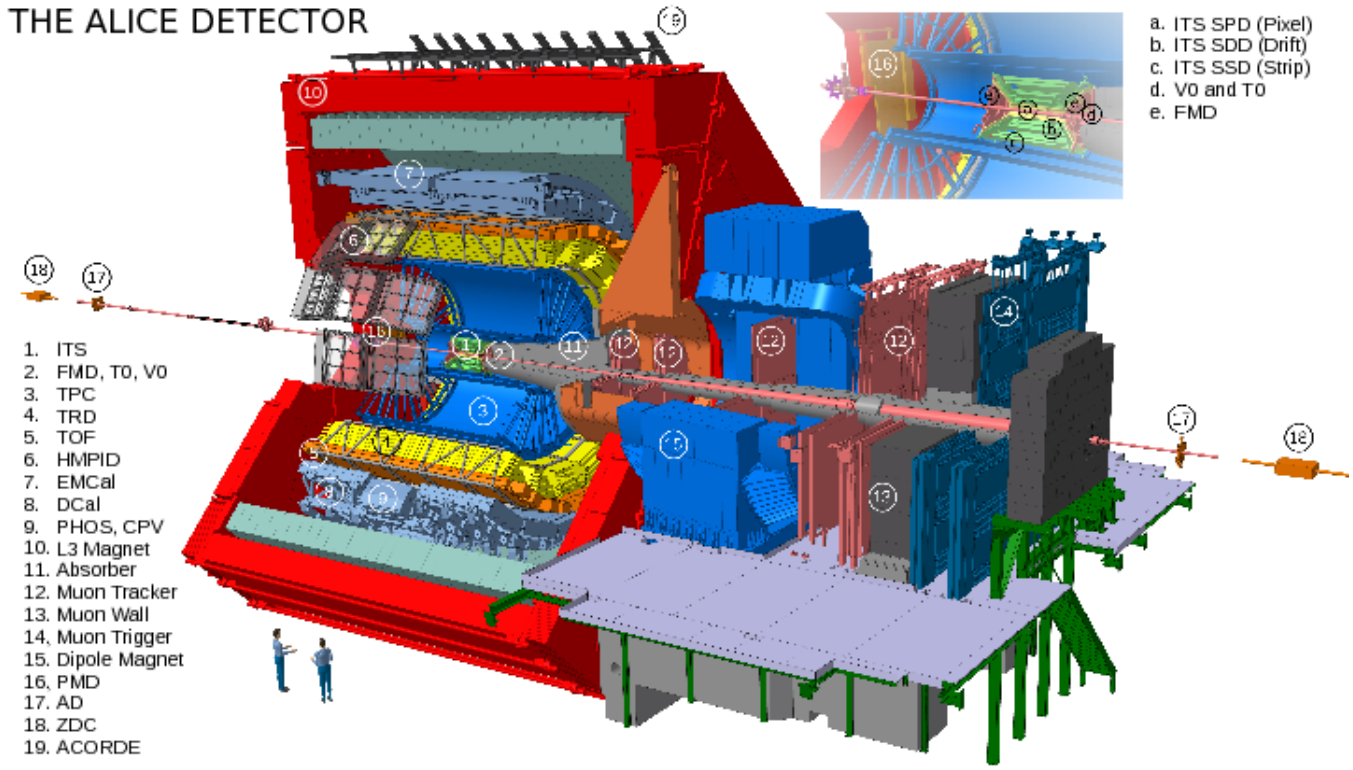


Figure 9: A schematic of the ALICE detector. The left part of the detector is the central barrel in which the collision takes place. The right part of the detector is the forward muon arm. The legend shows the abbreviations used to describe the components of the detector [12].

The central part of the detector covers the region between 45° and 135° where 90° is perpendicular to the beam. The detector is located inside a magnetic field with the a strength of 0.5 T [11]. The detector can be divided into two areas, the central barrel containing the tracking

and particle identification detectors and the forward muon arm. The collision occurs in the middle of the central barrel.

3.1.1 Inner Tracking System

The detector closest to the collision point is the Inner Tracking System (ITS). The ITS is a 6 layer silicon vertex detector whose primary goal is to locate the primairy vertex, the position where the nucleon or nuclei collision took place, and secondary decay vertices of fast-decaying heavy hadrons. An illustration of the ITS can be seen in figure 10. The first two layers are high resolution Silicon Pixel Detectors (SPDs). These layers record both x and y position of passing particles. The SPDs are the fastest triggering layers in LHC with a response time of less than 900 ns [11]. The SPD consist of 1200 pixel chips where each chip has 8192 individual cells [13]. Therefore the full SPD detector has $\approx 10^7$ pixels. When a particle passes trough the detector it interacts with the nearby pixels resulting in an electrical signal and due to the high resolution a precise position can be determined. Due to to extremely fast trigger the detector will also be used as an interaction trigger.

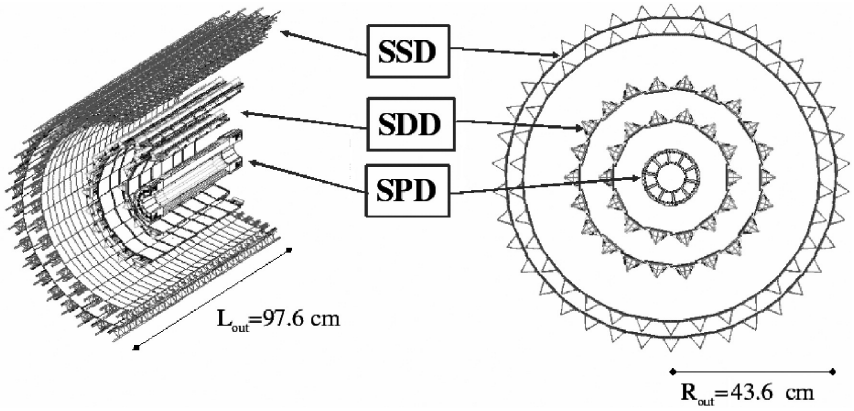


Figure 10: Illustration of the Inner Tracking System of the ALICE detector. The 2 most inner layers are the silicon pixel detectors, the 2 layers in the middle are the silicon drift detectors and the outer 2 layers are the silicon strip detectors [14].

The middle 2 layers of the ITS are the Silicon Drift Detectors (SDDs). These detectors consist of 2 regions in which particles drift to the opposite direction. When a particle hits these detectors it ionises parts of the silicon layers along its track. The electrons drift towards one of the anodes which have a voltage of -1800 V with a bias of -40 V to keep the biasing of the collection region independent of the drift voltage [15]. Both drift regions contain 256 anodes to collect the charges. The detector was carefully calibrated by injecting charges in over $1 * 10^5$ locations to account for systematic deviations caused by the non-linearity of the voltage divider. Furthermore the detector is calibrated every day or every 6 hours in three different ways. However discussing these methods is beyond the scope of this thesis. The SDDs also provide information on the energy loss of the passing particles which is important for particle identification. The last two layers of the ITS are Silicon Strip Detectors (SSDs). These detectors connect the track measured in the next detector to the track measured in

the ITS and also provides further information of the energy loss.

3.1.2 Time Projection Chamber

Next to the ITS is the Time Projection Chamber (TPC). It is chosen as the main particle tracker in ALICE [11]. Therefore it requires excellent precision in measuring momentum and energy losses [16]. The TPC has an acceptance of 2π in azimuthal angle and covers pseudorapidities $|\eta| < 0.9$. The TPC is a 90 m^3 large tube filled with a Ne-CO_2-N_2 mixture cooled down to temperatures below 0.1 K [16]. Inside the tube a homogeneous electric potential of 400 V/cm with distortions in the order of 10^{-4} is present [16]. An illustration of the TPC can be seen in figure 11. The readout plates are placed at the endplates of the tube and are divided in 18 sectors each consisting of an inner and outer readout chamber. When a particle passes trough the TPC it ionises the gas along its track. The electrons will drift towards the readout plates due to the electric field. Using the time difference of which the electrons hit the endplates the path can be reconstructed. Furthermore the energy loss of a particle can be determined via the Bethe-Bloch function, which can be seen in equation 9, if the track has at least 120 out of 160 possible hits.

$$\langle \frac{dE}{dx} \rangle = \frac{4\pi N e^4}{mc^2} \frac{Z^2}{\beta^2} [\ln(\frac{2mc^2\beta^2\gamma^2}{I}) - \beta^2 - \frac{\delta(\beta)}{2}] \tag{9}$$

Where dE/dx is the energy loss as function of position, N the number density of electrons, m the electron mass, Z the charge of the passing particle and I the mean excitation energy of the ionised atom in the gas.

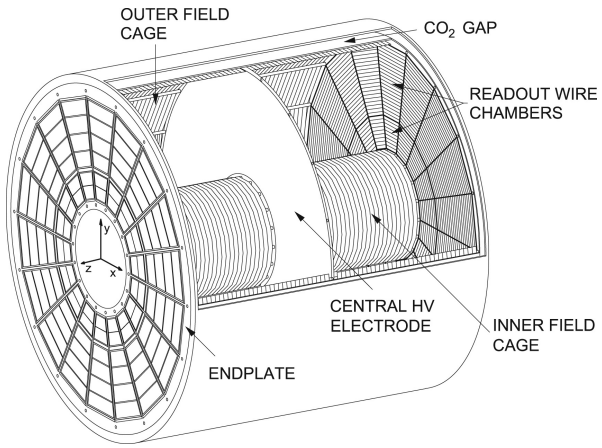


Figure 11: Illustration of the TPC [16]. On the left and right ends are the readout plates while the electrode is placed in the middle such that a potential of 400 V/cm is present.

As mentioned before the energy loss is an important parameter in particle identification. Together with momentum and velocity measurements it is possible to reconstruct the invariant mass of a particle, revealing its identity. For the ALICE TPC they use a parametrization of the Bethe-Bloch formula proposed by the ALEPH experiment [17].

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} [P_2 - \beta^{P_4} - \ln(P_3 + (\beta\gamma)^{-P_5}] \tag{10}$$

The parameters will be fixed once the exact gas mixture is known [16]. Figure 12 shows the energy loss for i.e. electrons, pions and protons in pp collisions as $\sqrt{s} = 7$ TeV.

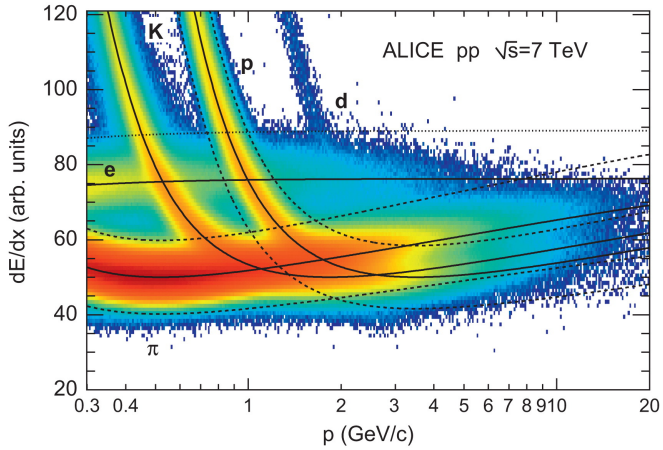


Figure 12: Energy loss in the ALICE TPC for $\sqrt{S} = 7$ TeV in pp collisions [16].

3.1.3 Time of Flight detector

The energy loss measurements from the TPC are sufficient for particle identification. However having other particle detectors allows for particle identification in higher p_T ranges. The first detector after the TPC is the time-of-flight (TOF) detector with an active area of approximately 140 m^2. Its main purpose is to distinguish pions, kaons and protons [11]. Due to the extremely high particle multiplicities it is required to have a time resolution of 50 ps. The main components of the TOF detector are the multigap resistive plate chambers (MRPCs). Figure 13 shows a schematic of the TOF components and the placement on the large frame in the central barrel. There are 2 detector systems attached to this metal frame, the TOF and transmission radiator detector (TRD). The purpose and working principle of this detector is beyond the scope of this thesis. The TOF is, similar to the TPC, based on the ionisation of a thin layer of gas, this time between a cathode and anode plate. The resistive plates are 0.5 mm thick and each chamber has 5 gaps with a size of 220 μm. In total the detector has almost 160.000 readout channels.

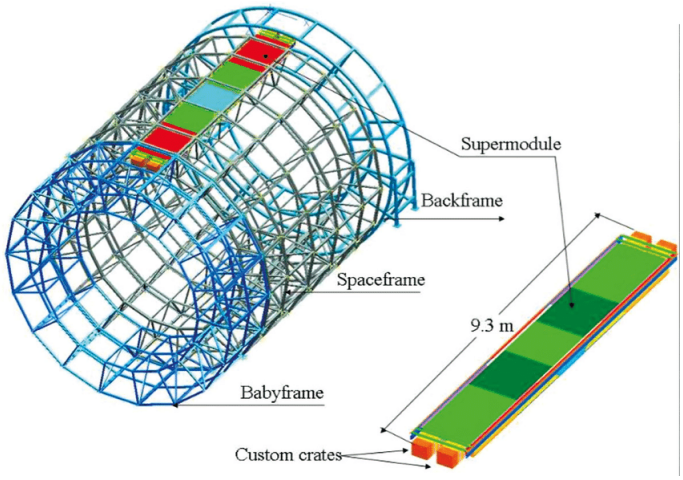


Figure 13: A schematic of one of the 18 TOF modules in the ALICE detector and the placement of the component on the spaceframe in the central barrel [18].

The TOF determines a particles velocity by measuring the time it takes for particles to travel a certain distance along a given track. Figure 14 shows the velocities β as function of momentum p for proton-proton collisions at $\sqrt{s} = 13$ TeV for different particles [19].

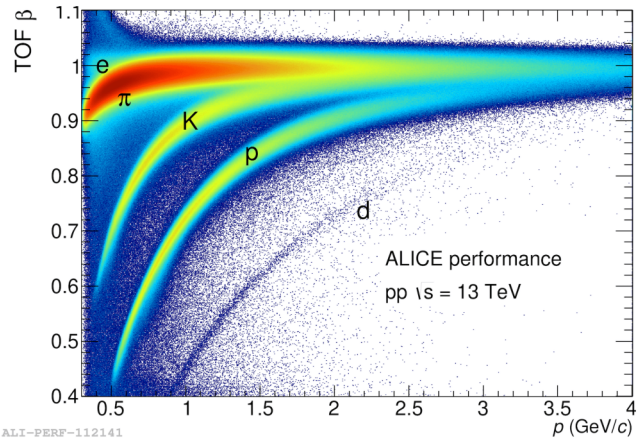


Figure 14: The velocity β measured by the ALICE TOF as function of momentum p for pp collision at $\sqrt{s} = 13$ TeV [19].

3.1.4 Forward detectors

There are also a few detectors placed in the forward region, outside of the central barrel. One of these detectors is the muon spectrometer. The muon spectrometer is placed in a magnetic field of 0.7 T with an opposite direction compared to magnetic field inside the central barrel. The muon spectrometer is shielded from most of the produced particles by a hadron absorber, lead-tungsten shield and an iron wall. The muons are measured by 10 cathode pad tracking chambers. Each pad consist of 2 planes to provide 2 dimensional space information [11].

Another detector in the forward region is the Forward Multiplicity Detector (FMD) measuring the amount charged particles. Furthermore a few small calorimeter (ZDCs) are placed over 100 meter inside the LHC tunnel to measure extreme forward neutral particles [11].

3.1.5 VZERO system

The VZERO system is a detector consisting of two arrays VZERO-A (V0A) and VZERO-C (V0C) covering rapidity ranges $2.8 < \eta < 5.1$ and $-3.7 < \eta < 1.7$ and is used to provide a trigger for minimum bias and high-multiplicity events. The VZERO components are scintillator detectors. Each array is made of BC404 plastic and has a thickness of 2.5 (V0A) or 2.0 (V0C) cm. The overall design of the arrays is adapted to the constraints given by the design of the full detector [20]. The photons produced in the array is transferred to a fine-mesh multiplier tube where the signal is split into two before being sent to the electronic read-out. One of the two channels is amplified by a factor 10 before being sent to the electronic readout while to other channel is transferred directly. This allows for 2 types of trigger algorithms which described in more detail in [20]. The minimum bias trigger is obtained by a time coincidence of the V0A and V0C triggers. The charged particle multiplicity selection uses the sum of V0A and V0C signals, which is denoted as V0M [21]. High multiplicity events are selected if the V0M signal exceeds 5 times the signal of the average minimum bias event.

4 Methodology

4.1 D^0 invariant mass analysis

In this thesis we study the prompt and non-prompt D^0 fractions by means of invariant mass analysis. The invariant mass is defined as $m_0^2 = E^2 - ||\boldsymbol{p}||^2$ where m_0 is the invariant mass, E the energy of the particle and $\boldsymbol{p}$ the momentum in spatial coordinates. The invariant mass of the D^0 meson is $M_{D^0} = 1864.83 \pm 0.05$ MeV when using the PDG fit value [22]. For our studies we have chosen to solely look at the $D^0 \to K^-\pi^+$ decay. The D^0 meson decays into a kaon and pion pair $3.89 \pm 0.04\%$ of the time [22]. We have chosen this decay channel because it is the hadronic decay mode with the largest branching fraction and the decay products can be reconstructed well in the ALICE detector. Even though there are (semi-)leptonic decay modes with a larger branching fraction it is much less suitable for invariant mass reconstruction due to the invisibility of the neutrino in the detector which results in missing energy and momentum which prevents an accurate invariant mass calculation. The first step in the reconstruction of the D^0 candidate is to reconstruct and pair the kaon and pion tracks using data from the ITS, TPC and TOF. Tagging the kaon and pion pairs reduces the recombinatorial background. Using the reconstructed tracks the location of the secondary vertex is determined. The smallest distance between the kaon and pion track is the distance of closest approach (dca). The ITS, TPC and TOF also provide information of the momentum the daughter particles. Using the conservation of momentum the momentum of the D^0 candidate can be reconstructed. Furthermore the daughter tracks are extrapolated. This is a very straightforward procedure since the tracks must be part of a circle with a certain radius due the Lorentz force, which is always perpendicular to $\overrightarrow{\beta}$ resulting in a circular motion. The ITS determines the location of the primary vertex by reconstructing tracks from primary particles to a single location. This completes the D^0 candidate reconstruction from primary vertex all the way to the identified daughter particles and their tracks. An illustration of the candidate topology can be seen in figure 15. Here DCA_{12} is the distance of closest approach. DCA_K and DCA_π are the impact parameters of the kaon and pion respectively and will be referred to as such. The shorthand notation used in this thesis for the impact parameters is $d_{0,K}$ and $d_{0,\pi}$.

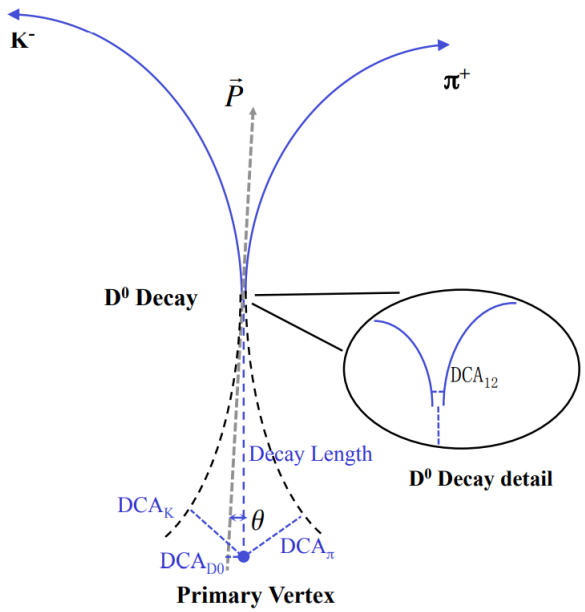


Figure 15: Decay topology of a D^0 meson illustrating the distance of closest approach (DCA_{12}) and the pointing angle $\cos(\Theta)$. The blue lines are the daughter tracks and the grey dotted line is the reconstructed momentum of the D^0. The black dotted lines are the extrapolated tracks of the daughter particles [23].

In the invariant mass analysis D^0 candidates are accepted or refused based on a selection of topological variable cuts. The candidate is only accepted when it matches the condition for all chosen selection variables. Both the dca and the impact parameters will be used as a selection variable. Another variable that can be determined from the decay topology is the pointing angle (Θ). The pointing angle is the angle between the reconstructed D^0 momentum, which is the grey dotted line in figure 15, and the straight line between the primairy and secondairy vertex (D^0 flight path), which is the blue dotted line. This angle can be determined either in a 2D projection of the tracks, often on the XY-plane, or in 3D. Since the pointing angle is expected to be very small due to the high resolution of the detector it is more convenient to use $|\cos(\Theta_{XY})|$ or $\cos(\Theta)$ when evaluating candidate. For our analysis we will evaluate both $|\cos(\Theta_{XY})|$ and $\cos(\Theta)$.

When considering the center-of-mass (COM) frame of the D^0 meson it possible to define another topological variable. In the COM frame the line on which the daughters travel is a straight line since they travel in the exact opposite direction. The angle between the line of flight of the D^0 and the lines on which the daughter particles move is Θ^*. For our candidate selection we will use $\cos(\Theta^*)$ as a selection variable. Other variables that we will use as selection variables are the momenta of the daughter particles, the decay length in 2D and 3D and the normalised decay length in 2D and 3D. The decay length is simply the distance between the primary and the secondary vertex. Again 2D means a determination of a variable using projections on the XY-plane.

4.2 AliPhysics

For this thesis we used the AliPhysics package from the ALICE collaboration. AliPhysics is a package built in ROOT, which is a C and C++ oriented framework to analyse data made at CERN [24]. AliPhysics is used to run analysis tasks using real data or MC simulations. The analysis task used for this thesis is AliAnalysisTaskSED0NonPromptFraction.cxx with corresponding file AddTaskD0MassNonPromptFraction.C to run the analysis task. This task performs the D^0 invariant mass analysis as described in section 4.1 and can also be used to store D^0 candidates in trees to use later for machine learning training. The task was modified such that when running simulations it would allow for the D^0 candidates to be stored in a prompt, non-prompt and background tree. These trees would later be used for algorithm training. Furthermore the task was modified such that machine learning algorithms could be used to evaluate the D^0 candidates. The unmodified versions of the AliPhysics files used for this thesis can be found in either AliPhysics/PWGHF/vertexingHF or AliPhysics/PWGHF/vertexingHF/macros. The mass peaks were obtained by using the output of the analysis task in FitMassSpectra.C. The secondary task used for this thesis was AliCFTaskVertexingHF with corresponding file AddTaskCFVertexingHF to run the secondary task. This task was used to study the accepted prompt and non-prompt D^0 fractions by the standard cuts and machine learning algorithms. The standard cuts were made using makeTFileCutsD0toKpi.

4.3 TMVA

The machine learning package used for this thesis was TMVA. TMVA is a supervised machine learning package with numerous features that simplify the data preparation, algorithm configuration and training and testing of machine learning algorithms. TMVA is a package built in ROOT and also works in Python using PyROOT. TMVA can be used to train all sorts of machine learning algorithms such as boost decision trees, neural networks, k-nearest neighbour and multilayer perceptrons. TMVA is compatible with various machine learning packages such as TensorFlow (Keras), PyTorch and Sci-Kit learn [25]. From the TMVA package the dataloader was used to read the D^0 candidate trees in the .root files and the factory was used as a tool to train and test the different algorithms used in this thesis.

4.4 Data and MC samples

All samples used for this thesis can be found on the Alice GRID Monitor website MonALISA under LEGO trains. The trains used for this thesis are the HF_D2H_pp for real pp data at a center-of-mass energy of $\sqrt{s}$ = 13 TeV and HF_D2H_pp_MC for Monte-Carlo (MC) simulations.

4.4.1 Data samples

The data used for this thesis belongs to the 2016 and 2018 LHC data campaign. 2018. These events are minimum bias events. The corresponding minimum bias triggers for these runs are the CINT7-[B, ACE] classes. From 2016 the full set LHC2016_AOD234_deghjop_13TeV

was used. From 2018 the l, p and k subsets of LHC2018_AOD264_bdefghijklmnop_13TeV were used.

4.4.2 Forced Monte-Carlo samples

In order to obtain a sufficient amount of prompt and non-prompt D^0 candidates to feed to the machine learning algorithms forced MC events were analysed. On top of that forced MC events were used to evaluate the machine learning algorithms and compare with the baseline. These simulations use the Pythia8 event generator the generate the events. In these events either a $c\bar{c}$ or $b\bar{b}$ is added to a minimum bias configuration to increase the heavy meson production. The probability that a $c\bar{c}$ or $b\bar{b}$ is added to an event is $p_{b\bar{b}} = p_{c\bar{c}} = 0.5$. We specifically used forced MC runs from LHC20f4a_2018_P8 which are runs connected to minimum bias data from 2018.

4.4.3 Minimum bias MC samples

To check whether our forced samples contain realistic physical features minimum bias MC samples were used. These simulations are produced using the Pythia8 event generator and mimic minimum bias events. The minimum bias MC samples used for this thesis are $\sqrt{s} = 13$ TeV events from the LHC19g6f2_XcP8_2017 runlist. These runs are connected to minimum bias data from 2017. These runs were chosen because they are between 2016 and 2018 and are therefore a reasonable average when taking the small changes of the ALICE detector due to aging into account.

5 Baseline analysis

The baseline for this thesis uses a set of cuts on the topological variables discussed in section 4.1 to reconstruct the invariant mass of the D^0 mesons. The baseline does not set a cut for the decay length in both 2 and 3 dimensions and the normalized decay length in 3 dimensions. In this section we will first look at these standard cuts and the corresponding prompt and non-prompt accepted fractions. Then we study the invariant mass peaks obtained using a forced MC sample. This same set is later used for machine learning algorithm validation and is ran to allow for a comparison with the baseline. Finally we will study the invariant mass peaks obtained by applying the standard cuts on real LHC data and compare the reconstructed invariant mass fits from MC and data.

5.1 Standard Cuts

The invariant mass reconstruction is performed separately for D^0 transverse momentum intervals. There are 11 intervals used in the analysis between $0 < p_T < 24$ GeV/c and the intervals have different widths. Tables 1 & 2 show the standard cuts for the topological variables measured in the $D^0 \to K^-\pi^+$ decay discussed in section 4.1. Here the type of the cut indicates whether a particle is selected if the value of the candidate is larger or smaller than the cut value. In these tables we can see that the cuts may vary for different p_T intervals. The type of cut means that the value of a candidate has to smaller or larger than the cut value. A candidate is only selected if it passes all 11 cuts.

Variable	**Type**	**[0,1] GeV/c**	**[1,2] GeV/c**	**[2,3] GeV/c**	**[3,4] GeV/c**	**[4,5] GeV/c**	**[5,6] GeV/c**
$\lvert$M-M$_{D^0}\rvert$ [GeV/c^2]	$<$	0.3	0.3	0.3	0.3	0.3	0.3
dca [cm]	$<$	0.03	0.025	0.03	0.03	0.03	0.03
$\cos(\Theta^*)$	$<$	0.08	0.8	0.8	0.8	0.8	0.8
p_{T_K} [GeV/c]	$>$	0.8	0.8	0.8	0.8	0.8	0.8
p_{T_π} [GeV/c]	$>$	0.8	0.8	0.8	0.8	0.8	0.8
$\lvert d_{0_K}\rvert$ [cm]	$<$	0.1	0.1	0.1	0.1	0.1	0.1
$\lvert d_{0_\pi}\rvert$ [cm]	$<$	0.1	0.1	0.1	0.1	0.1	0.1
$d_{0_K} d_{0_\pi}$ [cm^2]	$<$	-0.0004	-0.0003	-0.00026	-0.00015	-0.0004	-0.0001
$\cos(\Theta)$	$>$	0.9	0.9	0.9	0.85	0.9	0.85
$\cos(\Theta_{XY})$	$>$	0.998	0.998	0.998	0.998	0.998	0.998
NDL$_{XY}$	$>$	5	5	5	5	5	5

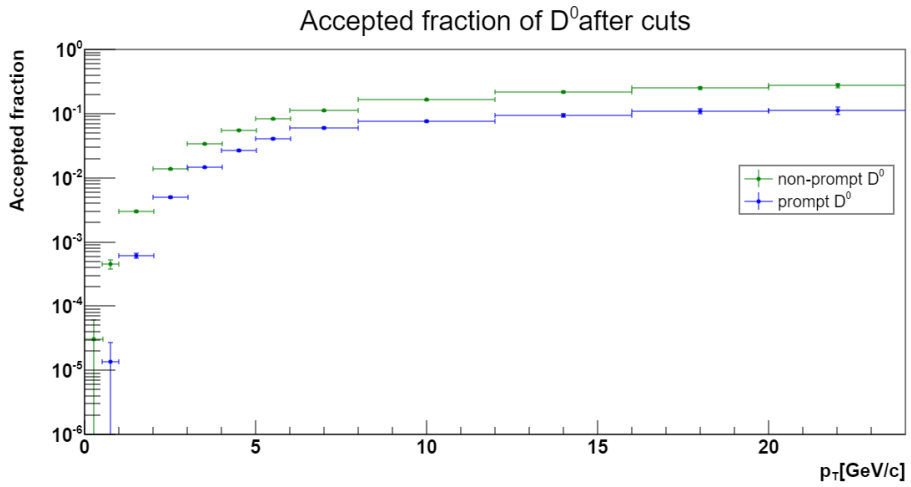
Table 1: The standard selection of topological variable cuts in the $0 < p_T < 6$ GeV/c interval.

Variable	Type	[6,8] GeV/c	[8,12] GeV/c	[12,16] GeV/c	[16,20] GeV/c	[20,24] GeV/c
$\|M\text{-}M_{D^0}\|$ [GeV/c^2]	$<$	0.3	0.3	0.3	0.35	0.3
dca [cm]	$<$	0.03	0.03	0.03	0.03	0.03
$\cos(\Theta^*)$	$<$	1.0	0.8	0.8	0.8	0.8
p_{T_K} [GeV/c]	$>$	0.8	0.8	0.8	0.8	0.8
p_{T_π} [GeV/c]	$>$	0.8	0.8	0.8	0.8	0.8
$\|d_{0_K}\|$ [cm]	$<$	0.1	0.1	0.1	0.1	0.1
$\|d_{0_\pi}\|$ [cm]	$<$	0.1	0.1	0.1	0.1	0.1
$d_{0_K} d_{0_\pi}$ [cm^2]	$<$	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004
$\cos(\Theta)$	$>$	0.9	0.8	0.9	0.9	0.9
$\cos(\Theta_{XY})$	$>$	0.998	0.998	0.998	0.998	0.998
NDL_{XY}	$>$	5	5	5	5	5

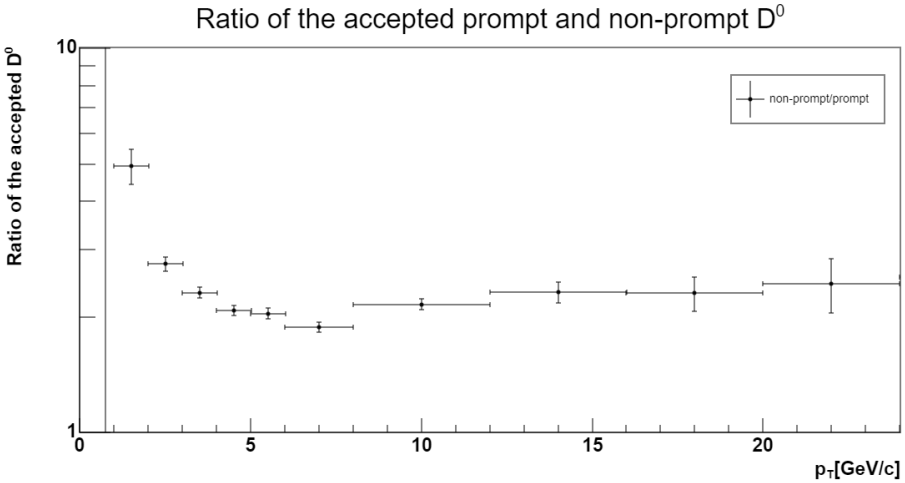
Table 2: The standard selection of topological variable cuts in the $6 < p_T < 24$ GeV/c interval.

5.2 Baseline - Validation

Using these cuts the efficiency for the different selection steps can be determined. For this thesis we are only interested in the accepted prompt and non-prompt D^0 fractions after all selection steps. This means both track and topological cuts are evaluated. Figure 16 shows the accepted fractions of prompt and non-prompt D^0 after passing track selection cuts and the topological cuts and the ratio of the accepted fractions. We can see that the non-prompt fraction is higher than the prompt fraction for each p_T bin. For values $p_T > 8$ GeV/c the acceptance of both prompt and non-prompt fractions barely increase and the ratio between the prompt and non-prompt fraction becomes more or less stable. The accepted non-prompt fraction is larger in the baseline because it is less challenging to reconstruct the D^0 via the daughter tracks for the non-prompt D^0. It is less challenging because the non-prompt D^0 decays further away from the primairy vertex than the prompt D^0. Further away from the primairy vertex it is easier to recombine the daughter tracks for non-prompt D^0 mesons which are required for the reconstruction.



(a) Accepted fractions



(b) Ratio of the accepted fractions

Figure 16: Top: Accepted fraction of prompt and non-prompt D^0 in the interval $0 < p_T < 24$ GeV/c after passing track and topological cuts. Bottom: Ratio of the accepted prompt and non-prompt fractions for the interval $0 < p_T < 24$ GeV/c.

The next step in the baseline analysis is to use these cuts in the invariant mass analysis. First we performed the invariant mass analysis in the region $0 < p_T < 24$ GeV/c for forced MC pp collisions at $\sqrt{s} = 13$ TeV. The amount of events analysed is $N_{events} = 7.4 \times 10^6$. Figures 17 and 18 show the results from the invariant mass analysis on this forced MC sample. In each p_T interval a well-defined peak occurs and the significances are very high (6.1 ± 0.4 to 74.5 ± 0.4). These very large significances (> 5) give the indication that when the analysis is now performed on real data mass peaks should occur even tough in real data the amount of background D^0 is significantly higher.

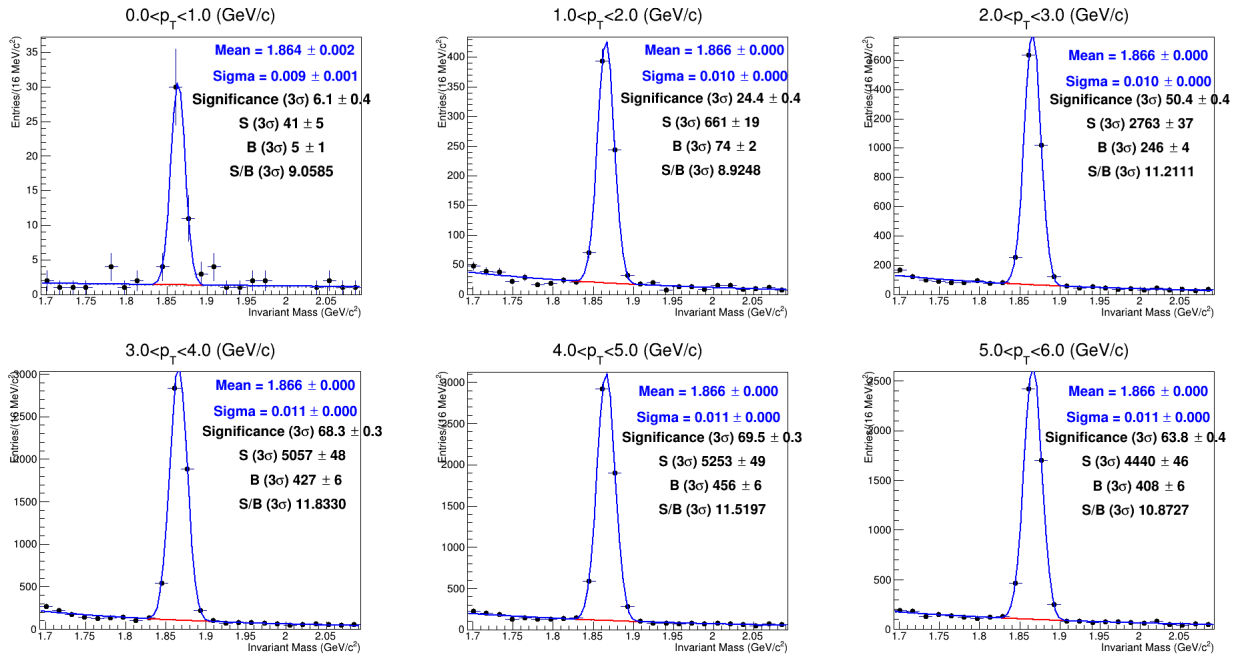


Figure 17: The reconstructed invariant mass in forced MC pp collisions at $\sqrt{s} = 13$ TeV obtained using standard cuts in the interval $0 < p_T < 6$ GeV/c.

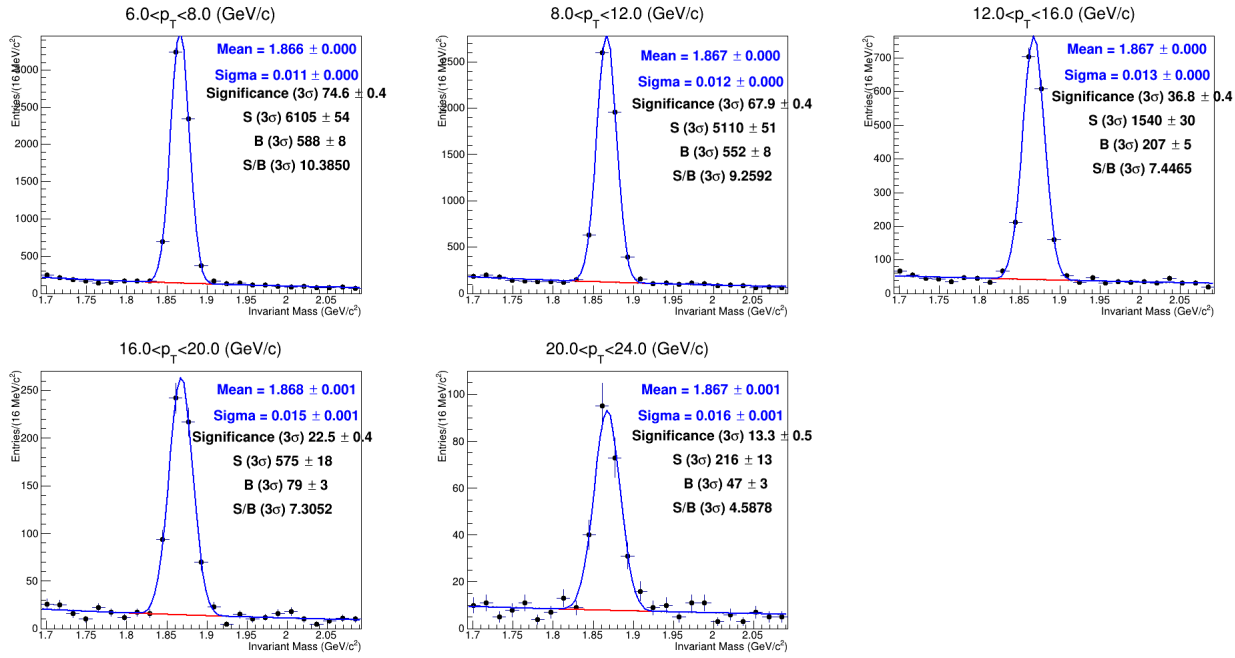


Figure 18: The reconstructed invariant mass in forced MC pp collisions at $\sqrt{s} = 13$ TeV obtained using standard cuts in the interval $6 < p_T < 24$ GeV/c.

5.3 Baseline - Implementation

The last step in the baseline analysis is use the standard cuts to perform the analysis on real data. In total 6.2×10^8 $\sqrt{s} = 13$ TeV pp events were analysed using the minimum bias trigger, given by the VZERO system, and the standard cuts. We used the data samples discussed in section 4.4. Figures 19 and 20 show the invariant mass peaks in the interval $0 < p_T < 24$ GeV from these analysed events. In the interval $0 < p_T < 1$ GeV/c the signal was not visible and therefore the invariant mass could not be determined and the points could not be fitted. The significances of the peaks are between 4.5 ± 0.9 and 29.0 ± 0.7. The only interval where the peak has a significance less than 5.0 is in the interval $20 < p_T < 24$ GeV/c, where the significance is 4.5 ± 0.9.

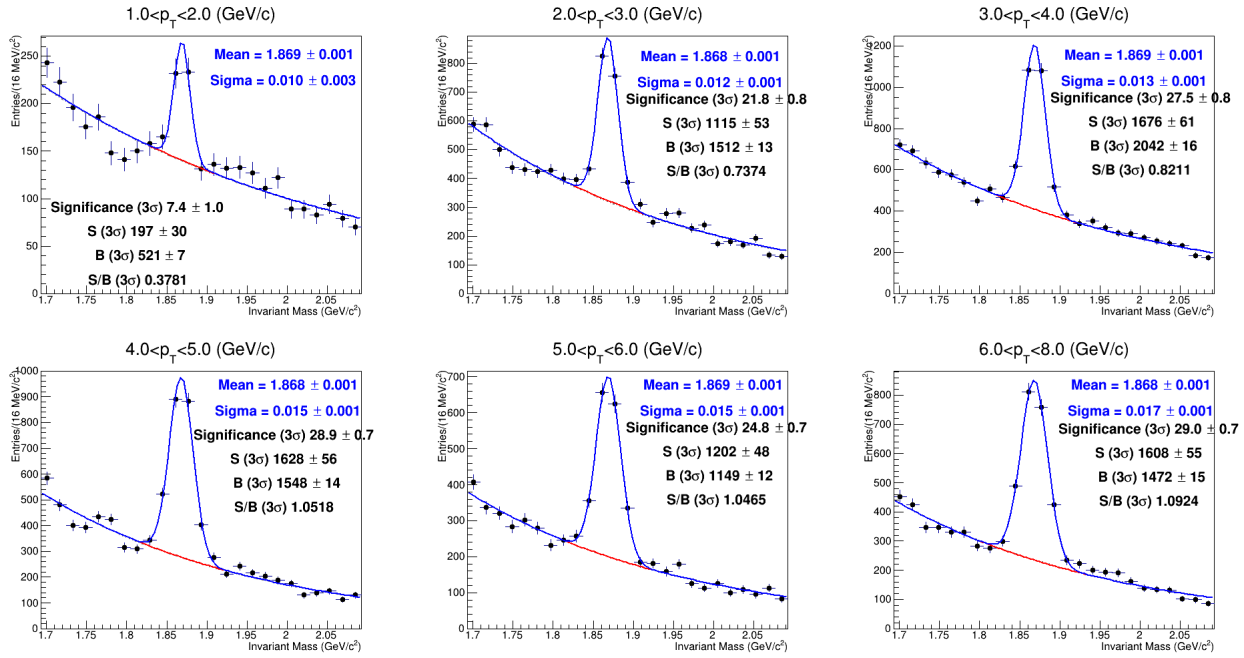


Figure 19: The reconstructed invariant mass in pp collisions at $\sqrt{s} = 13$ TeV obtained using standard cuts in the interval $0 < p_T < 8$ GeV/c.

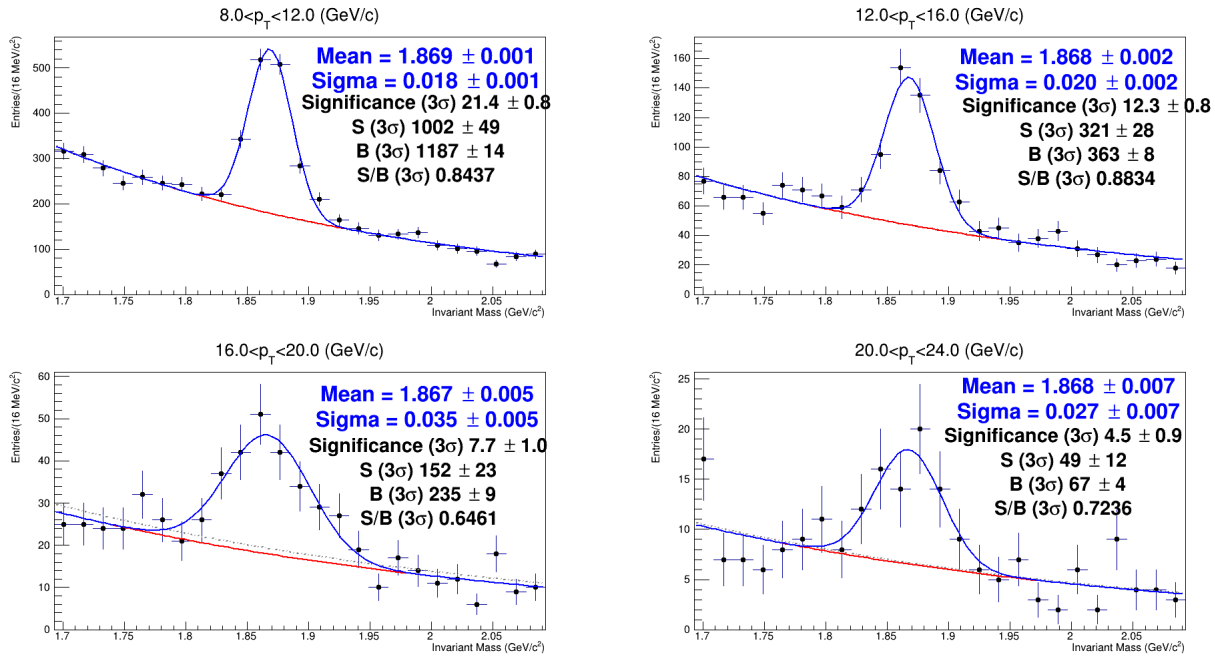
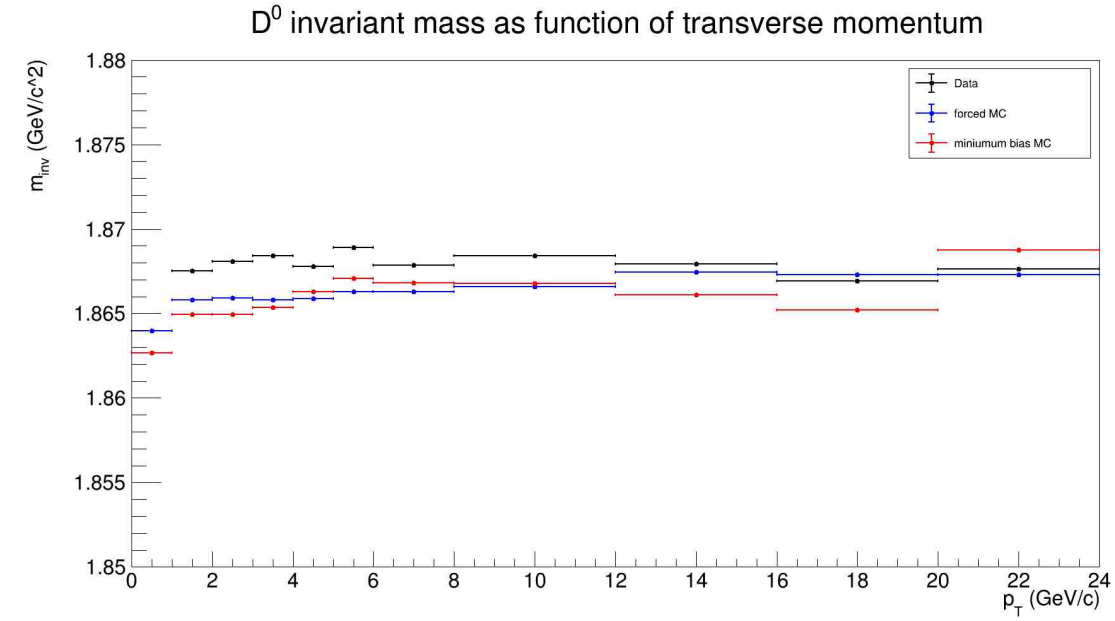
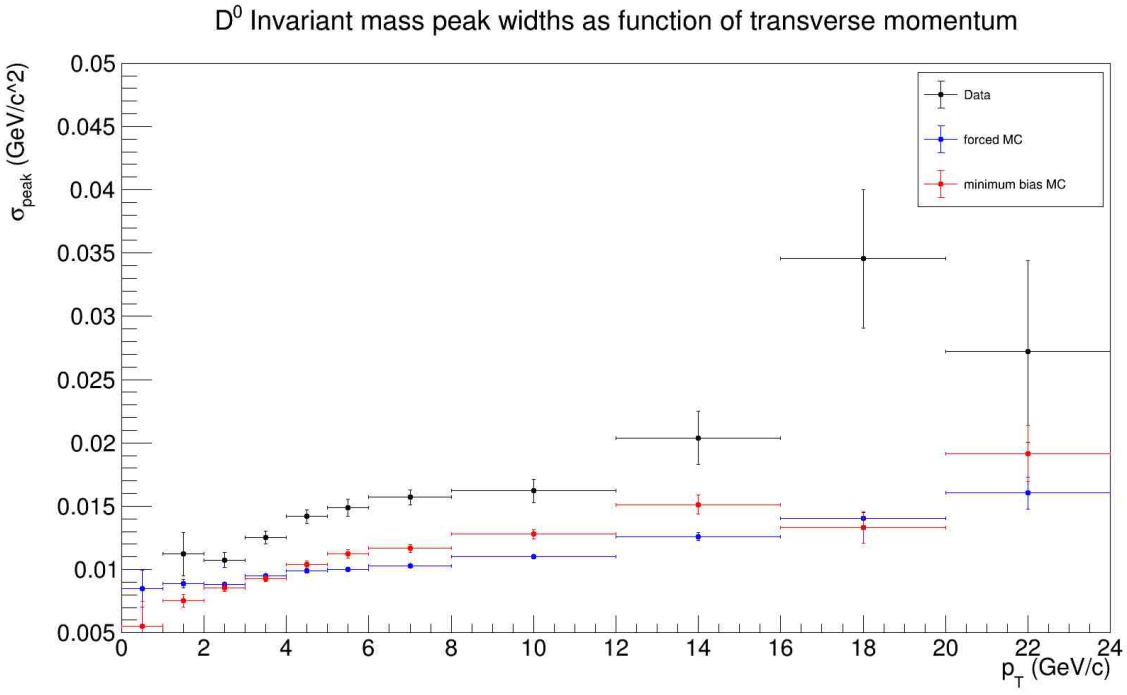


Figure 20: The reconstructed invariant mass in pp collisions at $\sqrt{s} = 13$ TeV obtained using standard cuts in the interval $8 < p_T < 24$ GeV/c.

Figure 21 shows a comparison of the invariant masses and peak widths between the forced MC sample, which is also used for validation, a minimum bias MC sample and data. Here we see that both MC samples agree. The fact that both MC samples agree means that even though the validation sample is forced it still has realistic physical features. Furthermore the figures show that the widths of the peaks are broader in data compared to the simulation. This effect is known within the ALICE collaboration and it is due to an interplay between the resolution effects and the misalignment that are not perfectly reproduced in the Monte-Carlo.



(a)



(b)

Figure 21: Top: Comparison of the invariant masses obtained from data, forced MC (validation sample) and minimum bias MC in the interval $0 < p_T < 24$ GeV/c. Bottom: Comparison of the invariant mass peak widths obtained for data, forced MC and minimum bias MC in the interval $0 < p_T < 24$ GeV/c.

It is important that the mass peaks are fitted carefully, especially in data. The simulations reproduce the invariant mass reconstruction perfectly such that the invariant mass distribution will peak around the invariant mass of the D^0 meson and therefore an accurate fit is practically guaranteed. However this is not necessarily true in data and one has to be careful that the fit corresponds with the data points. The invariant mass peaks are fitted using two methods. The first method uses an exponential + Gaussian fit. The second method is solely an exponential fit which only fits the exponential background data points far from the invariant mass peak. Figure 22 shows the relative fit accuracy between the 2 fit methods and the actual amount of counted signal. The first method, indicated in red, uses the exponential component of the full exponential + Gaussian fit as S_{fit}. The fit would be perfect if the relative difference is 0. In this figure we can see that for in every p_T interval both fit methods have a similar performance within uncertainty. In the interval $0 < p_T < 16$ the fits are quite accurate with the relative difference being smaller than 20%. Above $p_T > 16$ GeV the fits become less accurate which could be explained by the lack of sufficient statistic in those intervals. For each p_T interval both fitting methods agree within uncertainty which shows that the exponential component of both methods are almost identical and therefore the Gaussian component of the full fit solely fits the peak.

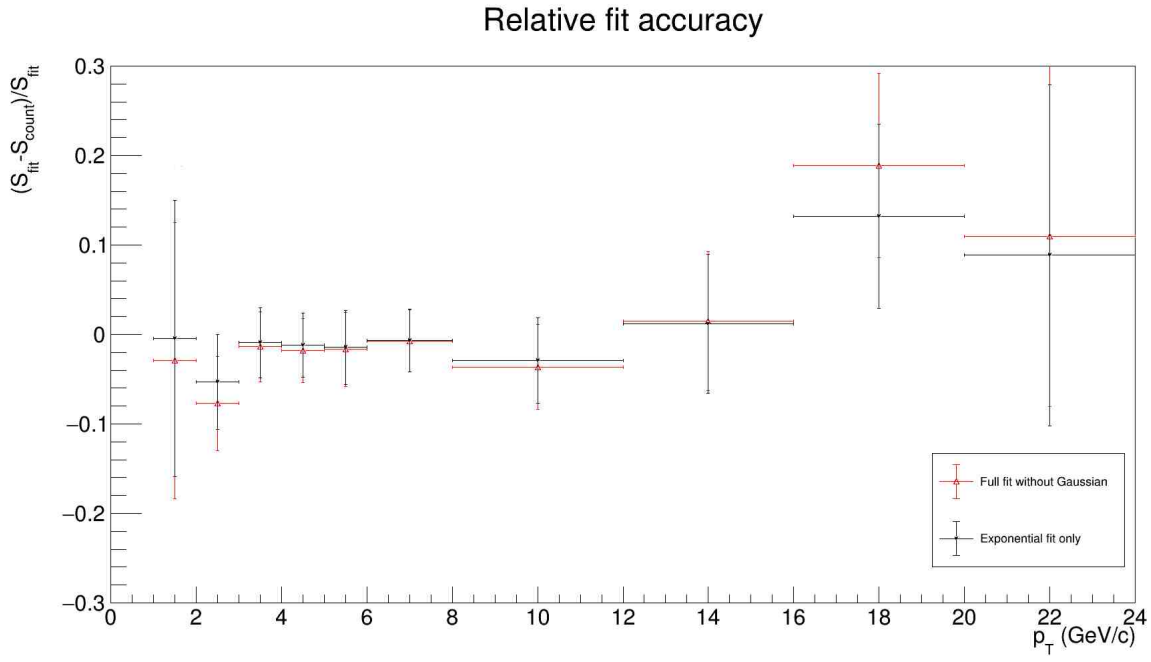


Figure 22: Relative difference between exponential components of the 2 fitting methods and the amount of counted D^0 mesons in the interval $0 < p_T < 24$ GeV/c. The red points represent the exponential component of the full fit while the black points represent the exponential fit performed on points far from the peak.

6 Machine learning analysis

This section describes all the steps that were performed in the machine learning analysis and the results of each step. We start with the discussion of the samples that are used to train the machine learning algorithms an. Then we will discuss the training of both the boost decision tree and the convolutional neural network for each p_T interval. Afterwards we compare the training results to chose the algorithm with which we want to proceed analysis. We validate the best performing algorithm on the same MC sample that was used in the baseline to check whether the algorithm performs well on a dataset which is not the training set. We chose to use the same MC sample here and in the baseline to make a comparison in section 7. After we discuss the validation we are going to discuss the results of the implementation of the BDT in the invariant mass analysis for the real data set to finalise this section.

6.1 Variable Distributions

The first step in the machine learning analysis is to prepare the training samples. The variables in [14] were used as a starting point for our variable selection. On top of that we chose to add the normalised decay length in 3 dimensions and the decay length in both 2 and 3 dimensions as training variables. This results in a total of 13 topological training variables. These variables are discussed in section 4.1. Since the analysis proceeds via invariant mass reconstruction the invariant mass is not used as a training variable. In this section we will discuss the samples that will be used in the machine learning analysis. The candidates are stored in ROOT trees in which all values of the variables are linked back to a single candidate. A separate tree is made for prompt, non-prompt and background D^0 mesons in a specific p_T interval.

The machine learning algorithms will be trained to separate non-prompt D^0 mesons from prompt D^0 mesons such that the prompt fraction reduces and the non-prompt fraction increases. The prompt and non-prompt D^0 mesons and their values of the topological variables are given to the algorithms by forwarding the ROOT trees to the algorithms via the TMVA dataloader. The normalized variable distributions of prompt (background) and non-prompt (signal) D^0 in the interval $2 < p_T < 3$ GeV/c can be seen in figures 23 and 24. Figures 25 and 26 show the normalized variable distributions for the interval $12 < p_T < 16$ GeV/c. These distributions were obtained by filling the ROOT trees with D^0 mesons from the forced MC samples described in section 4.4. In these figures we can see that the separation between prompt and non-prompt is smaller for D^0 with low transverse momentum. This is especially clear for the (normalized) decay variables in both 2D and 3D. Figure 27 shows a direct comparison of the decay length of prompt and non-prompt between the $2 < p_T < 3$ GeV/c and $12 < p_T < 16$ GeV/c intervals. From this figure it becomes very clear that the separation is smaller for low transverse momenta especially after we note that the x-axis have different ranges. It is important to note that the increase in separation for high transverse momenta is not the same for each variable. By comparing the p_{T_K} and p_{T_π} in figures 23 and 25 we see that the x-ranges are very different but in both bins the prompt and non-prompt distributions greatly overlap. There should not be a difference in the momentum of the daughters from a prompt and non-prompt D^0 meson which have the same momentum since

momentum is conserved and hence this observation was expected. The decay lengths have a larger separation because a non-prompt D^0 decays further from the primairy vertex because it originates from a parent which also decayed some distance away from the primairy vertex. The distance travelled by the parent is larger when the transverse momentum is higher and hence the separation is higher for high transverse momentum.

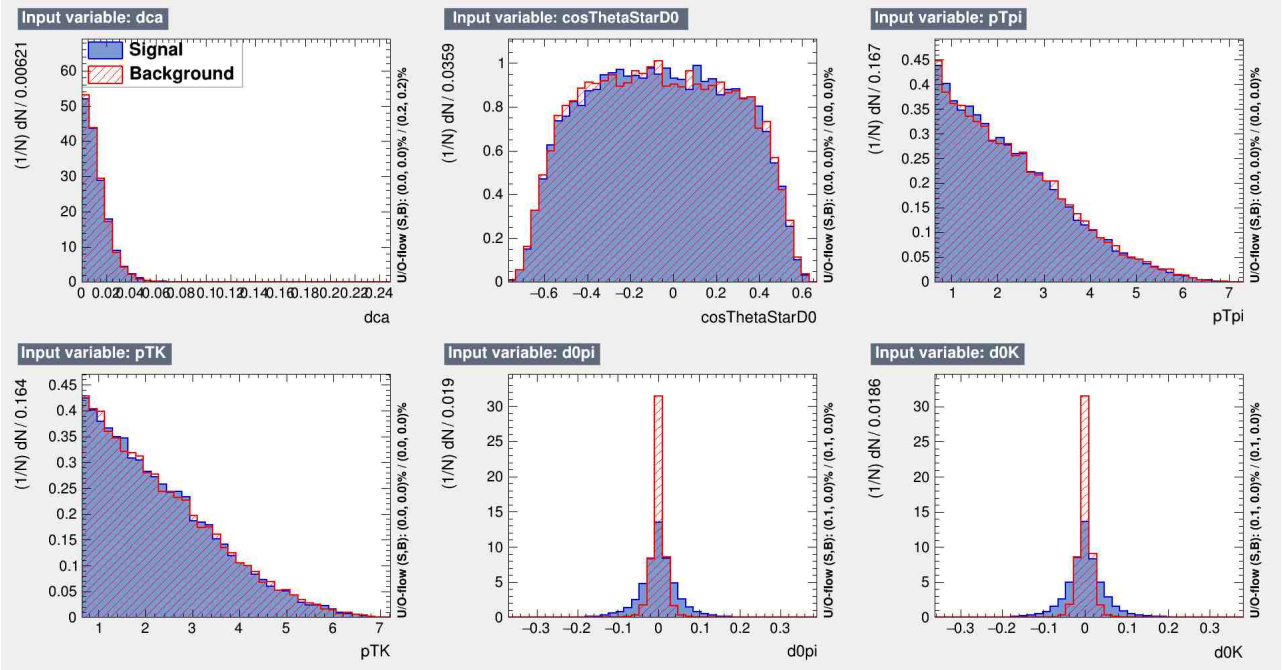


Figure 23: Normalized variable distributions of prompt and non-prompt D^0 mesons with $2 < p_T < 3$ GeV/c.

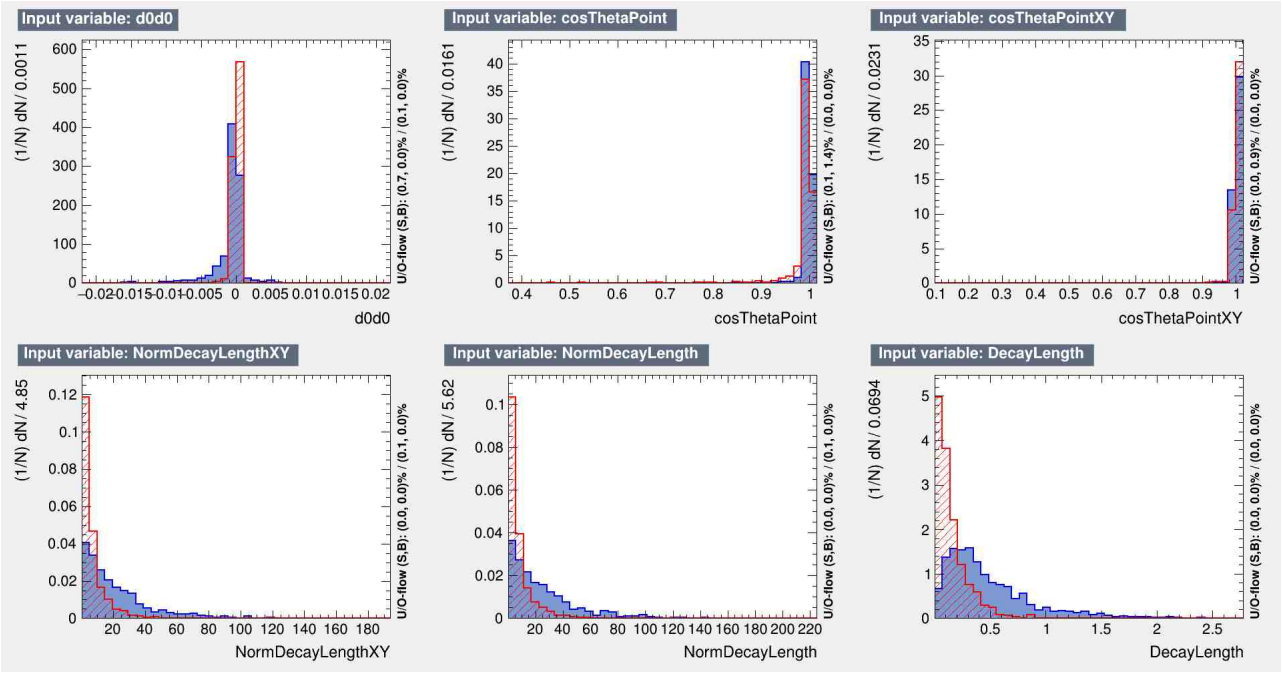


Figure 24: Normalized variable distributions of prompt and non-prompt D^0 mesons with $2 < p_T < 3$ GeV/c.

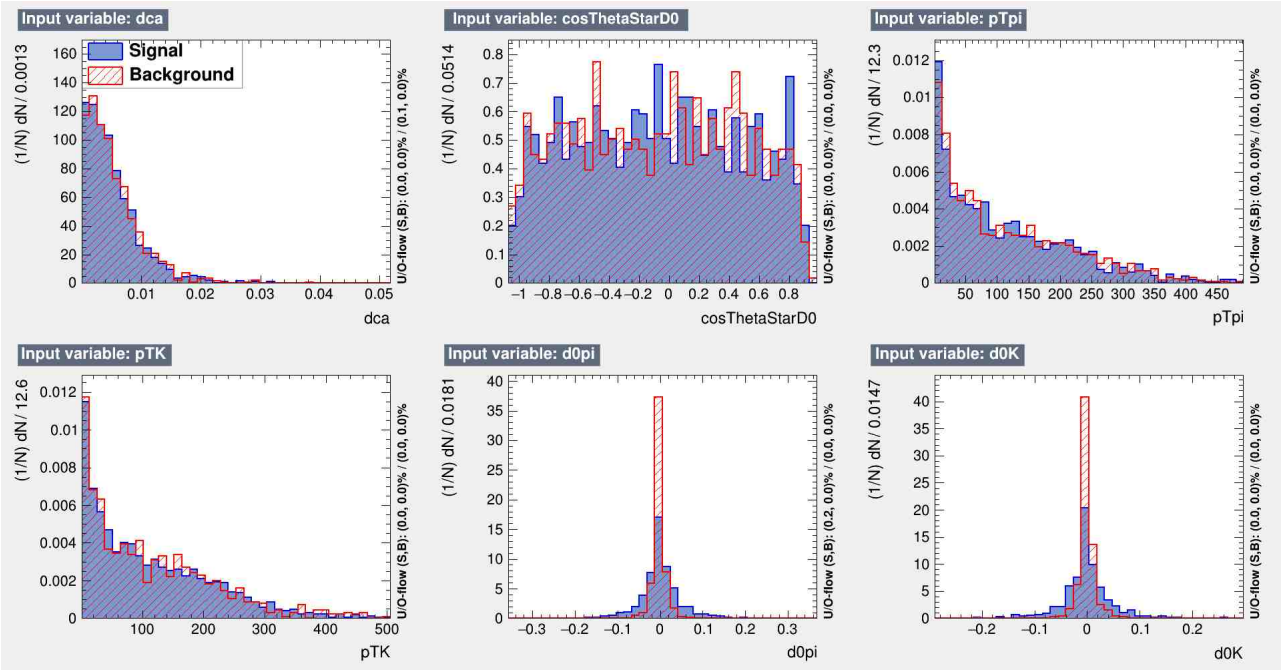


Figure 25: Normalized variable distributions of prompt and non-prompt D^0 mesons with $12 < p_T < 16$ GeV/c.

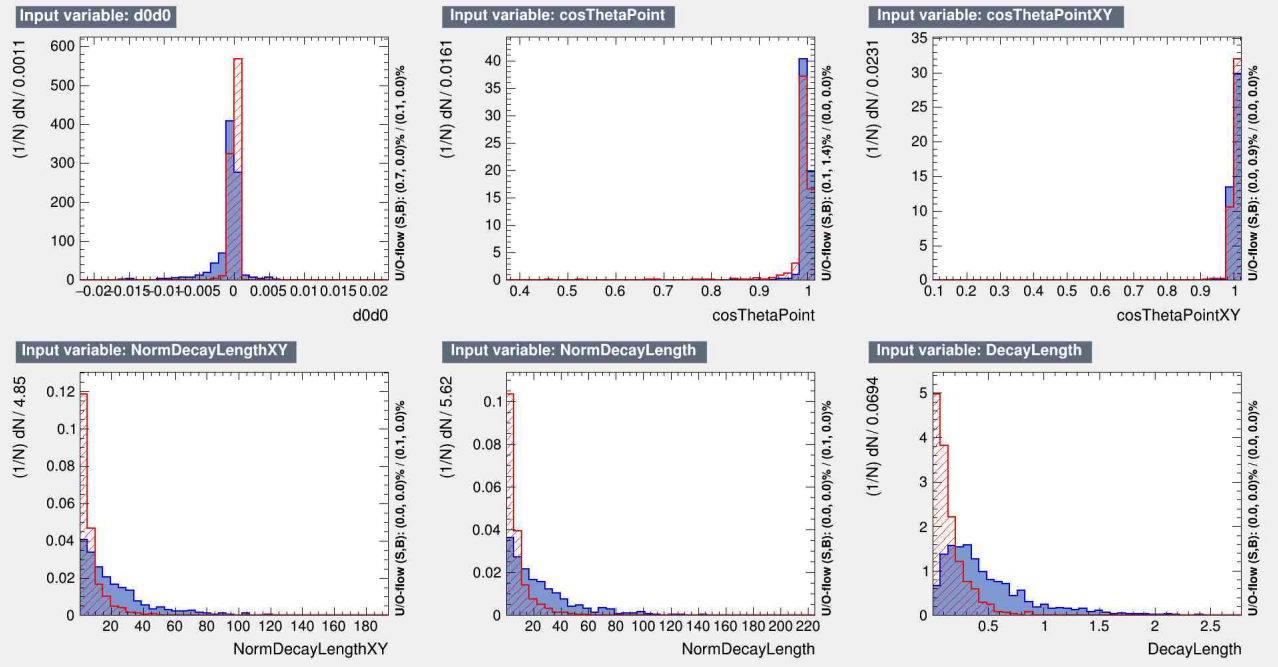
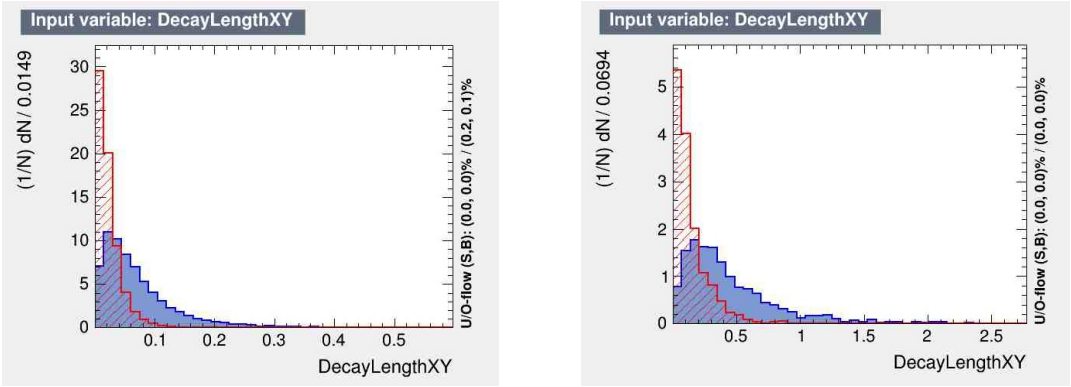


Figure 26: Normalized variable distributions of prompt and non-prompt D^0 mesons with $12 < p_T < 16$ GeV/c.



(a) $2 < p_T < 3$ GeV/c (b) $12 < p_T < 16$ GeV/c

Figure 27: Comparison between the decay length of prompt (red) and non-prompt (blue) D^0 in the intervals $2 < p_T < 3$ GeV/c and $12 < p_T < 16$ GeV/c.

After reviewing the distributions shown in this section it is estimated that the machine learning will be more difficult in the lower p_T regions while we expect better results for the higher p_T regions due to the large separation in the distributions.

6.2 Algorithm training

For our studies we use boost decision trees of type AdaBoost and convolutional neural networks to reduce the prompt D^0 fraction in our invariant mass analysis. In contrast to the baseline analysis candidates are now selected solely based on algorithm output. If the output value of the algorithm is greater than the respective cut value the candidate is selected. Therefore the original 11 selection variables used for the baseline will be substituted by a single selection based on the response of the learning model. In this section we first discuss the settings, configuration and training of the BDT. Then we discuss the same topics for the CNN. After the training of the algorithms we compare them to select the best performing one. For the selected model we study the effect on the invariant mass reconstruction for forced MC and data. In section 7 we compare its performance with respect to the baseline of this thesis.

6.2.1 Training - Boost Decision Tree

The first algorithm that we trained to separate prompt and the non-prompt D^0 mesons in our invariant mass analysis are the boost decision trees with AdaBoost as boosting type. Table 3 shows the used TMVA settings. In each p_T interval the same settings were used.

TMVA Setting	**Value**
NTrees	2000
MinNodeSize	2,5%
MaxDepth	2
BoostType	AdaBoost
UsedBaggedBoost	True
BaggedSampleFraction	0.5
SeparationType	GiniIndex
nCuts	-1

Table 3: ROOT TMVA settings used to train the boost decision trees.

For each of the 11 p_T intervals used a separate BDT was trained using the training set containing prompt and non-prompt D^0 stored a specific ROOT tree which was loaded using the TMVA dataloader. Figures 28 and 29 show the background rejection as function of signal efficiency for the different p_T intervals (ROC curves). In the optimal case the background rejection is 1 for every value of signal efficiency. We can see that for the BDTs in the intervals $0 < p_T < 1$ GeV/c and $1 < p_T < 2$ GeV/c, which are the red and black line respectively in figure 28, perform worse than the other BDTs because for the same values of signal efficiency these BDTs have lower background rejection compared to the BDTs trained on the interval $2 < p_T < 6$ GeV/c. This was expected because as we discussed in section 6.1 the separation between prompt and non-prompt D^0 mesons is smaller for lower transverse momentum. The background rejection is very similar for all BDTs trained on $p_T > 2$ GeV/c with only minor differences depending on signal efficiency.

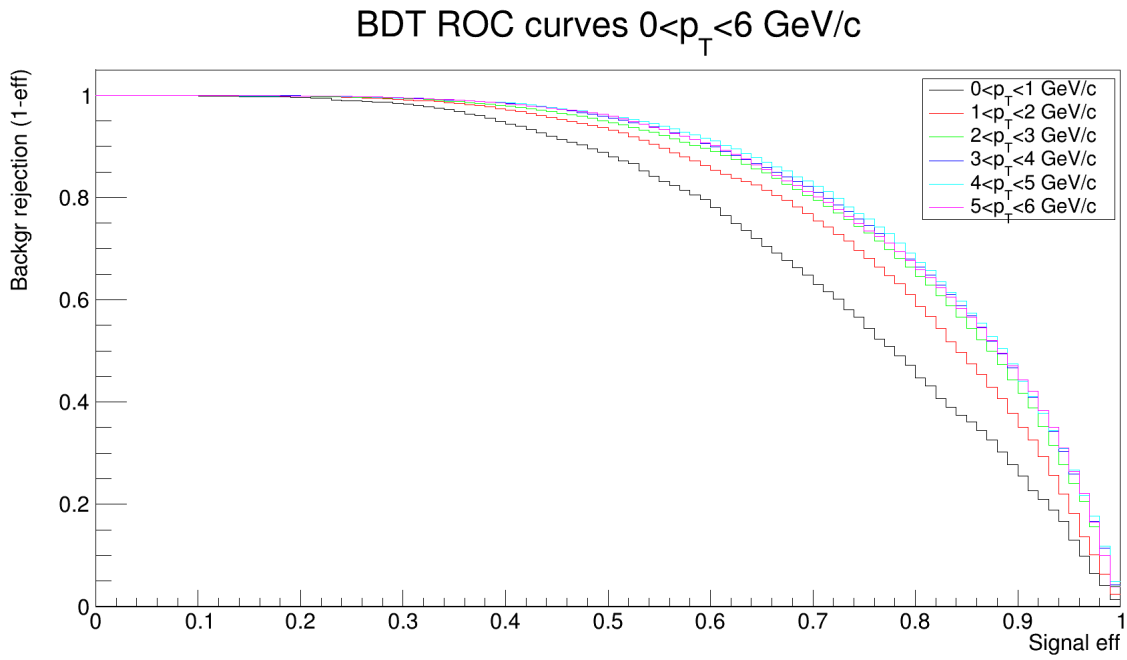


Figure 28: Background rejection as function of signal efficiency for the different BDTs in the interval $0 < p_T < 6$ GeV/c. The BDTs trained in the intervals $0 < p_T < 1$ GeV/c and $1 < p_T < 2$ GeV/c show a poorer performance compared to the other intervals.

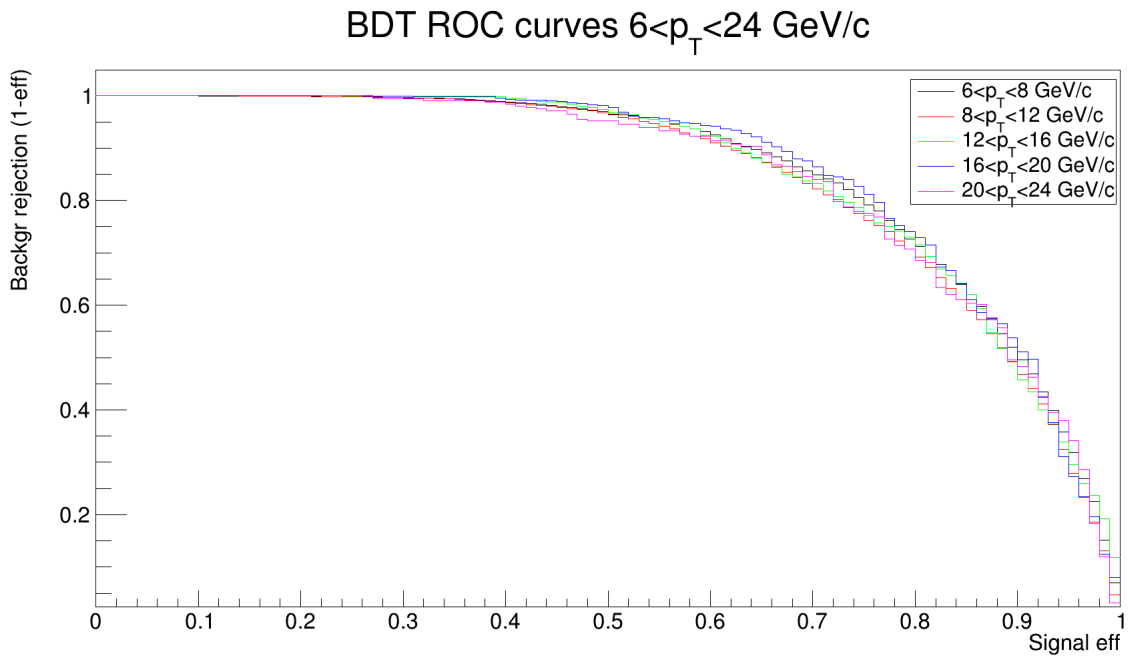


Figure 29: Background rejection as function of signal efficiency for the different BDTs in the interval $6 < p_T < 24$ GeV/c.

The TMVA package was used to select the optimal cut value for the algorithm. Since there is approximately 40 times more prompt than non-prompt D^0 in minimum bias data this ratio was taken into account during the selection of the optimal cut value. Table 4 shows the cut values for each p_T interval. Note that a cut value is dependent on the ratio between the amount of signal and background and corresponds to a fixed signal and background efficiency.

p_T **interval in GeV/c**	**BDT cut value for optimal significance**
$[0, 1]$	0.0991
$[1, 2]$	0.1458
$[2, 3]$	0.1297
$[3, 4]$	0.1229
$[4, 5]$	0.1456
$[5, 6]$	0.1380
$[6, 8]$	0.1114
$[8, 12]$	0.1105
$[12, 16]$	0.1394
$[16, 20]$	0.1395
$[20, 24]$	0.2210

Table 4: BDT cut values to obtain the optimal significance in the interval $0 < p_T < 24$ GeV/c.

6.2.2 Training - Convolutional Neural Network

The second type of machine learning algorithm studied in these thesis is a convolutional neural network. The specific layer configuration and the in- and output shapes of the model are illustrated in figure 30. It consists of 2 convolutional layers split by a max-pooling layer and 5 dense layers. The flatten layer is the transition between the convolutional part and the fully connected layers. The same layer configuration and settings were used in every p_T interval. Similar to the training of the BDTs the convolutional neural networks were also trained separately for each interval. The training sets are identical to the ones used for the BDT training. The specific settings are shown in table 5. The batchsize was set to 16 for the CNNs trained on $16 < p_T < 20$ GeV/c and $20 < p_T < 24$ GeV/c due to smaller train trees.

TMVA Setting	Value
Learning Rate	0.0005
Loss Function	MSE
Optimizer	Adam
N_{epochs}	50
Batchsize	32*

Table 5: ROOT TMVA settings used to train the convolutional neural networks. *For the CNNs trained on $p_T > 16$ GeV/c batchsize 16 was used instead of 32.

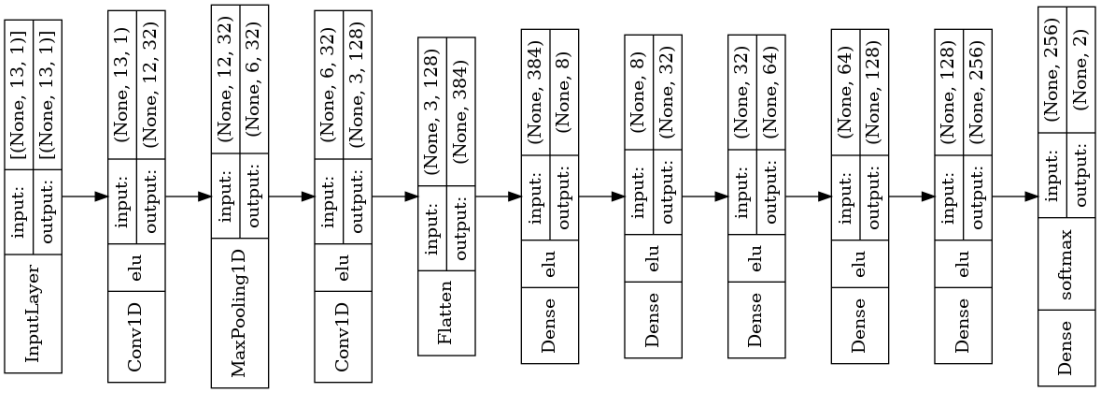


Figure 30: Illustration of the layers of the Keras model and their respective activation functions. Pooling and flatten layers do not have an activation function.

Figures 31 and 32 show the background rejection as function of signal efficiency for CNNs trained on the interval $0 < p_T < 24$ GeV/c. In figure 31 we see that the CNN trained on $0 < p_T < 1$ GeV/c performs worse than the other networks. The CNN trained on $1 < p_T < 2$ GeV/c performs slightly worse than the other networks. This is again due to the smaller separation of the variable distributions of prompt and non-prompt D^0 mesons. In figure 32 we see that the network trained on $16 < p_T < 20$ GeV/c has a higher background rejection for intermediate signal efficiency but for high signal efficiency it has a similar rejection to networks trained on other intervals.

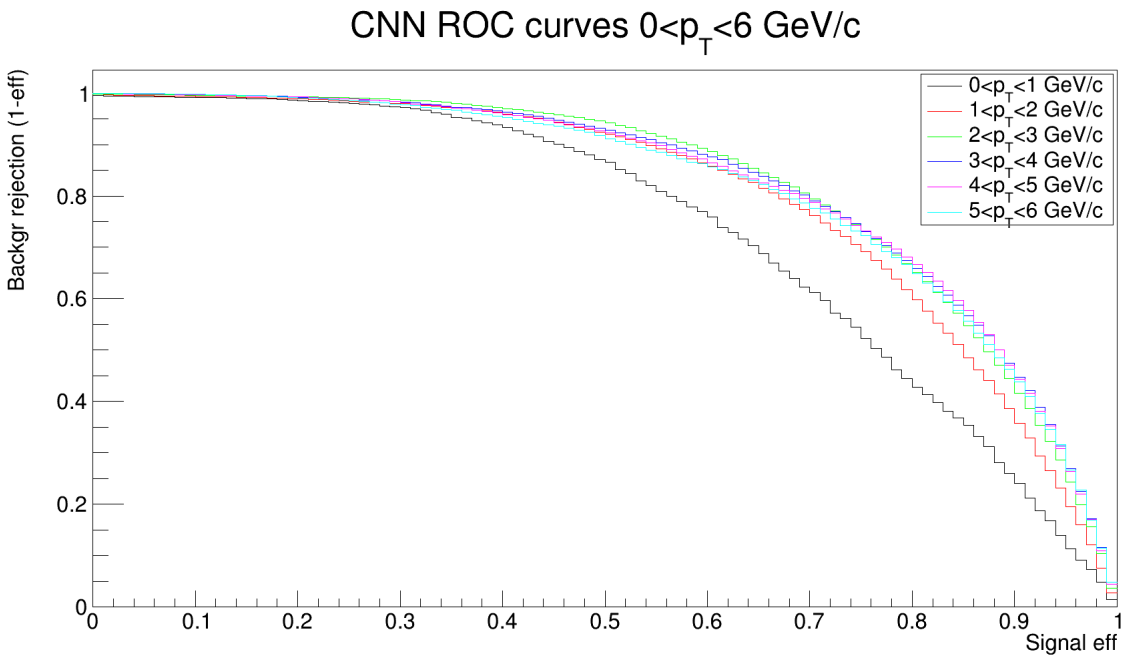


Figure 31: CNN ROC curves in the interval $0 < p_T < 6$ GeV/c.

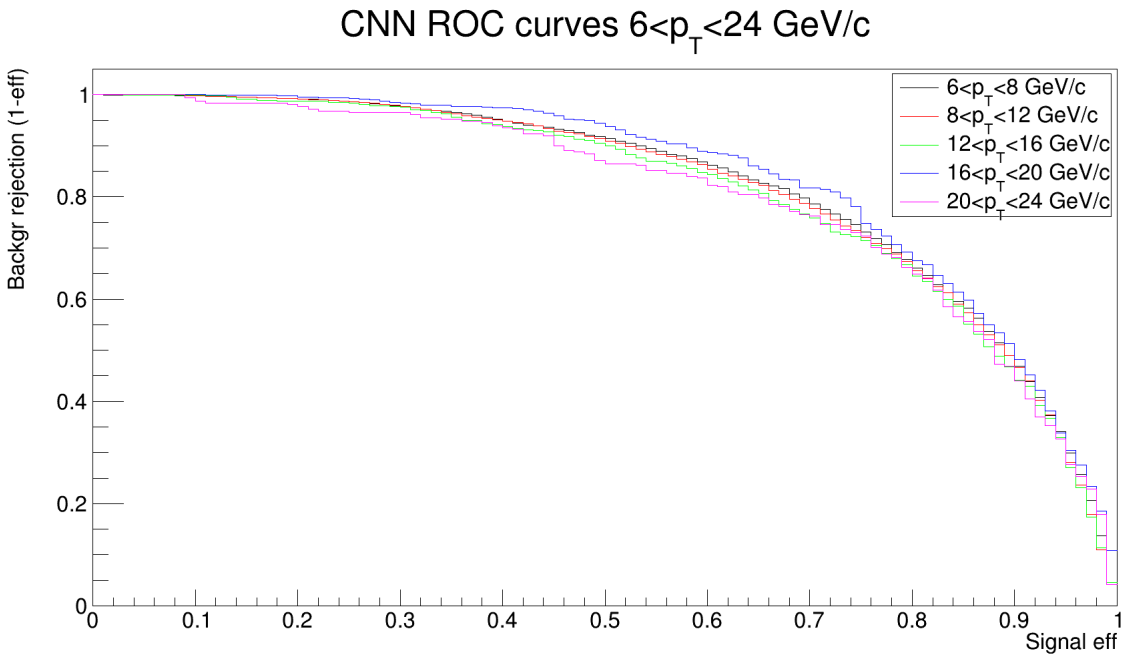


Figure 32: CNN ROC curves in the interval $6 < p_T < 24$ GeV/c.

Again the TMVA package was used to select the optimal cut value for the algorithm. The ratio of approximately 40 between prompt and non-prompt D^0 in minimum bias data was

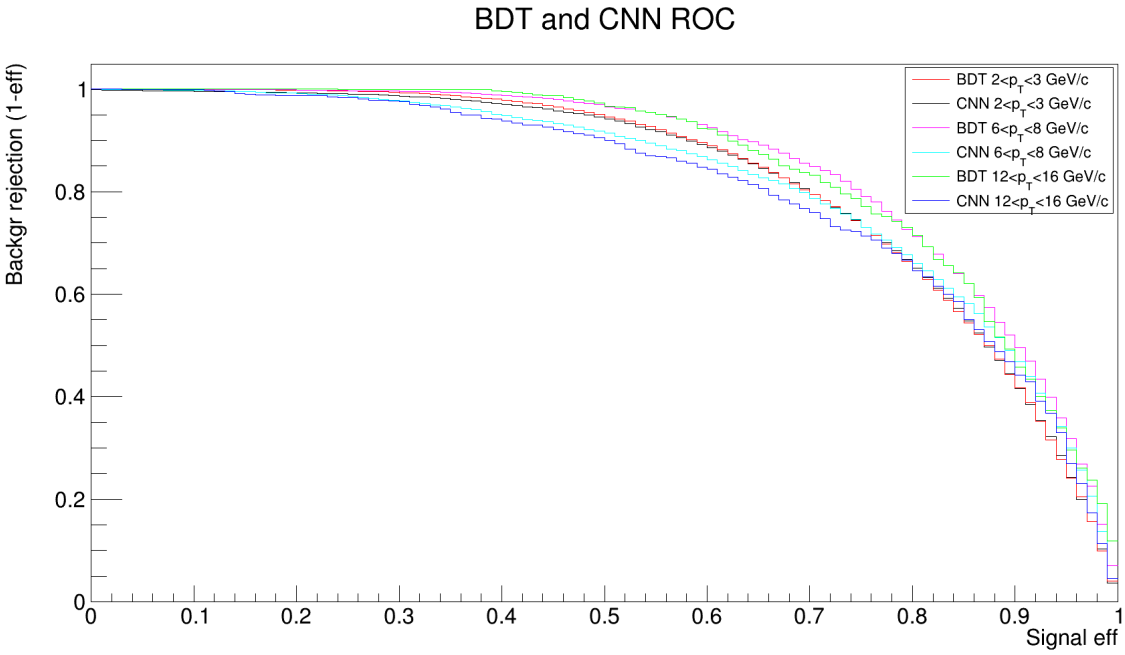
taken into account to select the optimal cut value. Table 6 shows the cut values for each p_T interval.

p_T **interval in GeV/c**	**CNN cut value for optimal significance**
$[0, 1]$	0.8243
$[1, 2]$	0.8616
$[2, 3]$	0.9329
$[3, 4]$	0.8809
$[4, 5]$	0.9046
$[5, 6]$	0.9379
$[6, 8]$	0.9111
$[8, 12]$	0.9077
$[12, 16]$	0.9630
$[16, 20]$	0.9150
$[20, 24]$	0.9795

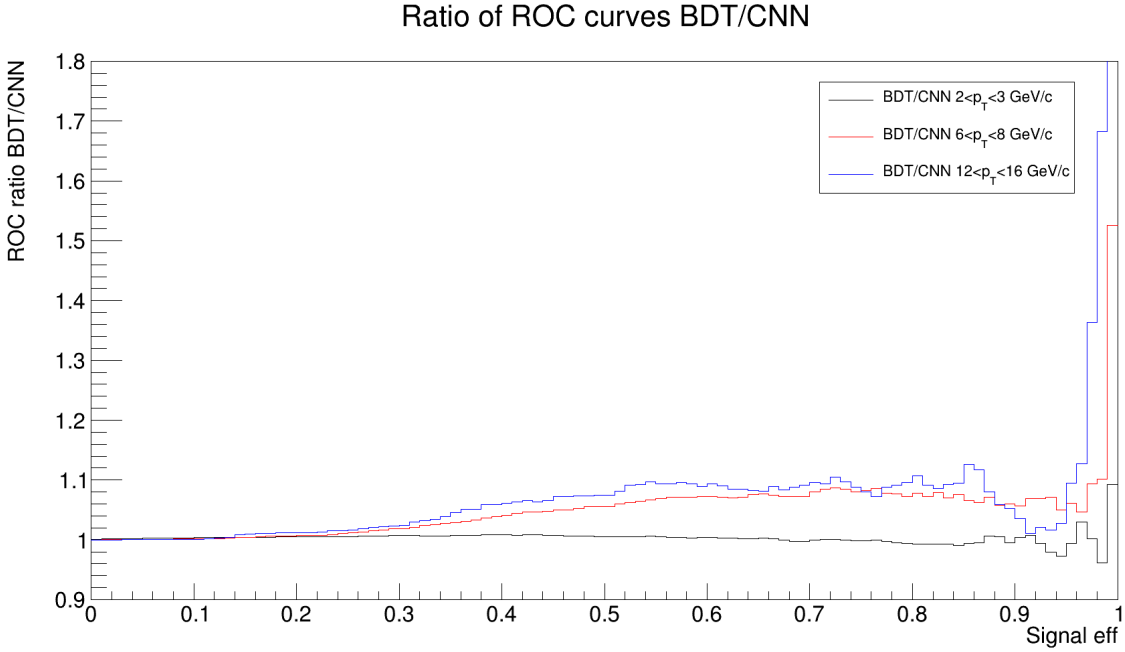
Table 6: CNN cut values to obtain the optimal significance in the interval $0 < p_T < 24$ GeV/c.

6.2.3 Algorithm comparison

In this section we will discuss and compare the training results of both the BDT and the CNN to select the most promising algorithm in terms of performance.
Figure 33 shows the ROC curves for three different (low, intermediate and high) p_T regions and the ratio between the ROC curves. In the top panel we see that in the region $2 < p_T < 3$ GeV/c the curves overlap. In the other 2 regions however we see that the BDT outperforms the CNN because for the same signal efficiency it rejects more background. From the bottom panel it becomes clear that there is a great overlap for $2 < p_T < 3$ GeV/c but for higher p_T the BDT rejects more background depending on signal efficiency. As we will see in table 7 the signal efficiencies we will use are roughly between 0.25 and 0.50 and for those signal efficiencies the BDT performs up to 12% better depending on the specific signal efficiency and p_T interval.



(a) BDT and CNN ROC curves



(b) Ratio of the ROC curves

Figure 33: Comparison of the ROC between BDT and CNN algorithms. Top plot shows the ROC of the BDT and CNN for three different p_T intervals and the bottom plot shows the BDT/CNN ratio as function of signal efficiency.

On top of the ROC curves the TMVA also provides information on the signal and background efficiency for a specific algorithm cut value. These efficiencies should not be confused with the

accepted fractions of prompt and non-prompt D^0 since the accepted fractions are also affected by track cuts and detector layer acceptance. Using the cut values for optimal significance from tables 4 and 6 we can make a comparison of the respective signal and background efficiencies using these algorithm cut values in different p_T intervals. Table 7 shows the efficiency for prompt (B_{eff}) and non-prompt (S_{eff}) candidates in a low, intermediate and high p_T interval for both machine learning algorithms. We can see that for $2 < p_T < 3$ GeV/c the signal efficiency of the CNN is approximately 5% higher when using the cut for optimal significance. However the background efficiency of the CNN is almost 3 times higher, which means there is less background rejection. So even though the signal efficiency is higher for the CNN in this interval the best performing algorithm is the BDT. For the other 2 p_T intervals the BDT performs better when cutting for optimal significance since the signal efficiency is higher than that of the CNN and the background efficiency is lower or similar.

p_T **interval in GeV/c**	$\mathbf{S}_{eff}$ **BDT**	$\mathbf{S}_{eff}$ **CNN**	$\mathbf{B}_{eff}$ **BDT**	$\mathbf{B}_{eff}$ **CNN**
$[2, 3]$	0.2901	0.3440	0.0052	0.0174
$[6, 8]$	0.3389	0.2279	0.0052	0.0098
$[12, 16]$	0.3868	0.1084	9.26×10^{-4}	9.26×10^{-4}

Table 7: BDT cut values to obtain the optimal significance in the interval $0 < p_T < 24$ GeV/c.

By combining the observations of the comparison of the ROC curves and the efficiencies it becomes clear the BDT generally performs better, although in some cases the performance could be similar since the ROC curves of the low transverse momentum overlap. In our case where we aim to cut the algorithms such that we achieve maximum significance the BDT has a better performance in all 3 discussed p_T regions. We proceed our studies by implementing the BDTs in our invariant mass analysis to replace the standard topological variable cuts.

6.2.4 Validation - Boost Decision Tree

The next step is to evaluate the performance of the BDT, which showed to be to most promising algorithm in the previous section. This is done by studying the effect of the BDT on a forced MC validation sample. This is the same sample that was used to evaluate the standard cuts in section 5 and is different from the sample that was used for algorithm training. First the accepted prompt and non-prompt D^0 fractions of the forced MC sample after track cuts and the BDTs are evaluated. In figure 34 we see that the non-prompt fractions are in the order of 10^{-1} in the interval $5 < p_T < 24$ GeV/c. For lower transverse momentum the accepted fractions are lower for both prompt and non-prompt D^0 which is due to the track cuts which were not modified in this thesis.

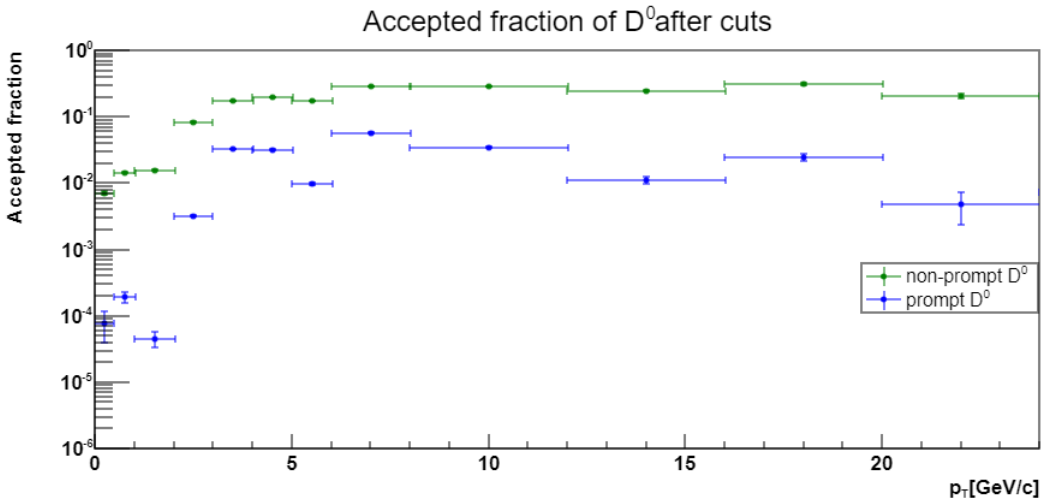


Figure 34: Accepted fractions of prompt and non-prompt D^0 in the forced MC sample in the interval $0 < p_T < 24$ GeV/c.

Figures 35 and 36 show the results of the invariant mass analysis on this MC sample and contains $N_{events} = 6.4 \times 10^6$ events. The amount of events is slightly lower than in the baseline due to temporal difference in successful runs on the ALICE Grid. Again we see very distinct mass-peaks in every p_T interval with significances ranging from 7.8 ± 0.5 to 70.1 ± 0.8.

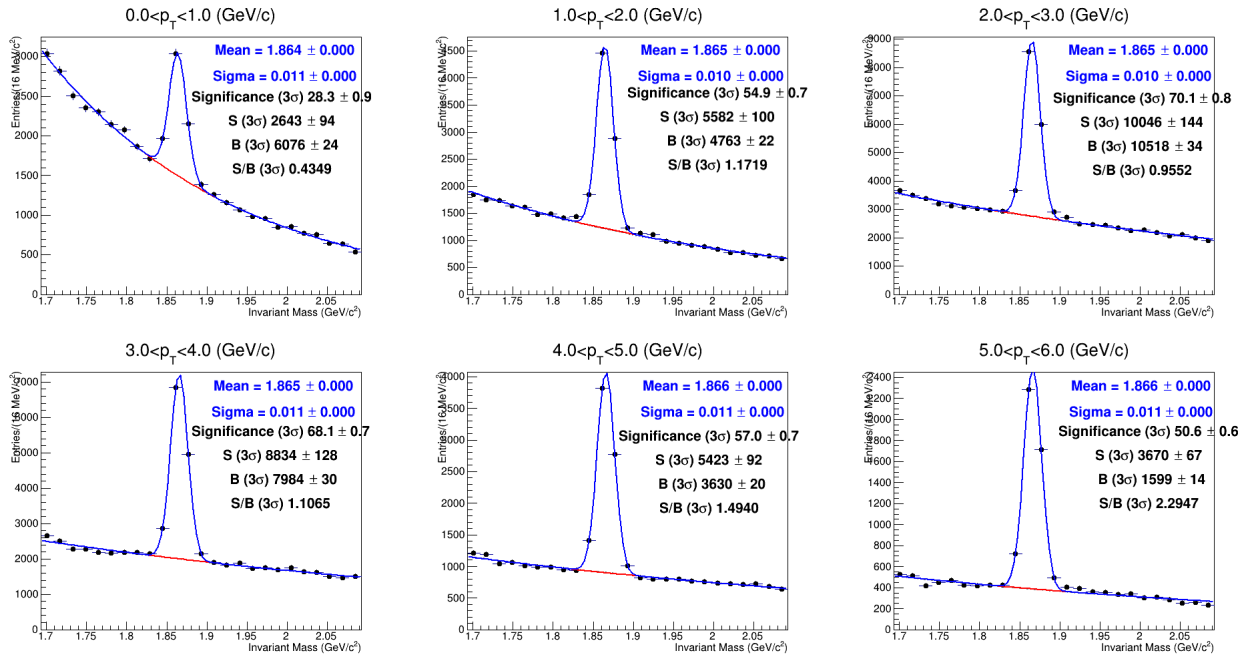


Figure 35: Invariant mass peaks obtained using forced MC pp collisions at $\sqrt{s} = 13$ TeV combined with the BDT cuts from table 4 in the interval $0 < p_T < 6$ GeV/c.

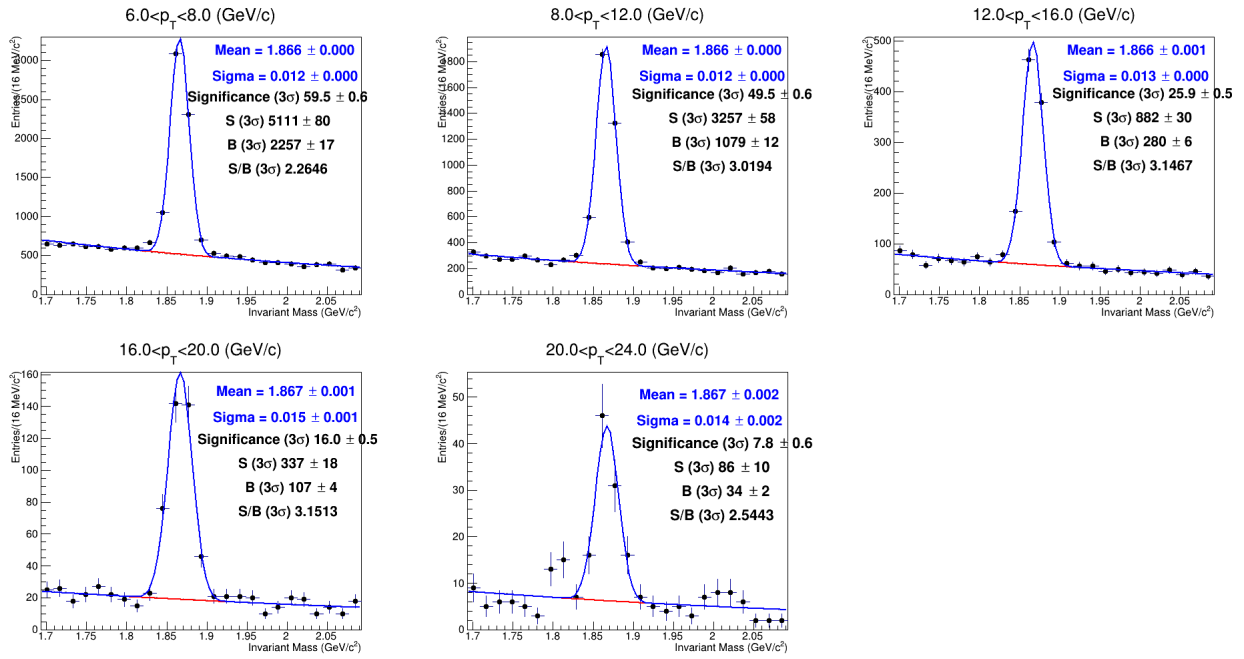


Figure 36: Invariant mass peaks obtained using forced MC pp collisions at $\sqrt{s} = 13$ TeV combined with the BDT cuts from table 4 in the interval $6 < p_T < 24$ GeV/c.

6.2.5 Implementation - Boost Decision Tree

The last step is to study the effect of using the boost decision trees and the corresponding cuts from table 4 on real LHC data. Figure 35 shows the reconstructed invariant mass in the interval $6 < p_T < 24$ GeV/c. To produce this figure 1.84×10^8 events were analysed which is significantly less than the number of events which was analysed in the baseline. Fewer events were analysed because of problems occurring while running the analysis with the implemented BDTs on the ALICE grid. This was due to using too many run numbers for a single master job which resulted in master jobs that were not fully submitted and therefore a lot of events were lost. As can be seen in the figure it was only possible to fit mass peaks in the interval $5 < p_T < 24$ GeV/c where the data from the interval $16 < p_T < 20$ GeV/c and $20 < p_T < 24$ GeV/c had to be combined to allow for a fit. The significances are between 2.4 ± 0.8 and 6.3 ± 1.1. The fits show that implementing the BDT allows for the reconstruction of the invariant mass.

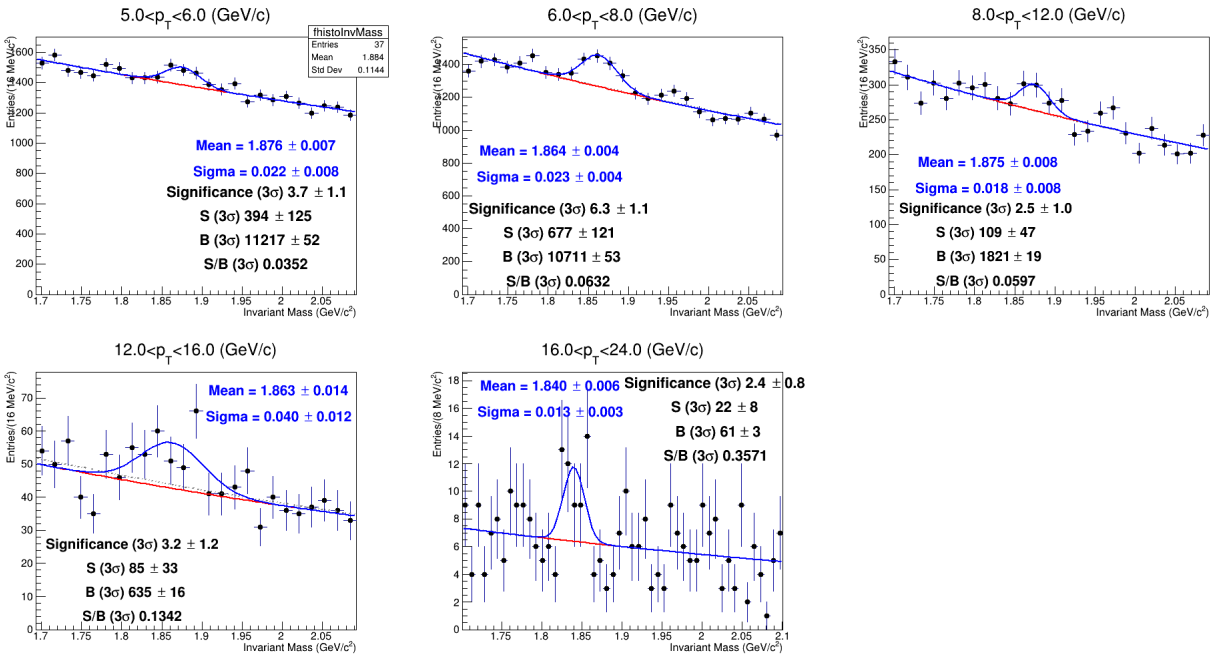


Figure 37: Invariant mass peaks obtained using LHC data from pp collisions at $\sqrt{s} = 13$ TeV combined with the BDT cuts from table 4 in the interval $5 < p_T < 24$ GeV/c.

Using that the significances scale with $\sqrt{N}$ we will be able to make a comparison with the baseline in section 7 once we scaled the significances. The scaling factor is $\sqrt{\frac{6.2\times10^8}{1.84\times10^8} \times 10^8} = 1.84$. In table 8 we can see that the significances in the intervals $6 < p_T < 8$ GeV/c and $12 < p_T < 16$ GeV/c are higher than 5. The other intervals have significances slightly below 5.0. With a simple calculation one can determine that the statistic required to obtain a significance of 5.0 in the interval $5 < p_T < 24$ GeV/c is achieved when $N = 8.0 \times 10^8$ events are analysed. The amount of available minimum bias events measured between 2016 and 2018 is larger than this number which means that it possible to obtain sufficient significances.

p_T interval in GeV/c	Obtained significance	Scaled significance
$[5, 6]$	3.7	4.8
$[6, 8]$	6.3	11.6
$[8, 12]$	2.5	4.6
$[12, 16]$	3.2	5.9
$[16, 24]$	2.4	4.4

Table 8: Scaled and unscaled significances obtained using the BDTs in the interval $6 < p_T < 24$ GeV/c. The scaling is performed to allow for a comparison between the significances obtained using the BDTs and the baseline.

The relative difference between the counted signal and the 2 methods can be seen in figure 38. Here we can see that the fit in the interval $6 < p_T < 8$ GeV/c is the only accurate fit since the relative differences are below 20%. The relative differences in the other p_T intervals are

between 30% and 60% which are poor. This can be explained by the lack of statistics. When the significances increase it is less challenging to fit them and the relative differences between the fit and the counted signal will decrease. For every p_T interval both fitting methods agree within uncertainty which means the fits are performed correctly given the available statistics.

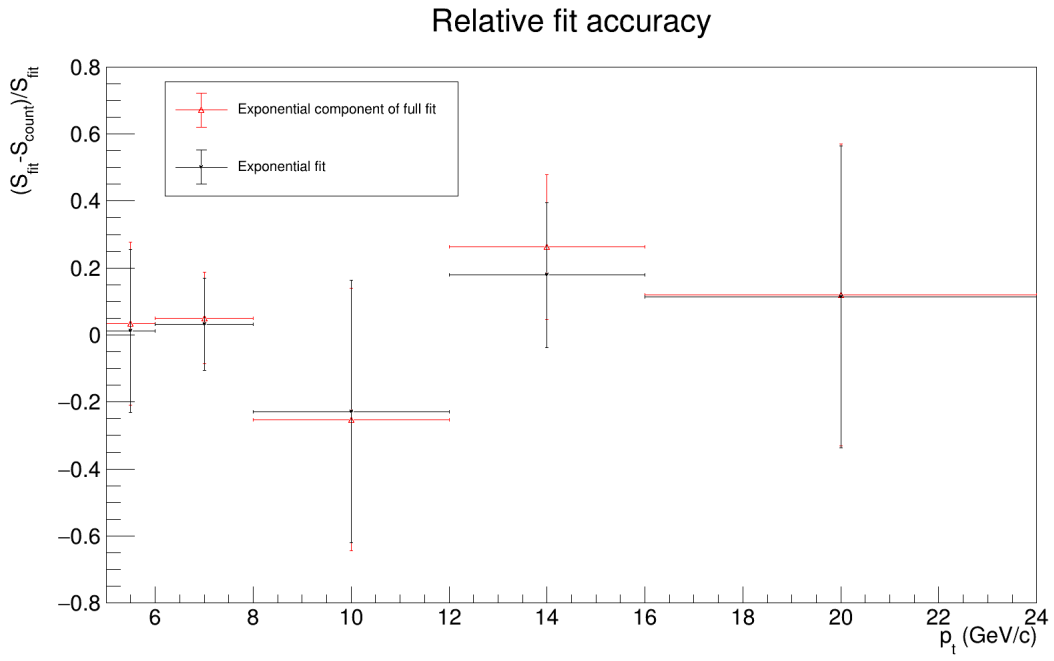


Figure 38: Relative difference between exponential components of the 2 fitting methods and the amount of counted D^0 mesons in the interval $5 < p_T < 24$ GeV/c. The red points represent the exponential component of the full fit while the black points represent the exponential fit performed on points far from the peak.

7 Comparison BDT & Baseline

In this section we compare the effect of replacing the standard cuts from the baseline by a single response cut on the BDT output. Figure 39 shows the accepted prompt and non-prompt D^0 fractions after the track and topological variables are evaluated by either the standard cuts or the BDTs. From this figure it becomes clear that for the BDTs trained on $0 < p_T < 12$ GeV/c the non-prompt fractions are increased compared to baseline and that for the BDTs trained on $p_T > 12$ GeV/c the non-prompt fraction is similar to the baseline. On top of that we can see that the prompt fractions have decreased for $p_T > 2$ GeV/c expect in the interval $3 < p_T < 5$ GeV/c and $6 < p_T < 8$ GeV/c. The prompt fractions fluctuate a lot more than the non-prompt fractions in for the BDTs. This can be explained by the fact that the BDT cut value is chosen to result in a maximum significance and that a separate BDT is trained for each p_T interval. It is possible that the highest significance is sometimes achieved by using a slightly higher prompt efficiency which automatically results in a higher non-prompt efficiency as can be seen in the ROC curves.

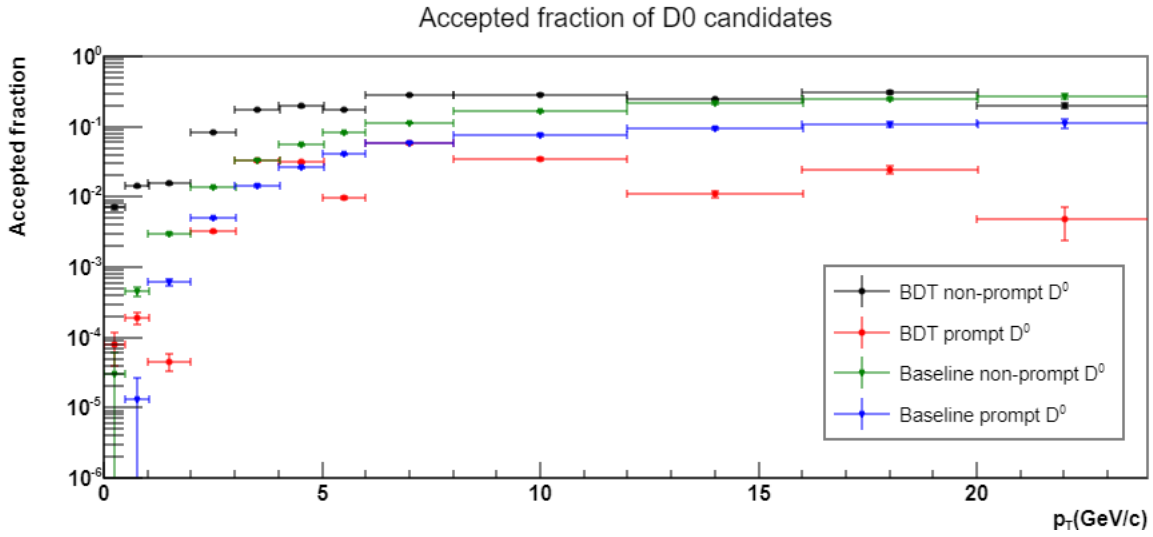
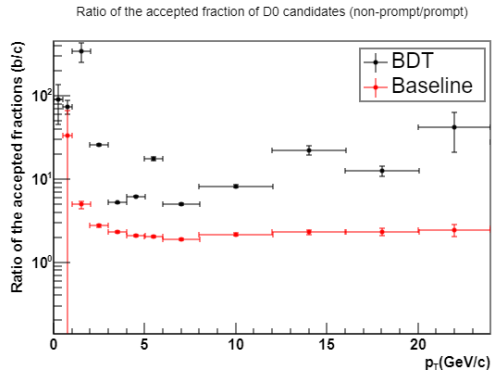


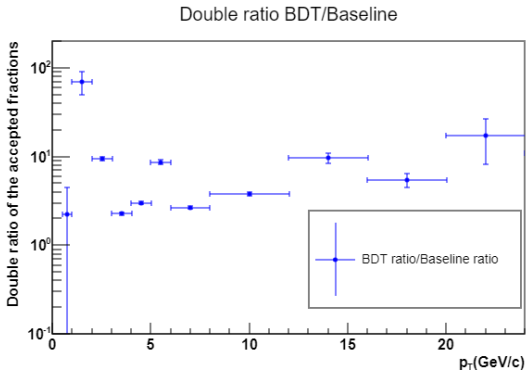
Figure 39: Accepted prompt and non-prompt D^0 fractions after evaluation of track and topological cuts in the interval $0 < p_T < 24$ GeV/c.

A better understanding of the performance of the baseline and BDTs can be achieved by also taking the non-prompt/prompt ratios into account. The ratios can be seen in the left panel of figure 40. Here we can see that the BDT outperforms the baseline for every p_T interval. This statement can only be made when the prompt fraction has a smaller or similar magnitude for the BDTs compared to the baseline since in minimum bias data the prompt fraction is expected to be approximately 40 times higher than the non-prompt fractions. For example an increase of factor 8 in the non-prompt fraction and an increase of factor 2 in the prompt fraction would result in a factor 10 increase of prompt D^0 for the total amount of D^0 that are selected. The factor can be eliminated by looking at the double ratio of the accepted fractions. This can be seen in the right part of figure 40. Here the ratio of the accepted fractions of the BDT is divided by the ratio of the accepted fractions of the baseline. We can

see that for every p_T interval using the BDT instead of the standard cuts results in increase of the non-prompt fraction between 2.268 ± 0.08 and 69.76 ± 20.1 depending on the p_T interval.



(a) Efficiency ratio (non-prompt/prompt)



(b) Double efficiency ratio

Figure 40: Left: The ratios of the accepted fractions (non-prompt/prompt) for the baseline and the BDTs for the interval $0 < p_T < 24$ GeV/c. Right: The double ratio of the efficiencies (BDT/Baseline) for the interval $0 < p_T < 24$ GeV/c.

In figure 41 we compare the results from the invariant mass analysis from the baseline and the BDT on the validation sample. In the left panel we see that the reconstructed invariant masses using baseline cuts are higher than those obtained using the BDTs. The differences are in the order of 10 MeV/c^2 which is very small compared to the invariant mass of 1.864 GeV/c^2. The reconstructed invariant masses in the interval $0 < p_T < 1$ are almost identical such that the blue data point is not visible in the figure. In the right panel we see that the widths of the mass peaks agree within uncertainty.

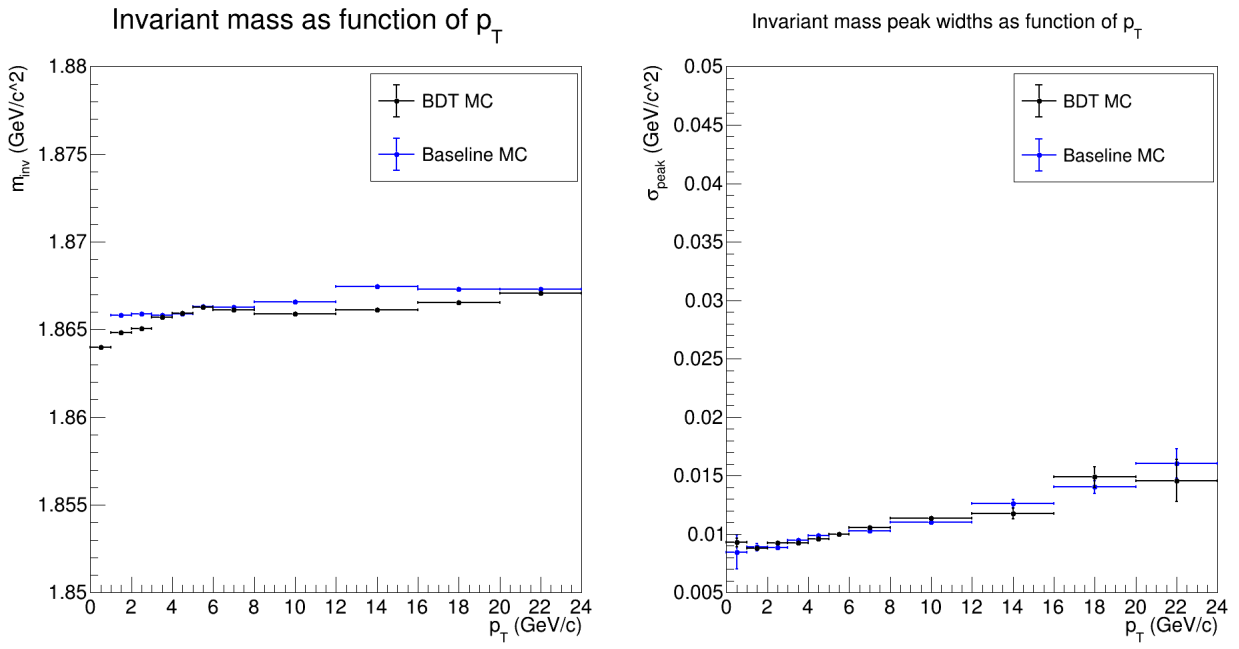


Figure 41: Left: Reconstructed invariant mass in the validation MC sample for baseline and BDT in the interval $0 < p_T < 24$ GeV/c. Right: The widths of the invariant mass peaks in the validation MC sample for baseline and BDT in the interval in the interval $0 < p_T < 24$ GeV/c.

Figure 42 shows a comparison of the reconstructed invariant masses and the peak widths obtained using LHC data between the BDT and the baseline. In the left panel we can see that in the interval $5 < p_T < 16$ GeV/c the values are similar but the reconstructed invariant mass in the interval $16 < p_T < 24$ GeV/c deviates. As we saw in 6.2.5 this was a very poor fit with low significance which could explain why it deviates so much from the other points. In the right panel we see that the widths of the peaks of the BDT agree within uncertainty with the baseline for $5 < p_T < 12$ GeV/c with the exception of $6 < p_T < 8$ GeV/c which almost agrees with the baseline. The widths in the intervals $12 < p_T < 24$ GeV/c certainly do not agree with the baseline. Again this could be explained by the poor fits.

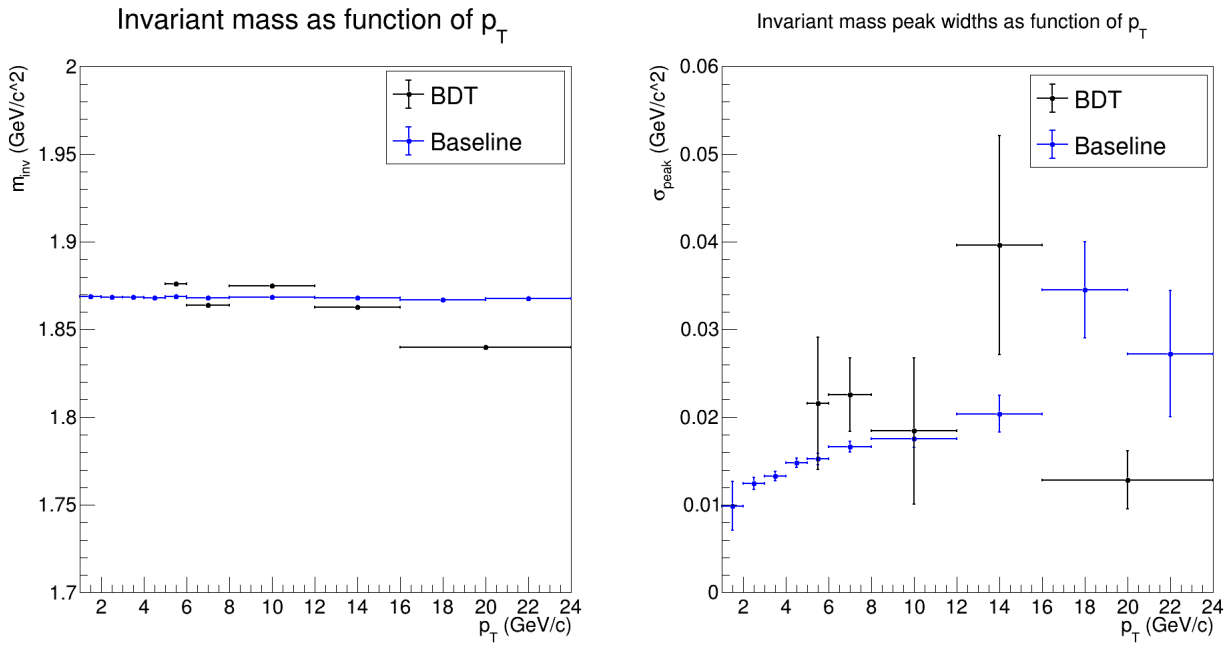


Figure 42: Left: Reconstructed invariant mass obtained using LHC data for baseline and BDT in the interval $0 < p_T < 24$ GeV/c. Right: The widths of the invariant mass peaks obtained using LHC data for baseline and BDT in the interval in the interval $0 < p_T < 24$ GeV/c.

8 Conclusion & Discussion

In this thesis we have shown that machine learning techniques can be used to increase the non-prompt D^0 fraction while decreasing the prompt D^0 fraction compared to the standard cuts made for candidate selection. We successfully developed a framework which allows for the training and implementation of these machine learning algorithms in ROOT. Using a boost decision tree with adaptive boosting we were able to increase the non-prompt fraction between 2.268 ± 0.08 and 69.76 ± 20.1 times depending on the p_T interval of the D^0 when using an algorithm response cut that results in the maximum significance. On top of that we have shown that using our configuration the boost decision tree performs better than the convolutional neural network for $p_T > 3$ GeV/c while the two algorithms are comparable for $p_T < 3$ GeV/c. We saw that both algorithms perform less effectively in the intervals $0 < p_T < 1$ GeV/c and $1 < p_T < 2$ GeV/c, which was expected because the selection variables of prompt and non-prompt D^0 are very similar in these momentum ranges. We were able to reconstruct the invariant mass of the D^0 with significance between 7.8 ± 0.5 and 70.1 ± 0.8 in a forced MC sample using the BDT for candidate selection. This showed that it was possible to reconstruct the D^0 invariant mass using the BDT response as a selection variable. The invariant mass obtained using LHC data has significances between 2.4 ± 0.8 and 6.3 ± 1.1 depending on the selected p_T range. The statistics used in the analysis was 1.84×10^8 which corresponds to 29.7% of the sample used for the baseline. This gives us confidence that significances between 4.4 and 11.6 can be reached once the full sample will be analysed. The position and widths of the invariant mass peaks are reasonably in agreement with the baseline within the large statistical uncertainties. To further improve the reconstruction of the invariant mass of the D^0 particle identification (PID) could be enabled within the analysis task for the TPC and TOF detectors. Using PID more constraints are set on the kaons and pions by evaluating their energy loss. The value should not deviate more than 3σ from the expected signal. Enabling the PID would result in decrease of the background leading to a higher significance. Unfortunately enabling the PID was not possible within the time available for this thesis.

9 Outlook

Further research can be performed on this subject. First of all the convolutional neural networks trained for this thesis can be validated using the same validation sample as was used to evaluate the BDT. The validation sample can also be used to determine the accepted fractions. Afterwards the effect of using the CNNs to separate the non-prompt from prompt D^0 mesons in real LHC data can be studied. The results of the accepted fractions, validation and implementation of the CNNs to analyse real LHC data can be compared to the baseline and to the usage of the BDTs in the analysis. Since we saw in section 6.2.3 that the CNNs performs worse we expect worse results it may be worth to investigate the CNNs further. We saw that for $20 < p_T < 24$ GeV that the CNN performance was worse than for lower transverse momentum while we did not observe the same for the BDT. It would be interesting to study the effect of using larger datasets for both algorithms, but especially for the CNNs. The CNNs can also be improved by varying network layers, layer sizes, optimizer and learning rates. Depending on the computational resources much larger networks can be tested. Studies to improve the performance of the BDTs can be performed. The BDT may improve when the tree depth is changed or when another boost type is used.

For both the BDTs and the CNNs different methods of determining the cut value can be studied. In this thesis we chose to use the cut value of the algorithm which results in an optimal significance. Within the TMVA framework it also possible to determine the cut value which results in a fixed signal efficiency or background rejection. It is even possible that for each p_T interval a different method of cut selection will result in the largest increase of the non-prompt fraction.

The increase of the non-prompt D^0 fraction results in a smaller peak significance if the relative increase is smaller than 40 since in minimum bias data the prompt fraction is 40 times higher. Therefore a possible extension to our studies is to train the machine learning algorithms to separate prompt and non-prompt D^0 mesons from background. The framework to test and evaluate these models is already presented in this thesis. Candidates from the minimum bias MC, which we used in the baseline analysis to test whether the forced MC contained realistic physical features, can be used to train the algorithms. Again the cut on the algorithm can be chosen such that this results in a maximum significance or another method to select the algorithm cut value can be used.

In this thesis we discussed the training of 2 types of machine learning algorithms to increase the non-prompt D^0 fraction in the invariant mass analysis. However there are various classifiers that show promising performance in other classification problems. One type of algorithm that has acquired more popularity as classifier is called a transformer. Transformers can weigh the significance of different parts of the input data. This property is called is called self-attention. Furthermore transformers process their previous output similar to recurrent neural networks (RNNs). In contrast to RNNs transformers have a relative 'long term memory' because all previously generated tokens are saved [26]. Transformers already have been used in particle physics in the field of jet tagging and show promising results [27].

References

[1] R. S. de Rooij. *Prompt D*+ production in proton-proton and lead-lead collisions, measured with the ALICE experiment at the CERN Large Hadron Collider*. PhD thesis, Utrecht U., 2013.

[2] The standard model. `https://en.wikipedia.org/wiki/Standard_Model`.

[3] P. Skands. Introduction to QCD. In *Searching for New Physics at Small and Large Scales*. WORLD SCIENTIFIC, sep 2013.

[4] John C. Collins, Davison E. Soper, and George Sterman. Factorization of hard processes in qcd, 2004.

[5] A large ion collider experiment. `https://home.cern/science/experiments/alice`.

[6] Antonin Maire. Phase diagram of QCD matter : Quark-Gluon Plasma. General Photo, 2015.

[7] Yann Coadou. Boosted decision trees. In *Artificial Intelligence for High Energy Physics*, pages 9–58. WORLD SCIENTIFIC, feb 2022.

[8] Intuition of adam optimizerconvolutional. `https://www.geeksforgeeks.org/intuition-of-adam-optimizer/`.

[9] Convolutional neural networks - basics · machine learning notebook. `https://mlnotebook.github.io/post/CNN1/`.

[10] Muhamad Yani, S Irawan, and Casi Setianingsih. Application of transfer learning using convolutional neural network method for early detection of terry's nail. *Journal of Physics: Conference Series*, 1201:012052, 05 2019.

[11] C. Fabjan and J. Schukraft. The Story of ALICE: Building the dedicated heavy ion detector at LHC. 1 2011.

[12] Arturo Tauro. ALICE Schematics. General Photo, 2017.

[13] M. Krivda et al. The alice silicon pixel detector readout electronics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 617(1):549–551, 2010. 11th Pisa Meeting on Advanced Detectors.

[14] Shreyasi Acharya et al. Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons. *JHEP*, 05:220, 2021.

[15] M. Sitta. The silicon drift detector of the alice experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 617(1):591–592, 2010. 11th Pisa Meeting on Advanced Detectors.

[16] Weilin Yu. Particle identification of the alice tpc via de/dx. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 706:55–58, 2013. TRDs for the Third Millenium.

[17] Walter Blum, Werner Riegler, and Luigi Rolandi. *Particle detection with drift chambers; 2nd ed.* Springer, Berlin, 2008.

[18] J. Adam, Dagmar Adamova, Muskaan Aggarwal, Gianluca Aglieri Rinella, Michelangelo Agnello, Nishika Agrawal, Zubayer Ahammed, Shakeel Ahmad, S. Ahn, Salvatore Aiola, A. Akindinov, Sartaj Alam, D. Albuquerque, D. Aleksandrov, Borri Alessandro, D. Alexandre, R. Molina, A. Alici, Anton Alkin, and Johann Zmeskal. Determination of the event collision time with the alice detector at the lhc. *The European Physical Journal Plus*, 132, 02 2017.

[19] Michael Karim Habib. Light (anti)nuclei production at the LHC measured in pp collisions at 13 TeV. Produktion leichter (Anti)Kerne am LHC gemessen in pp Kollisionen bei 13 TeV, 2022. Presented 05 Dec 2021.

[20] The ALICE collaboration. Performance of the alice vzero system. *Journal of Instrumentation*, 8(10):P10016, oct 2013.

[21] S. Acharya et al. Long- and short-range correlations and their event-scale dependence in high-multiplicity pp collisions at $$ \sqrt{s} $$ = 13 TeV. *Journal of High Energy Physics*, 2021(5), may 2021.

[22] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[23] Jaroslav Adam et al. Centrality and transverse momentum dependence of D^0-meson production at mid-rapidity in Au+Au collisions at $\sqrt{s_{\rm NN}} = 200\,\rm GeV$. *Phys. Rev. C*, 99(3):034908, 2019.

[24] About root. `https://root.cern/about/`.

[25] Machine learning with root. `https://root.cern/manual/tmva/`.

[26] Michael Phi. Intuition of adam optimizerconvolutional. `https://towardsdatascience.com/illustrated-guide-to-transformers-step-by-step-explanation-f74876522bc0`.

[27] Huilin Qu, Congqiao Li, and Sitian Qian. Particle transformer for jet tagging, 2022.