Master Thesis

Application of Machine Learning to Particle Identification for Dielectron Analysis in CBM

Henrik Schiller

Münster, August 2022

Contents

1 Introduction

The Compressed Baryonic Matter (CBM) experiment is a fixed-target experiment for heavy-ion collisions which is expected to provide new insights into the QCD phase diagram at high net baryon densities using high-energy nucleus-nucleus collisions. The experiment is located at the accelerator complex Facility for Antiproton and Ion Research (FAIR), which is currently under construction at the Helmholtz Center of the Gesellschaft für Schwerionenforschung (GSI) in Darmstadt. Nucleus-nucleus collisions can produce high baryonic densities that naturally exist only in very dense objects, such as neutron stars or in the early universe. New states of matter produced in the laboratory can probe the transition from baryonic matter to the quark-gluon plasma (QGP). The QGP is a state of matter in which quarks and gluons are not bound in nucleons and is reached only at very high temperatures and/or baryon densities. In addition to exploring the equation of state for nuclear matter at neutron star densities, as well as searching for phase transitions, the experiment is investigating the recovery of chiral symmetry and exotic forms of (strange) QCD matter [1]. Since the QGP is a very short-lived state, it can almost exclusively be probed by particles resulting from the freeze out after the expansion of the fireball created by the collision. The CBM detector is designed to measure resulting tracks with high precision and unprecedented statistics. It targets interaction rates of up to 10 MHz, which is two orders of magnitude higher than comparable existing heavy ion experiments such as BM@N or STAR [2].

"The unique combination of an accelerator delivering a high-intensity heavy ion beam with a modern high-rate experiment based on innovative detector and computer technology provides optimal conditions for a research program with significant discovery potential for fundamental properties of QCD matter." [1]

Thermal photons materializing from the collision fireball as e^+e^- or $\mu^+\mu^-$ pairs will be used to measure the time-averaged temperature of the QGP. For electron pair analysis, electrons must be distinguished from other particles using different variables measured by the detectors. The dielectron spectrum is obtained by calculating the invariant mass of all combinations of all electrons and positrons. With this spectrum the time-integrated temperature of the fireball can be determined. The challenge is to distinguish electrons from other particles such as pions, which is on one hand the task of the CBM detector setup and on the other hand the task of a suitable classification algorithm. In this thesis, different machine learning methods are utilized for particle identification (PID), improving the efficiency of the process. Unlike other experiments, CBM does not have an external trigger to record an event, but utilizes self-triggered readout system. Incoming data is, to some extent, directly reconstructed and analyzed online to extract relevant information from the raw data stream, which

exceeds 1 TByte/s [3]. Due to the high interaction rates, the particle classification method in CBM has to be fast, as it happens in real time.

The aim of this thesis is to use a modern machine learning method like XGBoost or Random Forest for the electron PID at the CBM experiment to increase the significance of dielectron spectra. For this purpose, a focus is set that the algorithm is sufficiently fast. Furthermore transparency of the algorithm is required, to ensure that the algorithm does not only produce good results in the simulation but also in the real experiment.

2 Measurement Goal and CBM Detector System

The strong nuclear force binds the majority of matter in the observable universe. It ensures that the components of the nuclei of our atoms, the quarks and gluons, are in stable bound states. The dynamics of these subparticles of a nucleon are described by the quark-gluon interaction. This is especially complex because the exchange particles, the gluons, themselves carry a color charge, in contrast to e.g photons in the electromagnetic interaction . At very high temperatures and/or densities, the bound objects "melt", and the constituents form a new state of matter called the quark-gluon plasma (QGP). The main feature of this state of matter is the fact that now the interaction of quasi free quarks is observable since here the confinement of quarks and gluons is removed. The main goal of the CBM experiment is to better understand the different states of hadronic matter. For this purpose, the CBM detector has been developed to measure the collective behavior of hadrons together with rare diagnostic probes such as multi-strange hyperons, charmed particles and vector mesons decaying into lepton pairs with unprecedented statistics.

2.1 Heavy Ion Collisions and the Quark-Gluon Plasma

The very short time scale of a heavy ion collision (5-10 fm/c) makes it impossible to measure effects in real time in experiments. However, the time evolution of the produced fireball has different stages with different characteristics of particle production, which can then be detected by the CBM detector. Figure 2.1 shows the different stages of the fireball evolution as simulated with Ultrarelativistic Quantum Molecular Dynamics(UrQMD).

In the following, the time evolution of a heavy ion collision from Figure 2.1 is described in 3 stages:

1. In the initial state, two Lorentz-contracted heavy ions are about to collide. In the process, they pass through a certain overlap region, creating a high energy density and (possibly) forming a new state of matter. This matter only survives for a very short time. It is expected that the charm production already happens in this very early state of the fireball. Likewise, in this state the production of prompt photons occurs, which later decay into dilepton pairs.

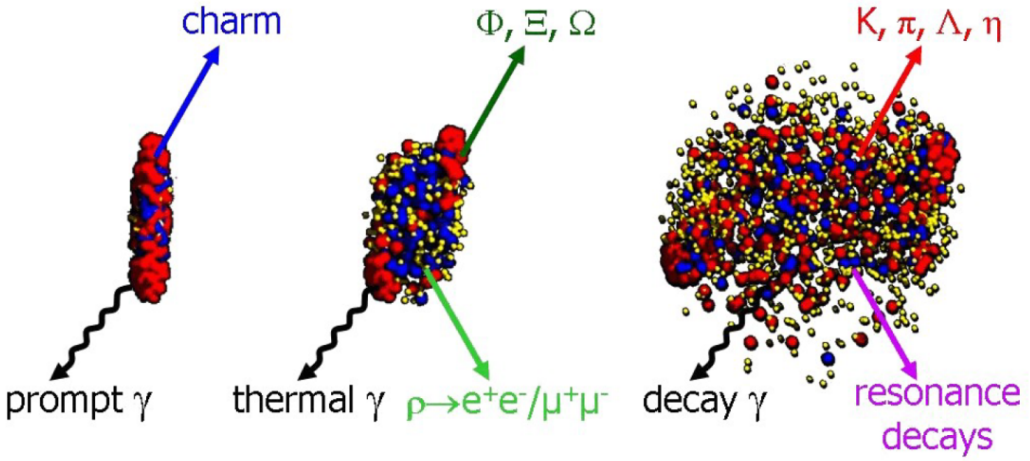


Figure 2.1: The time evolution of a heavy ion collision. The Lorentz-contracted ions (left) collide. In the overlap region between the two ions, the first hard collisions take place with large stopping and high energy density. This phase transitions into a hydrodynamic evolution of quarks and gluons (center). In the final state, the particles freeze out to hadrons (right). [4]

2. The second state is a very hot and dense phase of matter, in which J/Ψ production starts. The thermal radiation emitted in this phase yields information about the temperature of the fireball. The thermal photons are produced by blackbody radiation, and decay into dileptons via electromagnetic interaction. Since thermal radiation is one of the main research areas of this thesis, it will be discussed in detail in section 2.3. Beside the thermal radiation also the low mass vector mesons like ρ, ω, and ϕ can decay into lepton pairs. These are also produced in this hot and dense phase of the fireball.

3. The newly formed matter expands and freezes in the final stage to hadronic matter, which is also called chemical freezeout. The particles in this final state are called the cocktail after the chemical freezout and form the bulk of the particles after the collision. Multi-strange baryons, charmed particles, hypernuclei, and dilepton pairs are measured at CBM for the first time in the FAIR energy-range, which is why CBM is designed for a high significance measurement of these observables.

An important parameter for heavy ion collisions is the centrality, which indicates whether the ions collide head-on or rather just brush against each other. The centrality directly affects the momentum distributions and pressure gradients within the collision zone. Furthermore, heavy ion collisions and their final state are strongly dependent on the particle types of the collision system and the collision energy. Depending on these parameters, different types of matter can form during the collision.

2.2 The QCD Phase Diagram

The QCD phase diagram assigns a phase to the variables temperature and baryochemical potential. One of these phases is cold nuclear matter, as observed in the core of atoms. This cold nuclear matter consists of protons and neutrons bound to nucleons by the strong interaction. In Figure 2.2 a scheme of the QCD-phase-diagram is shown, where the hadronic phase, i.e. that of ordinary nuclei, can be found at small temperatures and small baryon chemical potentials. More precisely, normal nuclei are located at $\mu_B = 924\,\mathrm{MeV}$ at a baryon density $n_0 = 0.17\,\mathrm{fm}^{-3}$ and a temperature slightly above 0 MeV. Since many parts of this phase diagram are still widely unexplored, there is much effort to answer open questions about the phases and their transitions. One of these open questions is the search for a potential critical point which is shown as a red dot in Figure 2.2. For high temperatures the hadronic phase transitions to the quark gluon plasma which has already been predicted theoretically and shown experimentally. According to lattice QCD calculation this transition is a smooth crossover at $\mu_b \approx 0$ [5]. It is expected that the quarks are unbound after the transition and there is a restoration of the chiral symmetry such that $\langle q\bar{q}\rangle = 0$. At higher baryon densities a phase transition of the 1st order is suspected, which leads to the existence of a critical point in the prediction [6].

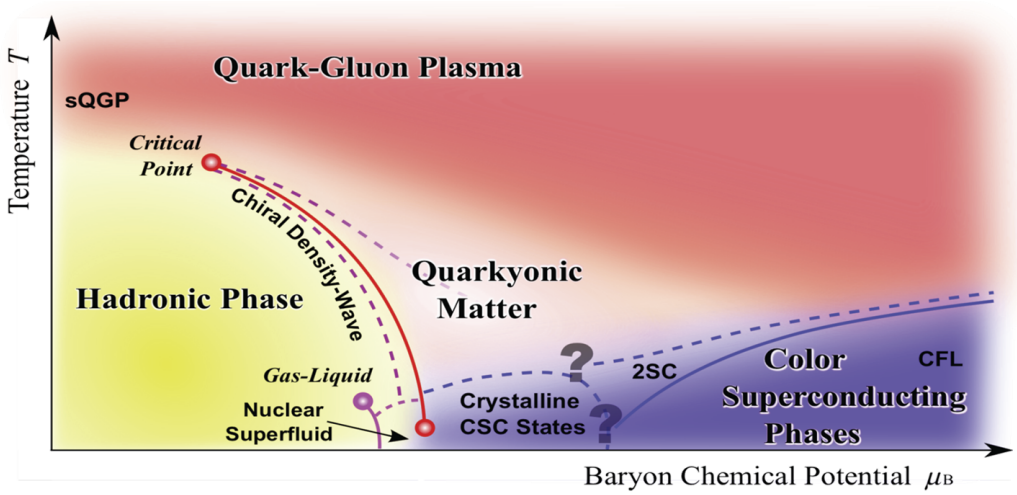


Figure 2.2: Schematic drawing of the phase diagram of hadronic matter [2]

2.3 Lepton Pairs

Lepton pairs are a key observable to measure the temperature of the fireball. Since the thermal photons do not interact via the strong interaction, they can leave the fireball undisturbed. Real and virtual photons are emitted throughout the evolution of the fireball and can decay into lepton pairs l^+l^-. Since these decays occur with a probability of 10^{-4} per collision, it will be a challenging measurement for the

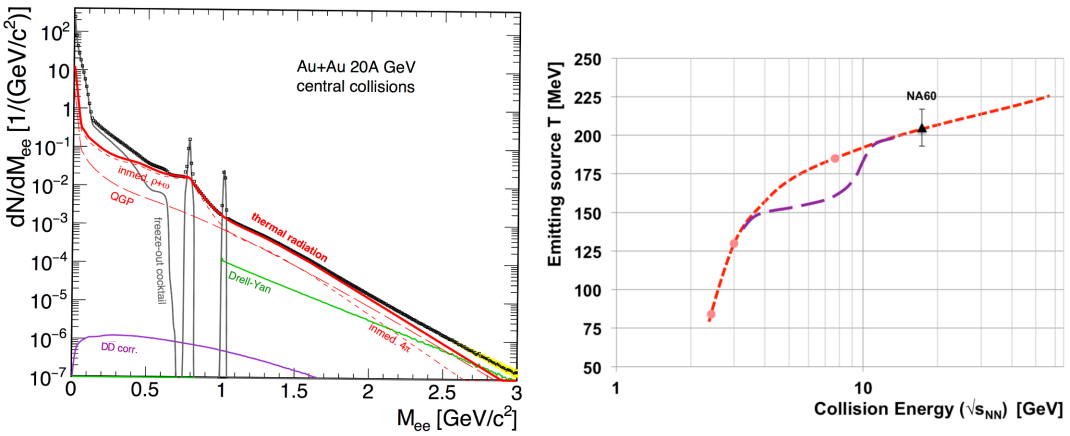


Figure 2.3: Left: Invariant mass spectrum of e^+e^- pairs radiated from central Au+Au collisions at 20 AGeV . The diagram includes different sources of dielectrons, which were simulated by [7] and [8]. The solid red curve shows the contribution of the thermal radiation which includes in-medium ρ, ω, $4-\pi$ spectral functions and QGP spectrum calculated using the many-body approach of [9] . Right: Sketch of the emitting source temperature as function of collision energy for two scenarios. The red dashed curve is the result of a calculation based on a thermal model assuming no phase transition or a crossover (see text). The purple line illustrates a hypothetical caloric curve reflecting a first order phase transition. In addition, the experiments result from NA60 [10] are shown [2].

experiment. Figure 2.3 shows the invariant mass spectrum for dielctrons in an Au+Au collision at 20 AGeV. The invariant mass of a lepton pair can be calculated using the reconstructed four-momentum:

$$M_{l+l-} = \sqrt{(E_{l+} + E_{l-})^2 - (\vec{p}_{l+} \cdot c + \vec{p}_{l-} \cdot c)^2} \tag{2.1}$$

Here E_{l+} and E_{l-} are the energies of the individual leptons which are determined using the momentum and the mass connected by the relativistic energy momentum relation.

The spectrum in Figure 2.3 (left) is normally divided into three mass regions:

1. The **Low Mass Region** (LMR) is below 1 GeV/c. The main source of dielectrons in this region are light vector-mesons like ρ, ω, ϕ and Dalitz decays like the decay of π^0 into a dielectron pair and a photon.

2. The **Intermediate Mass Region** (IMR) is the mass region between 1 GeV/c and 3 GeV/c , i.e. the region between the Φ and the J/Ψ peak. Since in the energy region of CBM the hidden charm production, which dominates this region for larger energies, is suppressed [2] it is possible to find out the temperature of the fireball by the slope of the curve. Since thermal photons arise from blackbody radiation of the fireball, the time-integrated temperature of the fireball can be determined using a Boltzmann fit (equation 2.2) for the

invariant mass M_{ee} [11].

$$f(M_{ee}) \propto M_{ee}^{3/2} \cdot e^{-M_{ee}/T} \tag{2.2}$$

If the temperature is measured for different collision energies, conclusions can be drawn about a possible phase transition and the restoration of chiral symmetry. [2]

3. The **high invariant mass range** concerns masses above about 3 GeV/c. At energies higher than those of CBM, this region is dominated by quark-antiquark annihilation processes. In the CBM energy regime, this is not the case, and the other signal contributions are mainly due to the Drell-Yan process (see Figure 2.3 on the left).

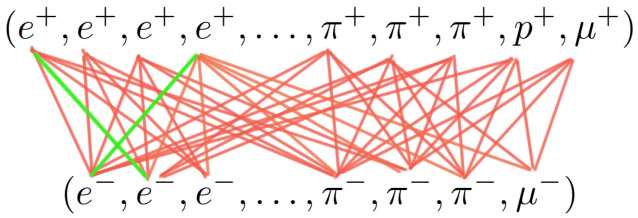


Figure 2.4: Schematic representation of the process of pair analysis, all positive tracks are paired with all negative tracks. Only a fraction of electron pairs contribute to the signal (in green). Background pairs are shown in red.

To get the invariant mass spectra for the electrons from the nucleus collisions, the electrons have to be filtered from the rest of the detected particles first. Any combinations of an electron and a particle falsely identified as an electron contributes to the background (see Figure 2.4). Errors are made by the not perfect detectors and imperfect classification methods and therefore a lot of e.g. pions are classified as electrons. Considering an event with one correlated e^+ e^- pair and n as the number of positive tracks of the event and k as the number of negative tracks, the result is a number of $(n-1) \cdot (k-1)$ uncorrelated combinations. The background contribution to the spectrum should therefore be kept as small as possible using reliable electron identification methods, which are presented in this thesis. One of the main challenges for the dielectron analysis is, due to the large hadronic contribution, that the CBM detector has a sufficient electron identification in the whole momentum region.

2.4 Facility for Anti Proton and Ion Research (FAIR)

Well-known accelerator facilities such as RHIC and LHC explore the phase diagram towards higher temperatures and near-zero net baryon densities due to their high energy but comparably low interaction rates [12]. The CBM experiment, in contrast, explores the regions of higher baryon chemical potential at moderate temperatures [13]. For these reason, construction of a new accelerator facility, FAIR started in 2017 next

to the GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt/Germany [14]. FAIR will expand the existing GSI complex by building a new ring accelerator, the SIS100, and new experiment facilities. This accelerator will have a magnetic rigidity of 100 T·m and a circumference of 1100 m. In Figure 2.5 a sketch of the planned accelerator complex can be seen. To the left of SIS 100 is the existing SIS18, which will serve as a pre-accelerator for SIS100. SIS100 will be able to accelerate protons, antiprotons and nuclei of all kinds up to uranium [13]. Energies of 29 GeV for protons and, for example, 11 AGeV for gold will be achieved. [15].

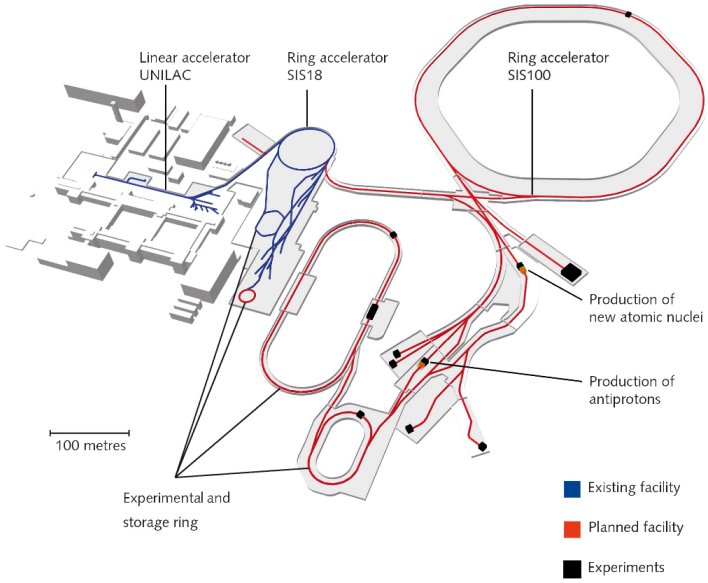


Figure 2.5: Illustration of the FAIR facility with the accelerators under construction such as the SIS 100 in red. The CBM experiment is represented by the larger black rectangle on the right. Other planned experiments and storage rings such as the high-energy antiproton ring are shown in red. The SIS 100 uses the already built SIS18 (in blue) as a pre-accelerator. [16]

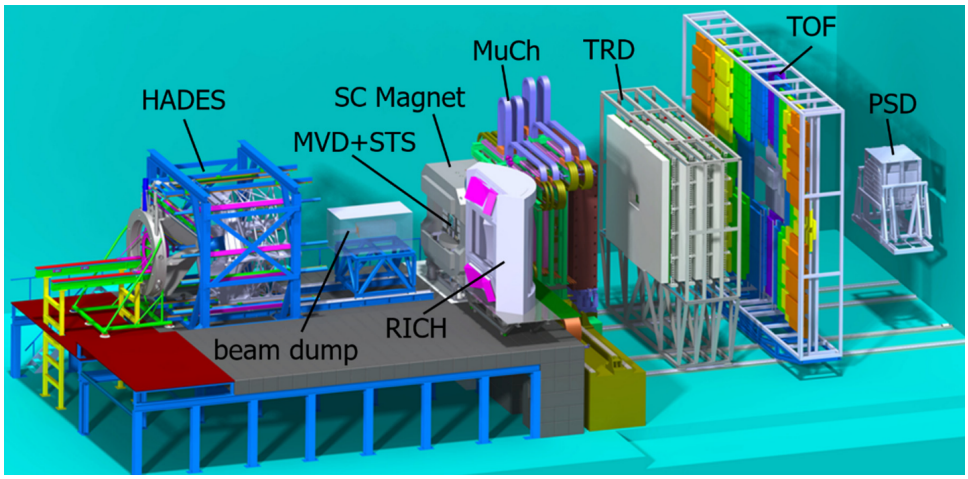


Figure 2.6: The HADES detector (left) with its beam dump, which will be removed during CBM operation. The CBM experimental setup comprises the following components (from left to right $\hat{=}$ beam direction): the superconducting dipole magnet, the Micro-Vertex Detector (MVD) and Silicon Tracking System (STS) located in the magnet gap, the Muon Chamber system (MuCh) in measuring position, the Ring Imaging Cherenkov (RICH) detector in parking position, the Transition Radiation Detector (TRD), the Time-Of-Flight (TOF) detector and the Projectile Spectator (PSD) detector. The CBM subdetectors are described in the text. For muon measurements, the RICH will be exchanged with the MuCh which is shown in a parking position to the right of the beam direction. [17]

2.5 The CBM Subdetectors

At FAIR, heavy ion collisions are studied not only with the CBM experiment but also with the HADES experiment. The two experiments are operated alternately and are shown in Figure 2.6. The HADES detector system has a large polar angular acceptance between 18 and 85 degrees and is therefore well suited for reference measurements with proton and intermediate mass ion beams such as Ag at low SIS100 energies. The HADES setup will measure hadrons and electron-positron pairs. For the measurements with the CBM setup, the beam dump shown in Figure 2.6 will be removed. With an interaction rate of up to 10 MHz and beam energies of e.g. 11 AGeV for Au+Au collisions, 10^9 charged particles (hadrons, electrons and muons) per second have to be measured. This requires fast and radiation hard detectors and a novel data readout and acquisition system. The CBM detector system is suitable for polar angles of 2.5 to 25 degrees, which is sufficient to cover the forward velocity range including the mean velocity.

2.5.1 Silicon Tracking System (STS)

The 8 layers of the Silicon Tracking System (STS) are used to determine the momentum of charged particles. These are deflected by the magnetic field in a momentum-dependent manner as shown in Figure 2.7 on the left. Since the particles leave hits in

the layers of the STS with a high spatial resolution, the track of the particle can be reconstructed. A technical drawing of the STS is shown in Figure 2.7 on the left. The STS is located 30 cm to 100 cm downstream of the target directly behind the micro vertex detector [4].

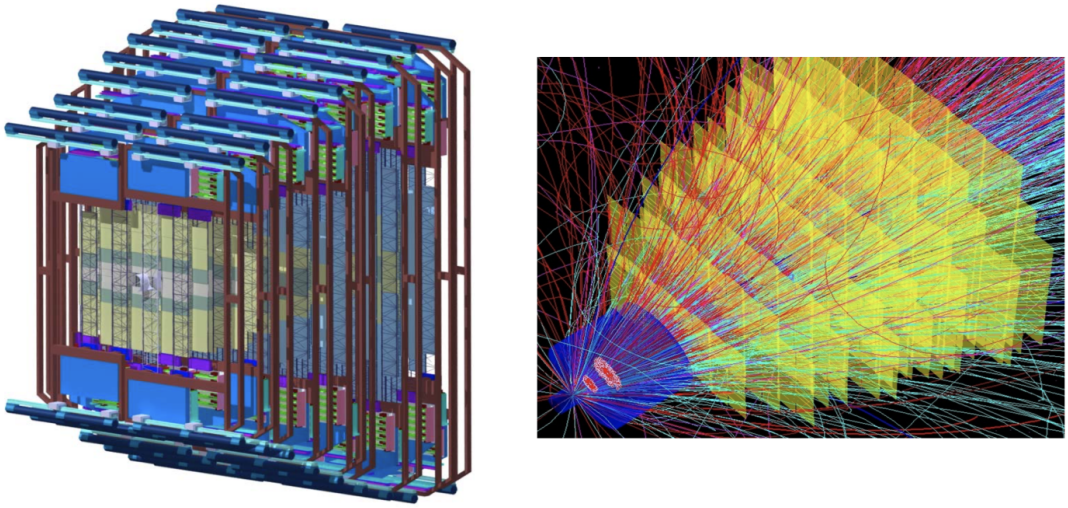


Figure 2.7: Left: Geometry of the CBM-STS with 8 tracking stations [18]. Right: GEANT simulation of a 10% most central Au+Au collision at 25 A GeV with the tracks in the STS [18]. [4]

2.5.2 Micro Vertex Detector (MVD)

The MVD will consist of four layers of Monolithic Active Pixel Sensors (MAPS) with a pixel size of 27 - 30 µm^2 [19]. Like the STS, its main task is tracking, which is why it is located inside the magnetic field. The study of open charm requires a very high resolution of the decay vertex of the particle. For example, the D^0 meson decays after 123 µm, so the micro vertex detector is positioned very close (5 cm to 20 cm) to the target. Since the position resolution can be drastically disturbed by multiple scattering, it is important that the detector has a small material budget and is in vacuum. Figure 2.8 shows a technical drawing of the MVD.

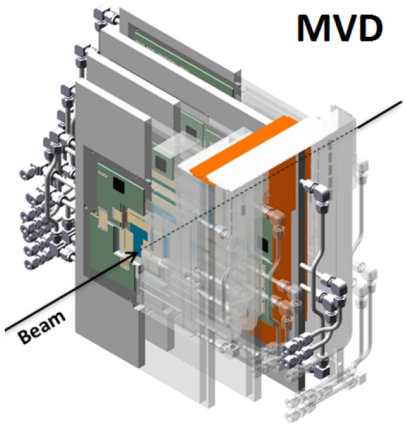


Figure 2.8: Design of the CBM-MVD [18]

2.5.3 Ring Imaging Cherenkov Detector (RICH)

After the beam has left the MVD and STS tracking detectors, it reaches the Ring Imaging Cherenkov (RICH) detector about 1.6 m to 3.6 m downstream of the target outside the magnet. Its main task is to identify electrons and positrons in the low momentum range up to momenta of about 8 GeV/c using Cherenkov light. This is especially important for electron physics, since large contributions from pions can thus be extracted from the observed tracks. A technical sketch of the detector is shown in Figure 2.9 on the right.

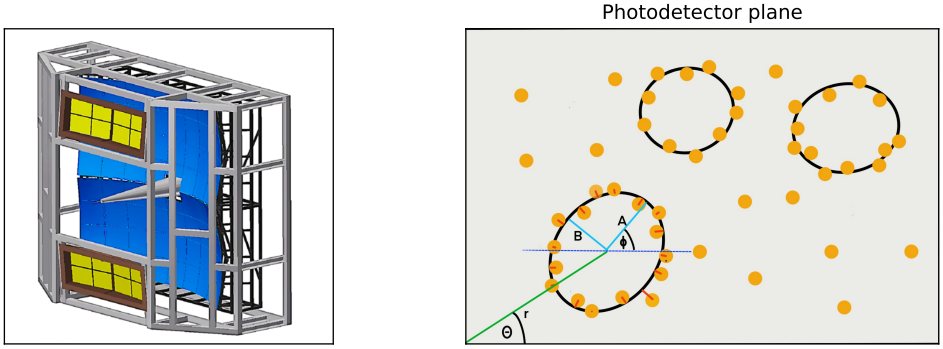


Figure 2.9: Left: Technical drawing of the CBM-RICH detector. The two mirror arrays are shown in blue. The photomultipliers are shown in yellow. In grey is the aluminum support structure [4]. Right: Illustration of the photon hits (orange) on the photodetector plane. Some of the hits form an ellipse which can be fitted with five parameters (Θ, r, A, B, ϕ).

When a particle moves faster than the speed of light inside a medium, Cherenkov photons are emitted. Thus, this process depends on the refractive index of the material. Suitable media for the RICH detector are gases like CO_2, because on the one hand a high probability for Cherenkov radiation production must be ensured and on the other hand a low absorption of this radiation by the medium itself is needed. The convex mirrors, shown in blue in the technical sketch, focus the photons on photon detection planes, where they can be read out by photomultiplier tubes. Since the Cherenkov light propagates like a cone around the track, photon ellipses result when projecting them onto the photodetector plane (see Figure 2.9 right). After the detection of the photons, a ring finding algorithm is used to fit ellipses around the data. Each fit of an ellipse has 5 degrees of freedom: The center of the ellipse (r, Θ), the major and minor semi-axis (A,B) and the tilt angle ϕ (see Figure 2.9 right). In chapter 6 the identification of electrons is investigated in more detail using the RICH data.

2.5.4 Transition Radiation Detector (TRD)

Behind the RICH detector, the particles reaches the four layers of the TRD detector after about 4 m to 6 m. The TRD enables distinction of electrons from pions with higher momenta ($p > 1 - 2\,\mathrm{GeV/c}$) by utilization of transition radiation. Thus, RICH and TRD converge to an electron identification over the whole momentum range. Electrons have a larger Lorentz factor (γ-value) than pions for the same momentum, since pions have a larger mass. Since transition radiation (TR) depends on the γ-value, electrons produce more TR for the same momenta than pions, which is measured by the detector. The TRD is a multi-chamber gas detector which uses mirror charges to reconstruct the location of a hit. Therefore the hits in the TRD layers contribute to the tracking. Figure 2.10 shows column-wise normalized two-dimensional histograms for the energy deposition in one of the TRD layers for electrons and pions. Figure 2.10 left and middle show the energy deposition for different momenta. It can be seen that the deposition does not change significantly but the peaks smear out a bit more. Figure 2.10 on the right shows the histograms momentum integrated where the momenta are weighted with the respective frequency. From Figure 2.10 the probability that a particle with a given momentum and a given energy deposition in the detector is an electron can be calculated (see Figure 2.11). For the calculation the counts from the electron histogram for each pixel are divided by the sum of the electron and the pion histogram. The momentum integrated probability is shown in Figure 2.11 b). Furthermore, it can be seen that larger energy depositions in the detector layers lead to a 100% probability of electron hits. This result is expected due to the transition radiation emitted by electrons.

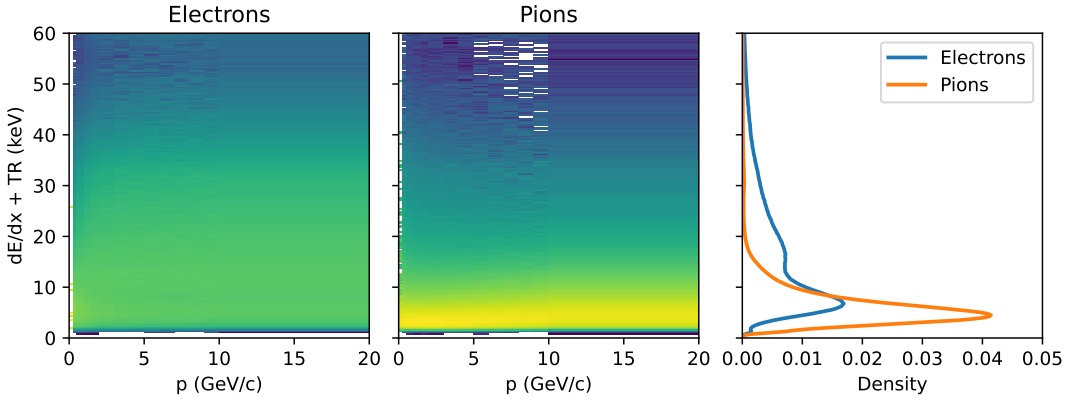


Figure 2.10: Left: Column-wise normalized histogram of the simulated energy output to the TRD for different momenta of the electrons and pions (middle), respectively. Right: Momentum integrated histograms. Electrons have more energy output to the TRD due to the generated transition radiation.

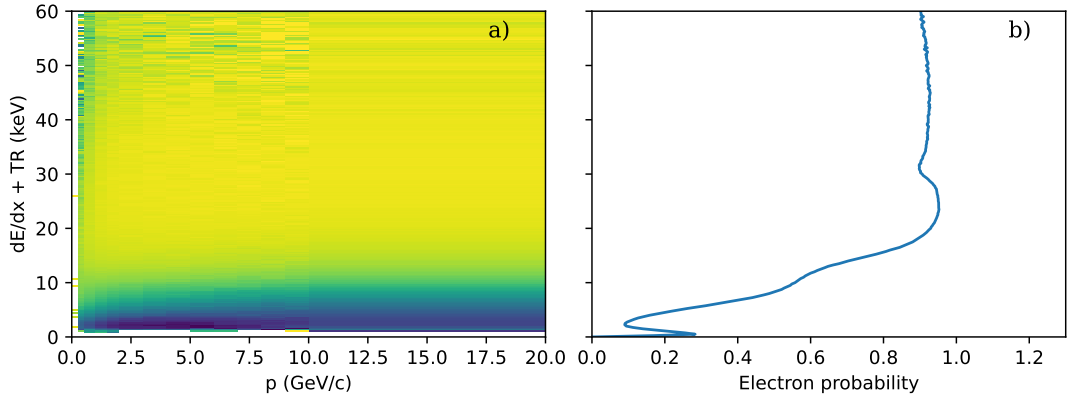


Figure 2.11: Column-wise normalized two dimensional probability histogram for the correct classification of a particle as electron with the respective momentum and energy deposition. b) shows the probability that a particle with a certain energy deposition is an electron. The curve is integrated over all momenta.

2.5.5 Time of Flight Detector (TOF)

To determine the mass of tracked particles, a Time-Of-Flight (TOF) wall is installed approximately 7 m downstream of the target, with an active area of approximately 120 m^2 [19]. The TOF wall consists of a matrix of multi-gap resistive plate chambers. The total temporal resolution of the TOF detector is 80 ps. The TOF is mainly responsible for the identification of hadrons but can also contribute to electron identification.

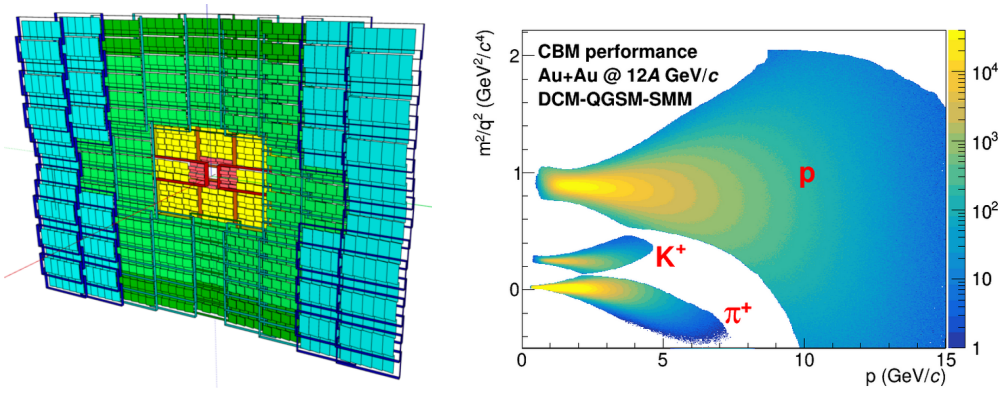


Figure 2.12: Left: Technical drawing of the TOF wall [15]. Right: Simulation of the determined masses of protons ,kaons$^+$ and positive pions selection with 90% purity requirement determined by TOF. [20]

Particle identification is done by calculating the mass from the reconstructed momentum and velocity (see equation 2.3). The velocity is calculated from the track-length and the time of flight. The momentum p is determined by the curvature of the track in the magnetic field determined by the STS.

$$m = \frac{p}{\gamma\beta c} = \frac{p\sqrt{1-\beta^2}}{\beta c} \qquad \text{with} \quad \beta = \frac{v}{c} \tag{2.3}$$

A simulation of the masses determined by the TOF using equation 2.3 is plotted in Figure 2.12 against the corresponding momentum for pions, protons and kaons.

3 Theory of Classification Methods

A number of machine learning methods are suitable for the classification of particles based on detector data. Among these are for example “Nearest Neighbors”, “RBF SVM”, “Gaussian Process”, “Decision Tree”, “Random Forest”, “Neural Net”, “AdaBoost”, “Naive Bayes” and “QDA”. The classifiers can assign a type probability to each point in the input space. Figure 3.1 is intended to illustrate the nature of the decision boundaries of different classifiers. The figure shows the decision boundaries of selected classifiers which are arranged between two classes, blue and red, in two dimensional space. For this purpose, simulated data, such as two crescents interlocking (Figure 3.1 row 1), points lying inside each other (row 2) or points distributed next to each other (row 3) are compared. The input data sets are examples and are not based on real data, which can have significantly more complicated arrangements. The examples in Figure 3.1 are two-dimensional. In higher dimensions, data can be more easily linearly separated, so classifiers such as Naive Bayes or linear SVM can produce better results than those presented in Figure 3.1.

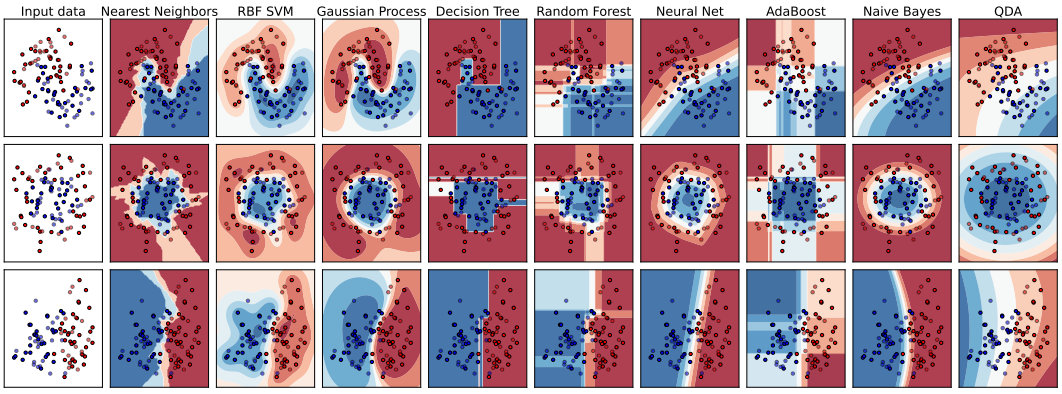


Figure 3.1: An illustrative comparison of several classifiers on synthetic datasets.

Since not all of these classification methods for the distinction of electrons and hadrons in the CBM experiment yield good results, only the most important 4 Methods are presented in detail: “Decision Tree”, “Random Forest”, “Neural Net” and “XGBoost”. After the introduction of these machine learning methods, methods for the investigation of the classification performance are presented.

3.1 Decision Tree

A Decision Tree is a tree-like directed diagram for decision making. It consists of root, nodes, branches and leaves. The nodes form the decision-dependent branch points. Typical application of decision trees are classification tasks. In order to introduce the theory of decision trees, a thought experiment is considered, which is suitable on the one hand to briefly explain the functioning of a decision tree, but also to introduce the concept of information gain by a decision.

> In a box are the letters A, B, C and D each once. Now the game leader chooses one letter at random. What is the best strategy to find out which letter it is with as few yes/no questions as possible?

Two frequently mentioned decision trees are now examined:

1. It is asked: "is it A or B?", if yes it is asked "is it B?", if no "is it C?". Here 2 questions are used to find out which letter the game master has chosen.
2. It is asked "Is it A?", if no, it is asked "Is it B?", if no, it is further asked "Is it C?". In 1/4 of the cases one question is needed, in 1/4 of the cases two questions and in 1/2 of the cases 3 questions. That makes an average of 2.25 questions.

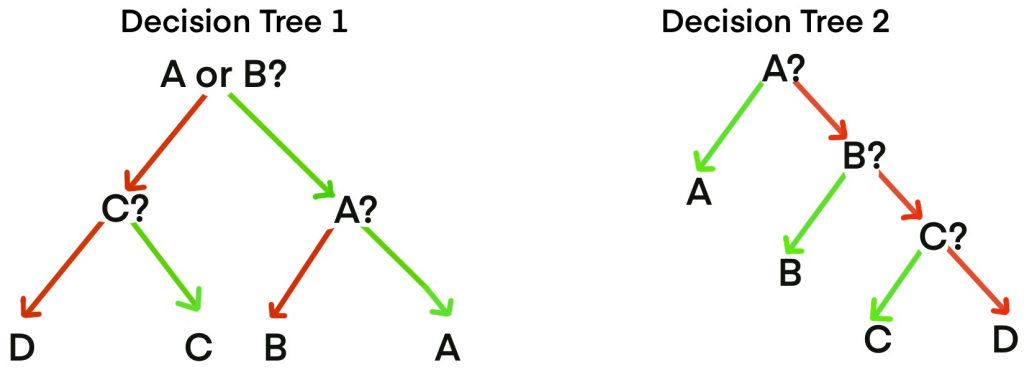


Figure 3.2: Different decision trees for asking a randomly selected letter from the set A,B,C,D. Green arrows indicate the answer "yes", red arrows indicate the answer "no".

Decision tree 1 in Figure 3.2 needs fewer questions for a classification on average. An average of two questions is also called 2 Bit, i.e. two yes/no questions (1 or 0), which are needed to find out which letter it is. By asking the question "is it A or B?" an information gain of 1 Bit is obtained, because after the answer another question must be asked to identify the letter and 2 Bit-1 Bit = 1 Bit is valid. If the first question is: "is it A?" and the answer is no, there are on average another (1+2+2)/3 = 1.66 questions that need to be asked and thus there is a lower information gain. In this thesis, the goal is to find a decision tree that distinguishes between electrons and pions with as few questions as possible. An example of such a decision tree for

the data of the RICH detector in the CBM experiment presented later in this thesis is shown in Figure 3.3. A difference to the decision trees in Figure 3.2 is that the decision boundaries are real numbers like for example if the number of hits is smaller than 26.5.

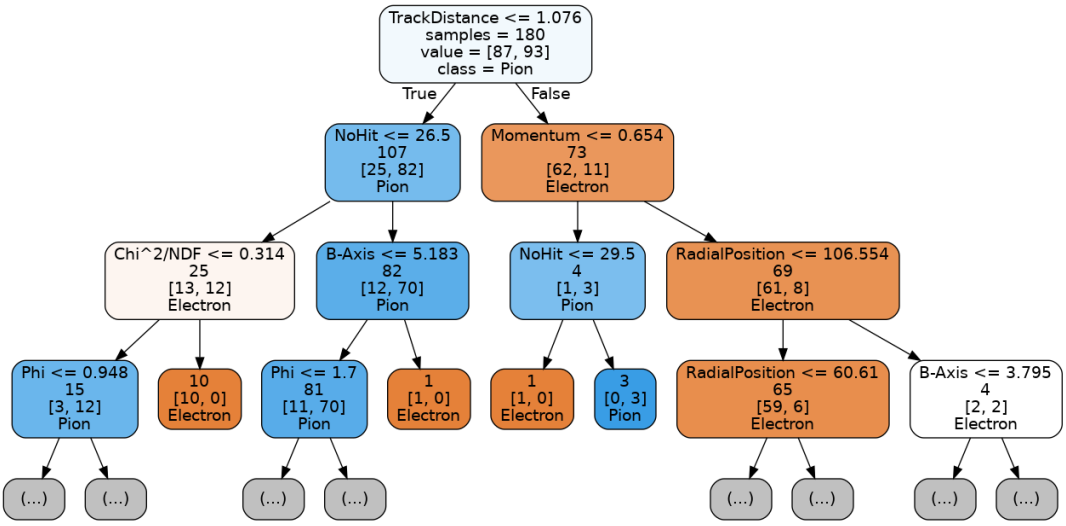


Figure 3.3: Graphical representation of a decision tree to distinguish between electrons and pions with the RICH detector. The maximum depth of the representation is 3. The first line states the question which is to be answered by the data with True of False. If the answer is true, move to the left in the diagram (otherwise to the right). The values in the diagram show how many electrons and pions fall into this node $[N_{\text{Electrons}}, N_{\text{Pions}}]$. The color of the boxes is determined by the ratio of electrons to pions at the node. Blue belongs to more pions, white is a similar proportion and red belongs to more electrons.

The nodes represent the decision-dependent branch points and lead to the next decision level via the branches. For true-false decisions, so called binary decision nodes, branching occurs in two branches. The number of decision levels can vary from decision tree to decision tree. In Figure 3.3, the decision tree has been mapped up to the third decision level. The final level is the result level, this is shown in Figure 3.3 without further arrows. The advantage of the decision tree is that decision paths can be represented clearly and comprehensibly.

Information Gain

To make the right decisions with which the decision tree can most effectively make a prediction, the concept of entropy is needed. After each question in the decision tree in Figure 3.3 information is obtained. In the A,B,C,D example gained information is calculated by the number of yes-no questions that are needed on average to classify the letters. The missing information can also be calculated with the definition of the entropy, equation 3.1.

$$\mathrm{H}(T) = -\sum_{i=1}^{J} p_i \log_2 p_i \tag{3.1}$$

$p_1, p_2, \ldots$ are fractions that add up to 1 and represent the percentage of each class present in the child node that results from a split in the tree. When changing the decision variables x like "NoHits <= x" in the decision tree from Figure 3.3, the gained information changes. The tree-generation algorithm computes it as a function of x and maximizes it. The information that can be computed at each node of the tree represents how many yes-no questions must be asked on average to make a classification.

$$\overbrace{IG(T,a)}^{\text{information gain}} = \overbrace{\mathrm{H}(T)}^{\text{entropy (parent)}} - \overbrace{\mathrm{H}(T \mid a)}^{\text{sum of entropy's (children)}} \tag{3.2}$$

In the following, the information gain of a decision is calculated on the basis of the classification of electrons and pions. Figure 3.4 shows a sample of equal numbers of pions and electrons distributed in the two-dimensional space. Since 50% of the data are electrons and 50% are pions, the missing information H becomes 1, which means that on average there is one question to ask (like: "is it an electron?") if a random particle with the two detector properties is labeled as an electron. The missing information H of the parent node is calculated using the equation 3.1 as follows:

$$\mathrm{H} = -0.5 \cdot \log(0.5) - 0.5 \cdot \log(0.5) = 1 \tag{3.3}$$

The question whether it is an electron can not be asked, therefore it must be asked for the measured values of the particle. If in Figure 3.4 it is asked whether the feature RICH ANN is greater than 0.6 (see red line), then the answer will be "yes" for 72.2 % of electrons and for 14.7% of pions. The answer "no" is obtained for 27.8 % of electrons and 85.3 % of pions. If the answer to the question is "yes", then it is most likely a pion, but there are other classification decisions to be made (yes-no questions). If the answer is yes then the probability P_e that it is an electron $\frac{0.722}{0.722+0.147} = 0.83$ the probability that it is a pion is 0.17.

$$\mathrm{H}_{\mathrm{Yes}} = -0.83 \log(0.83) - 0.17 \log(0.17) = 0.66 \tag{3.4}$$

$$\mathrm{H}_{\mathrm{No}} = -0.25 \log(0.25) - 0.75 \log(0.75) = 0.82 \tag{3.5}$$

The answer yes to the question "RICH-ANN >0.6?" is given in 43.45 % of the analyzed data points, the answer no is 56.55 %. Therefore the total entropy H results in

$$H = 43.45\% H_{Yes} + 56.55\% H_{No} = 0.75 \tag{3.6}$$

With the help of equation 3.2 the information gain is calculated to $1 - 0.75 = 0.25$. This value can be read from the red line in Figure 3.4 on the right.

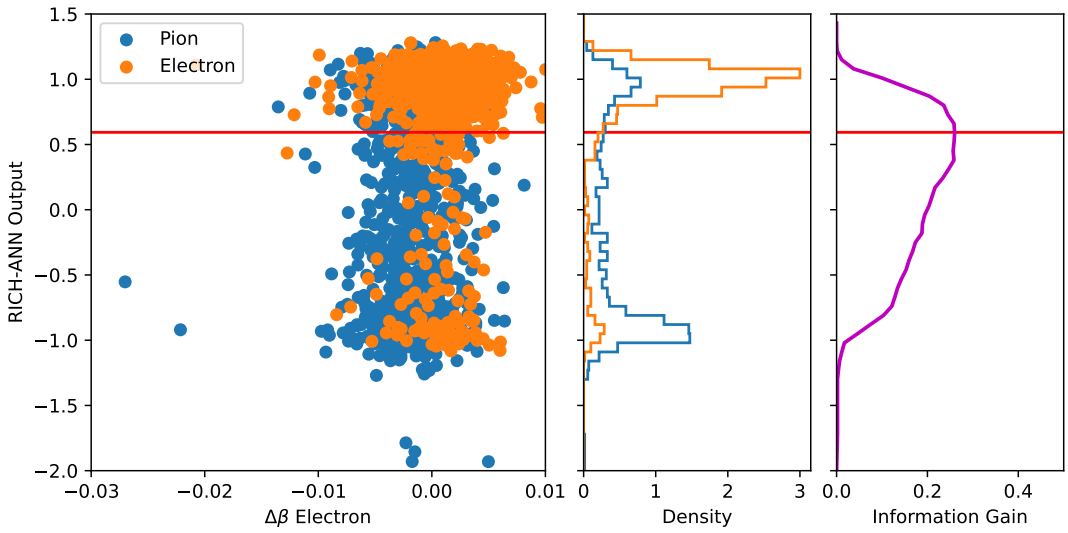


Figure 3.4: The left panel shows the distribution of electrons and pions in the two dimensional detector space. The RICH-ANN output is a detector variable presented in chapter 4. The red line shows a cut that is applied so that particles below the line are classified as pions and particles above the line are classified as electrons. The red line shows the cut of the maximum information gain. The middle panel shows a projection of the histogram along the y axis. The right panel shows the information gain of a cut at the respective position.

On the basis of the histogram in Figure 3.4 it can be found already with intuition that a cut, behind (under) the maximum for the pions (in orange), is suitable. Using the information gain discussed in equation 3.3 as a metric, the position of the cut can be optimized.

3.2 Random Forest Classifier

A Random Forest is a classification and regression method consisting of several uncorrelated decision trees. Each decision tree handles exactly one basic classification task, in order to represent more complex issues and to increase the quality of the classification. Each decision tree by itself does not have to provide a perfect classification

in a decision forest. By cleverly combining the decisions of many Decision Trees, an optimized classification is created. The method is illustrated in Figure 3.5.

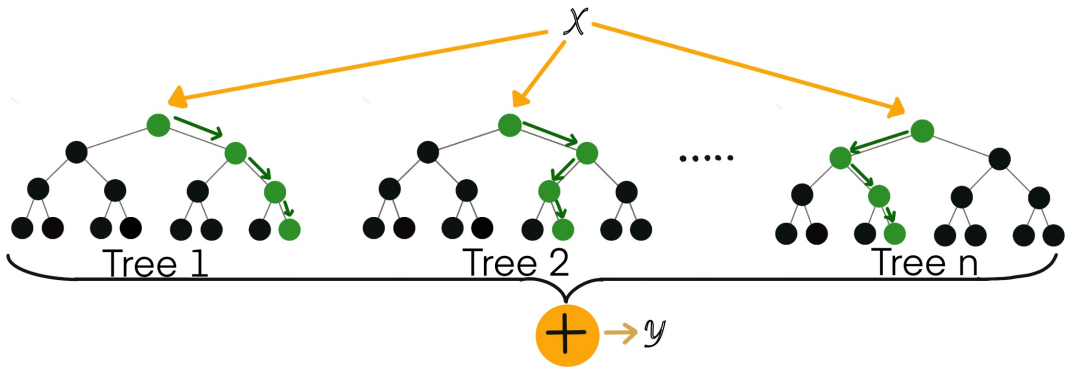


Figure 3.5: Concept of Random Forest Classifiers. n uncorrelated decision trees vote on the label of an input x. The sum of the votes is the output y of the classifier.

There are different models for training the individual random trees, the simplest is the **bagging**. Bagging repeatedly (B times) selects a random sample of the training dataset with replacement (an element may appear multiple times in the sample) and fits trees to these samples. This procedure leads to better model performance than a single decision tree because it decreases the variance of the model, without increasing the bias. This means that while the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees is not, as long as the trees are not correlated. Simply training many trees on a single training set would give strongly correlated trees (or even the same tree many times, if the training algorithm is deterministic); bagging is a way of de-correlating the trees by showing them different training sets. Some advantages (in green) and disadvantages (in red) of Random Forest are listed below.

- Random Forest usually does not overfit, which means more decision trees lead to better prediction. This makes fine tuning the classifier particularly easy.
- The classifier trains very fast: This advantage results from the short training or construction time of a single decision tree and from the fact that the training time for a Random Forest increases linearly with the number of trees.
- The evaluation of a test sample happens on each tree individually and can therefore be parallelized.
- It is very efficient for large data sets (many classes, many training examples, many features).
- While Random Forests often achieve higher accuracy than a single decision tree, they sacrifice the intrinsic interpretability present in decision trees. This interpretability is one of the most desirable qualities of decision trees. It allows developers to confirm that the model has learned realistic information from the

data and allows end-users to have trust and confidence in the decisions made by the model. For example, following the path that a decision tree takes to make its decision is quite trivial, but following the paths of tens or hundreds of trees is much harder.

3.3 XGBoost Classifier

Gradient boosting and XGBoost are machine learning techniques that have already been successfully applied to high energy physics and in particular have contributed to the discovery of the Higgs boson. These methods usually outperform the Random Forest method and therefore have high potential for the particle identification investigated in this thesis. Just like the Random Forest, the XGBoost algorithm combines different decision trees. A single decision tree is called a weak learner. However, there are other even simpler weak learners, such as separating data in N-dimensional data space by an N-1 dimensional surface. For the two-dimensional case, this N-1 dimensional surface is a simple line in the parameter space. With the help of such a simple weak learner, the XGBoost algorithm is explained step by step. The search for a suitable classifier is an **optimization problem in function space**, since a classifier is a function that labels each input, for example whether it is an electron or a pion. XGBoost uses the Newton optimization method for functions where a cost function $L(x)$ that outputs the quality of the classification is to be minimized. The Newton optimization method finds the zeros of the derivative of the function $L(x)$ (see Figure 3.6). Gradient boosting, unlike XGBoost, uses the root finding algorithm in the function space to minimize the loss function.

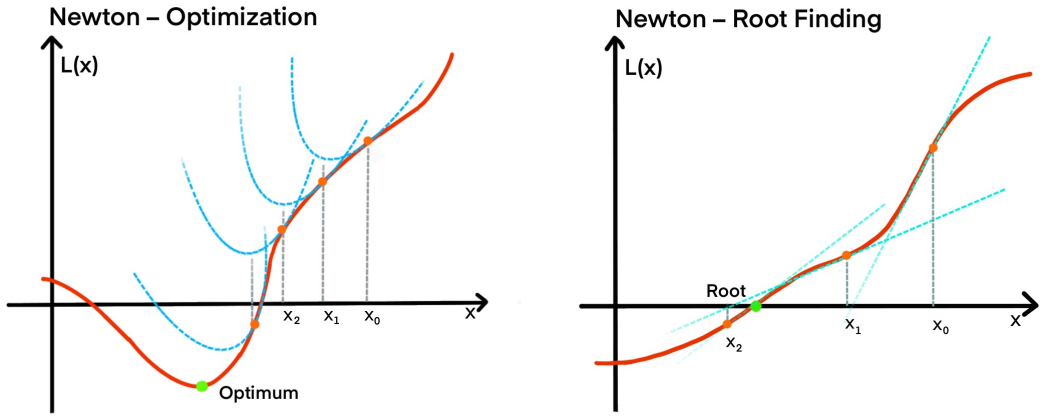


Figure 3.6: Concept of the Newton method for optimization (left) and finding zero points (right). For finding extreme points, parabolas are fitted to each iteration point. For finding roots of the function $L(x)$, straight lines are fitted.

The Newton optimization method performs the iteration in equation 3.7. This also

appears in the XGBoost algorithm and will be shown later.

$$x_{k+1} = x_k - \frac{L'(x_k)}{L''(x_k)}. \tag{3.7}$$

The complete XGBoost algorithm is explained in the following by means of a relatively simple classification problem.

1. First, a training data set is needed: $\{(x_i, y_i)\}_{i=1}^{N}$. In the example, the x_i are two dimensional data points which are distributed in Figure 3.7 left as orange and blue points across the surface. y_i is the corresponding label of the points, here the value 1 is assigned to electrons and the value -1 to pions. Furthermore a differentiable loss function $L(y, F(x))$ is needed. This is defined in this simple example as $L(y, F(x)) = (y - F(x))^2$ (a least square function). $F(x)$ is the classification function and is shown as a colorcode on the left in Figure 3.7. In the left graph of Figure 3.7, a weak learner has been trained to minimize the loss function. The weak learner classifies particles above the line as electrons and particles below the line as pions. The classifier makes some errors and, for each particle it misclassifies, the loss function in this first example increases by $(y_i - F(x_i))^2 = 4$. The function of a single weak learner is called ϕ, the set of all possible weak learners is the set of all possible degrees in the 2-dimensional space and is abbreviated as $\boldsymbol{\phi}$. Figure 3.7 on the right shows possible example classifiers. Furthermore, the number M of weak learners to be combined and a learning rate α must be defined at the beginning. In this example $M, \alpha = (40, 0.1)$ is used.

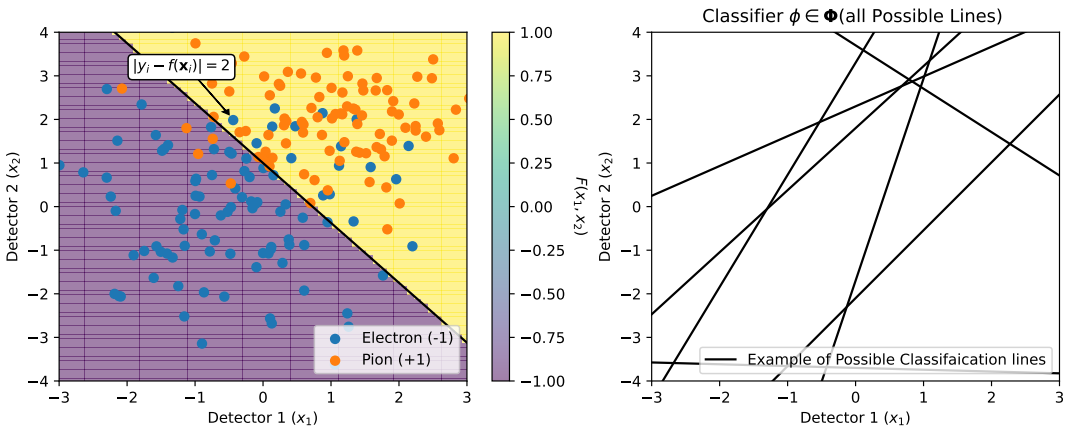


Figure 3.7: Left: Electrons and pions distributed in two-dimensional space. A weak classifier, here a dividing line, classifies electrons and pions. Particles above the line are assigned to electrons, particles below the line to pions. Electrons have the value -1 and pions the value 1. Right: example for possible weak classifiers which are combined in the XGBoost algorithm. Of all possible separating lines, the one used on the left is the optimum.

2. At the beginning of the algorithm, the model is initialized with a constant value. A constant function is searched which minimizes the loss function. In this example it

is the 0 function, because as many electrons (+1) as pions (-1) should be classified.

$$\hat{f}_{(0)}(x) = \arg\min_{\theta} \sum_{i=1}^{N} L(y_i, \theta). \tag{3.8}$$

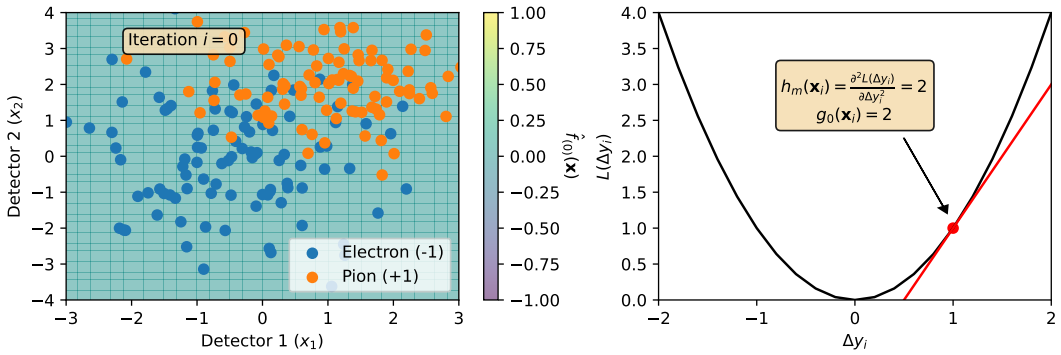


Figure 3.8: Left: Constant initialization function $\hat{f}_{(0)}(x)$ assigns the value 0 to each data point. Right: Loss function of the example. The greater the difference between the classified value and the actual value y_i, the greater the loss ($\Delta y_i = f(x_i) - y_i$). The first and second derivatives at position x_i are shown in the box.

3. The following steps are iterated for $m \leq M$. In each step the Hessian $\hat{h}_m(x_i)$ and the gradient $\hat{g}_m(x_i)$ are calculated for each data point (blue and red). In the example, the difference between the classification function $f(x_i)$ and y_i must be calculated for each data point. For the initial case, shown in Figure 3.8, this value is either -1 or 1. This must then be inserted into the loss function and the 1st and 2nd derivative of the function at the point 1 or -1 must be calculated. By choosing a least square loss function the second derivative is always 2.

$$\hat{g}_m(x_i) = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=\hat{f}_{(m-1)}(x)} \tag{3.9}$$

$$\hat{h}_m(x_i) = \left[\frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2}\right]_{f(x)=\hat{f}_{(m-1)}(x)} \tag{3.10}$$

4. Next, a weak learner is fitted to the training set $\left\{x_i, -\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)}\right\}_{i=1}^{N}$ by minimizing the same loss function. The loss function can now be calculated faster with a 2nd order Taylor using the equation 3.11.

$$\hat{\phi}_m = \arg\min_{\phi \in \mathbf{\Phi}} \sum_{i=1}^{N} \frac{1}{2} \hat{h}_m(x_i) \left[-\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)} - \phi(x_i)\right]^2. \tag{3.11}$$

This is analogous to the Newton optimization problem. A "parabolic step" is to be taken in the direction of the optimum. In equation 3.7 the step is simply subtracted from the previous iteration value. Here $\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)}$ cannot be simply subtracted from the previous step, because the previous step is not a real number as in equation 3.7 but a function. Therefore, following equation 3.11, the function which can go this next step in the direction of the optimum has to be searched first.

5. As a last step, the model is updated using the learning rate alpha:

$$\hat{f}_{(m)}(x) = \hat{f}_{(m-1)}(x) + \alpha\hat{\phi}_m(x). \tag{3.12}$$

   Steps 3-5 are repeated iteratively.

6. The final classification function is now the model at the m-th step $\hat{f}_m(x)$

$$\text{Output: } \hat{f}(x) = \hat{f}_{(M)}(x) \tag{3.13}$$

Figure 3.9 shows how the output function $\hat{f}(x)$ adapts more and more to the data after the different iterations and outputs a good classification. The progressive increase in quality with the clever combination of weak learners can be observed.

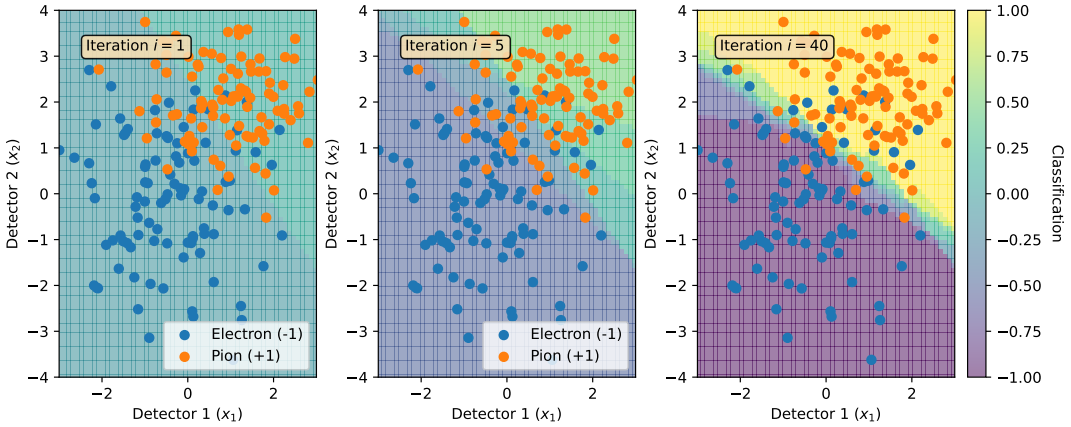


Figure 3.9: Output of the XGBoost algorithm for the discrimination of electrons and pions with a combination of linear discriminators after different iteration steps.

The XGBoost classifier used in this thesis does not use the linear discrimiator but decision trees which are fit to the data. The algorithm is exactly the same, so that the optimal function can be found as a combination of decision trees. Since the whole classifier is always a sum of many individual trees, the classification function can be executed in parallel. This means that the individual trees decide at the same time and the result is then summed up. Parallelizability is especially important for the CBM experiment because of the high data stream and the corresponding high demands on the computing time.

3.4 Artificial Neural Networks (ANN)

The basic idea of artificial neural networks is based on a principle inspired by biology: elements that are very simple in themselves can be combined to create constructs of arbitrary complexity. In this model, our brain consists of a multitude of networked neurons, each of which assumes one of two states: activated or not activated, or the neuron sends a signal or remains silent. A neuron is activated when a sufficient number of the neurons connected to it are in turn activated. When the model is generalized, a neuron can assume not only two states, but any value between one and zero. A neuron is a function that has a number of weighted inputs and an output. Neural networks are particularly suitable for the processing and classification of **images**, **audio** and **text** [21]. So in general for classification algorithms where the input space is very large. In essence, a neural network is a complicated function with many free parameters that is fitted so that the output label matches the inputs. Since the reconstructed detector information in the CBM experiment contains about 20 variables that are mapped to one output (the particle label), the input is relatively small compared to the input of image or text files. In this thesis neural networks do not contribute to the improvement of the classification. Nevertheless, modern methods like XGBoost are compared with an already used neural network. The architecture of this network is shown in Figure 3.10: nine input neurons are followed by a hidden layer with 18 neurons which converge in one output neuron.

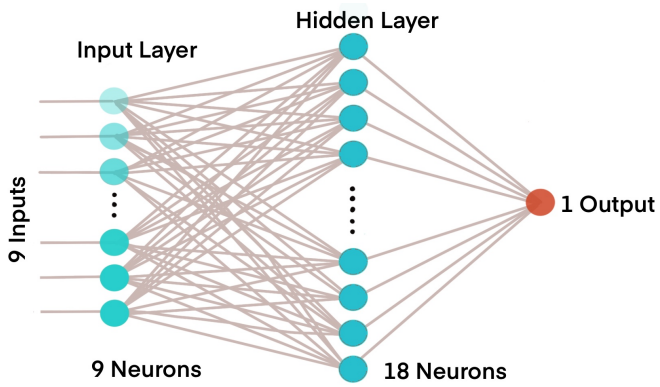


Figure 3.10: Graphical representation of a feed forward neural network with 9 inputs and a hidden layer with 18 neurons. The network has an output layer with one neuron.

A neural network is a function $f(x)$ that maps an input, here a vector with 9 detector variables, to an output. Figure 3.10 is a graphical representation of 3.14. Here $\mathbf{W_1}$ is a 9x18 matrix and $\mathbf{W_2}$ is a 9x1 matrix. $\mathbf{b_1}$ is an 18 dimensional vector and $\mathbf{b_2}$ is a 1 dimensional vector. These matrices and vectors contain the fit parameters.

$$f(x) = \sigma\left(\mathbf{W_2}\sigma(\mathbf{W_1}x + \mathbf{b_1}) + \mathbf{b_2}\right) \tag{3.14}$$

$$\text{Sigmiod Function } \sigma(r) = \frac{1}{1+e^{-r}} \tag{3.15}$$

In principle, the neural network from Figure 3.10 is a fit function with (9·18+9·1+18+1) = 190 variables. At this point there are many words to be said about the function of fitting, convolution neural networks, graph neural networks etc. In this thesis, the neural network is mainly considered as a given function $f(x)$ to which other methods are compared. Therefore, the most important information is that the neural network presented in this thesis is a fit function with nine input and one output value, which has in the order of 100 fit parameters.

3.5 Receiver Operating Characteristic (ROC)

The Receiver Operating Characteristic (ROC) is a method for the evaluation and optimization of classification methods. Each classification method is tested by calculating the ROC, which is generated from test datasets. The histogram in Figure 3.11 shows electrons and pions mapped by the TOF to a $\Delta\beta_{\text{electron}}$ value. The TOF detector is thus a classifier because it can distinguish electrons from pions by determining a cut value. In Figure 3.11 above the cut value is set to zero and particles having $\Delta\beta_{\text{electron}}$ values greater than 0 are classified as electrons. Particles having a value less than 0 are classified as pions.

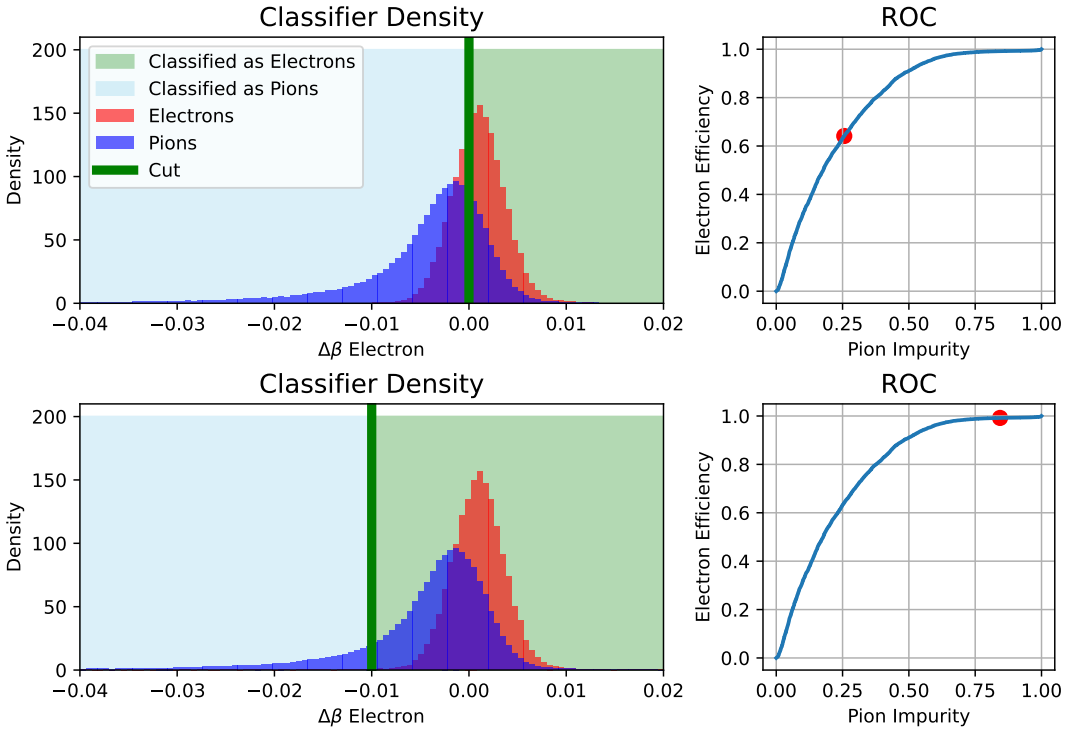


Figure 3.11: Left: Different distributions for the $\Delta\beta_{\text{electron}}$ values for electrons and pions measured with the TOF. Values right of the green line (cut) are classified as electrons. Values left of the line are classified as pions. Right: ROC for different cut values. The red dots belong to the cuts shown on the left.

In the upper part of Figure 3.11 it can be seen that the classifier erroneously identifies some pions electrons. Some of the blue bars of the histogram are in the green side. There are two important quantities that indicate how well the classification worked: First, how many of the electrons are correctly designated as electrons (electron efficiency), and second, how pure the data classified as electrons are from pions (pion impurity). The electron efficiency is the proportion of the summed red bars that are in the green area. In the upper case this is a little over 60%. The pion impurity is the percentage of blue bars in the green area. In the upper case this is about 25%. If now the cut value is shifted, new pairs of electron efficiency and pion impurity will result. If the cut value shown in green is varied through the entire length of the x-axis (from right to left), the two values can be calculated for each cut value and entered in the diagram (see Figure 3.11 on the left). The red point which belongs to the respective cut value then moves along a curve, the so-called ROC from bottom to top. This curve runs inevitably through the points a) (0,0) and b) (1,1) because these are the two extreme cases: a) all particles are called pions (cut rightmost) or b) all particles are classified as electrons (cut leftmost). The ROC is convex which is an important property that will be needed in the course of the thesis.

The general term for electron efficiency is True Positive rate (see Figure 3.12). Since electrons are labeled True (pions False). The prediction that a particle is an electron is labeled positive. Therefore the pion impurity results in the False Positive rate. If the ROC is close to a diagonal (in orange), the classifier cannot distinguish well between two classes. This would be the case, for example, if the blue histogram overlapped with the red one. Any prediction is therefore random. An optimal classifier recognizes all electrons correctly without designating a pion as an electron. The electron efficiency would be at 100% and the pion impurity at 0. This point is shown in blue in Figure 3.12. The further the ROC runs at this point the better the classifier is. When classifiers are compared in this thesis, each classifier has a ROC. In the example of the Figure3.12 the classifier with the light blue ROC is better than the one with the green or orange ROC.

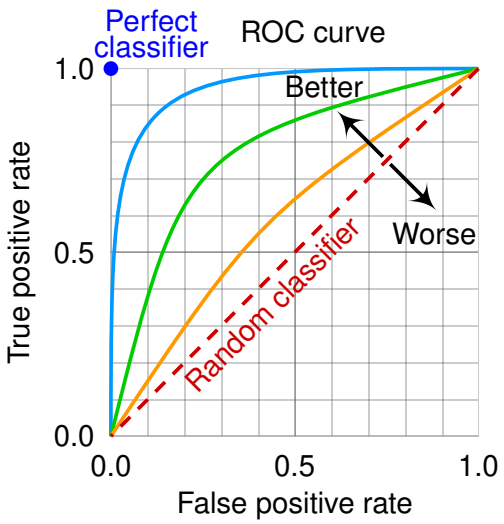


Figure 3.12: Diagram illustrating the evaluation and interpretation of an ROC curve [22]

4 Conventional Classification Methods

In this thesis, modern machine learning classifications methods are compared with the conventional classifications methods, as they are currently implemented in CBMROOT. The conventional electron ID method consists of two main steps. First, so-called **quality cuts** are applied to the tracks in order to use only those of sufficient quality for particle identification. **Particle identification cuts** (PID-cuts) are used to isolate electrons from other particles. The cuts are chosen, so that as few background particles like pions and protons are used for the pair analysis. The three most important quality cuts for the tracks are the following:

1. χ^2 **to Vertex Cut**: The hits of a track in STS and MVD are fitted with a track function depending on 5 parameters. Two of these variables encode the distance to the collision point. Since dielectrons are generated from thermal photons and decays of vectormesons directly at the vertex (point of collision), the reconstructed track should match the position of the vertex within the variance. Since conversion electrons generally do not originate directly from the vertex, this cut also separates conversion electrons from signal electrons. In the conventional methods this cut is set to 3.

$$\chi^2 \hat{=} \frac{\text{Distance to Vertex}^2}{\text{Variance}} < 3 \tag{4.1}$$

2. **Quality Tracking** : Since, as mentioned above, the track fit function has 5 free parameters, at least 3 hits are needed, each of which provides 2D information to accurately identify the shape of the track. The more hits a track generates in STS and MVD the higher its quality. Therefore it is only suitable to select tracks that have at least 4 hits in the tracking detectors.

3. **Quality PID Cut**: In the conventional methods it is required that each track hits all 3 particle identification detectors TRD, RICH and TOF. When more hits belong to a RICH ellipse, the quality of the fit of the ellipse increases. Usually at least 6 hits are required because an ellipse has 5 free fit parameters. For the TRD at least 3 hits are required. The point of discussion in this thesis will be whether tracks that have only been seen by two or one PID detector can be classified.

After the quality cuts have selected the best tracks, the particle identification cuts decide whether a track is identified as an electron or as a different particle. The electrons have to be separated from the large background of pions and protons. For

the classification it is less important which background particle it is, but only if the track is an electron and can be used for the pair analysis. In the conventional methods each PID detector RICH, TRD and TOF decides separately if the particle is an electron or not. Only if all are in favor of an electron, it is classified as such. In the following, the individual decision mechanisms of the PID detectors will be discussed.

4.1 PID with RICH

For the particle identification of the conventional method, first the 8 measurement variables of the RICH detector together with the momentum of the particle are mapped to values between -1 and 1 with the help of a neural network. Hereby values closer to -1 belong rather to a pion and values closer to 1 to an electron. Besides the 5 degrees of freedom of the ellipse, the momentum, the number of photon hits and the distance of the photon ring to the track are given as input to the network. The network used is described in section 3.4. The network was implemented in the software of the experiment (CBMROOT) in 2008 and therefore has great potential for improvement due to the immense advances in machine learning in the last decade.

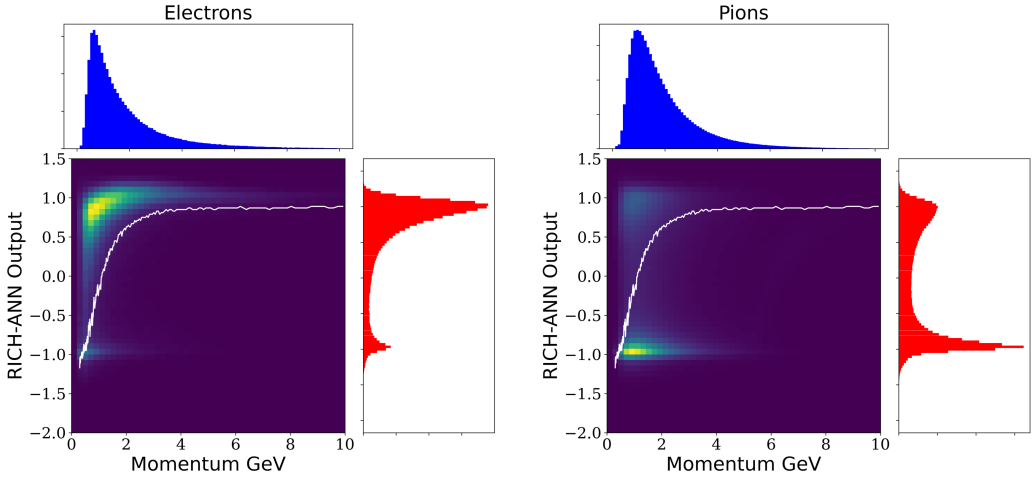


Figure 4.1: Two-dimensional histograms of the output of the RICH ANN vs. the momentum. The left panel shows the output for electrons and the right panel for pions. The projections on the one-dimensional histograms for momentum and ANN output are added at the sides. The white line shows the two-dimensional cut. Entries above the line are classified as electrons. Entries below the line are classified as pions.

Figure 4.1 shows two-dimensional histograms of the output of the ANN implemented in CBMROOT versus the momentum of the particles. On the left of Figure 4.1 is the output of the ANN for electrons and on the right is the output for pions. Since due to the similar masses of electrons and pions, these are the two most difficult particles to

distinguish for the PID detectors. Therefore it is sufficient as a good approximation if the RICH focuses on the separation of pions and electrons and ignores the results of protons in the neural network. Protons can also be filtered out by the TOF. Particles having momentum and ANN-output above the white line in Figure 4.1 are classified as electrons, particles below the line as pions. The right figure shows that, in contrast to the left one, the pions cluster below the white cut line. The cut line is calculated so that for each momentum an electron efficiency of 90% is achieved. This means that the RICH detector misclassifys 1 in 10 electrons as pions. The RICH detector distinction ability for electrons and pions decreases with an increase in momentum of the particles. Figure 4.2 shows the histogram from Figure 4.1 with normalized columns. It can be seen that for low momenta the network assigns the value 1 to electrons and the value -1 to pions. However, this changes for high momenta so that the network also assigns the value 1 to pions and the particles can no longer be distinguished. The information that the RICH detector loses its classification ability for larger momenta cannot be read directly from Figure 4.2. It can only be seen that the classification with the ANN trained in 2008 does not work well for higher momenta and may therefore also be a consequence of an imperfect classifier. In chapter 6 the same plot is investigated for a more modern classification method presented in this thesis.

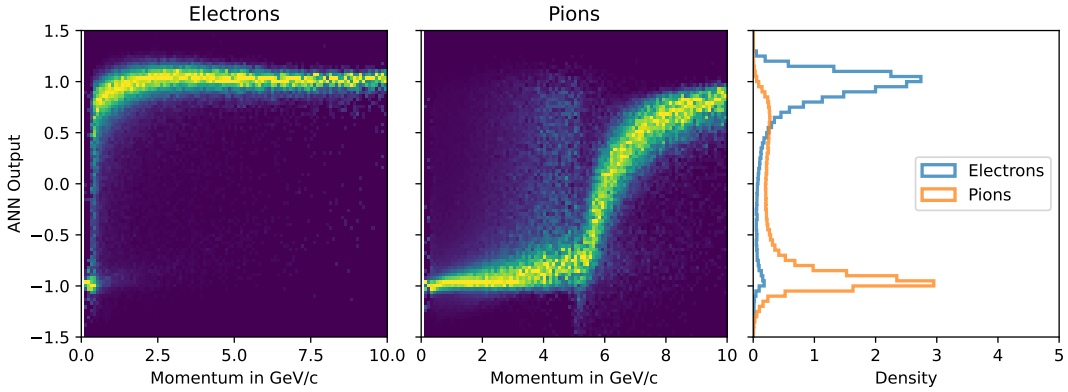


Figure 4.2: Electron identification value, which represents the RICH data of electrons (left) and pions (center) by a ANN implemented in CBMROOT, for different momenta. The assigned rings for both distributions require at least 6 hits on the photon plane. Both distributions are normalized to unity for each momentum bin individually. On the right is the momentum integrated spectrum of the ANN output for electrons and pions. Here the individual columns were weighted with the predicted frequency of the respective momentum.

4.2 PID with TRD

Using the different energy deposition for electrons and pions in the TRD (see Figure 2.10) the probability that a given TRD track belongs to an electron can be calculated. Figure 4.3 shows a correlation matrix of the energy depositions in the 4 layers based on simulated data. Note the negligible off-diagonal correlation, showing that The variables can be treated as independent. The probability that a particle Φ has the energy depositions E_1 to E_n can be written as a product of single probabilities:

$$P(E_1, E_2, \dots E_n|\Phi) = \prod_{i=1}^{n} P(E_i|\Phi) \tag{4.2}$$

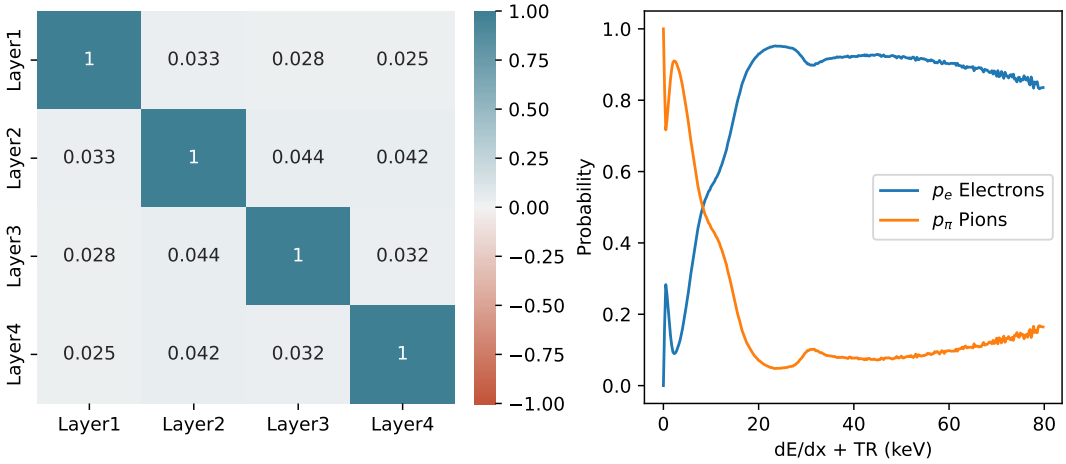


Figure 4.3: Left: Correlational matrix of the energy deposition in the individual detector layers of the TRD. Note that the layers are only slightly correlated. Right: Probability for finding an electron (p_e) and for finding a pion p_π for different energy depositions in one layer.

For the probability of measuring an electron p_e in a single TRD layer, the probability of measuring an electron must be divided by the size of the total possibility space. This results from the fact that the particle is either a pion and an electron to $p_e + p_\pi$.

$$L_e = \frac{p_e}{p_e + p_\pi} \tag{4.3}$$

The probability that a certain energy deposit in the TRD belongs to a pion p_π or to an electron p_e can be seen in Figure 4.3 on the right. For the probability that all n hits in the TRD layers belong to an electron, the product of the probabilities of the individual layers must be divided by the total possible space, which is again the probability that the whole track belongs to an electron or to a pion.

$$L_e = \frac{\prod_{i=1}^{n} p_{ei}}{\prod_{i=1}^{n} p_{ei} + \prod_{i=1}^{n} p_{\pi i}} \tag{4.4}$$

On the calculated probability of an electron track from the four energy depositions in the TRD, a cut can be applied. Figure 4.4 shows the same figure as Figure 4.1, but this time with the electron likelihood instead of the ANN output. The white line shows the 2D cut boundary, which classifies particles above it as electrons and particles below it as pions. The two-dimensional histogram has logarithmic entries.

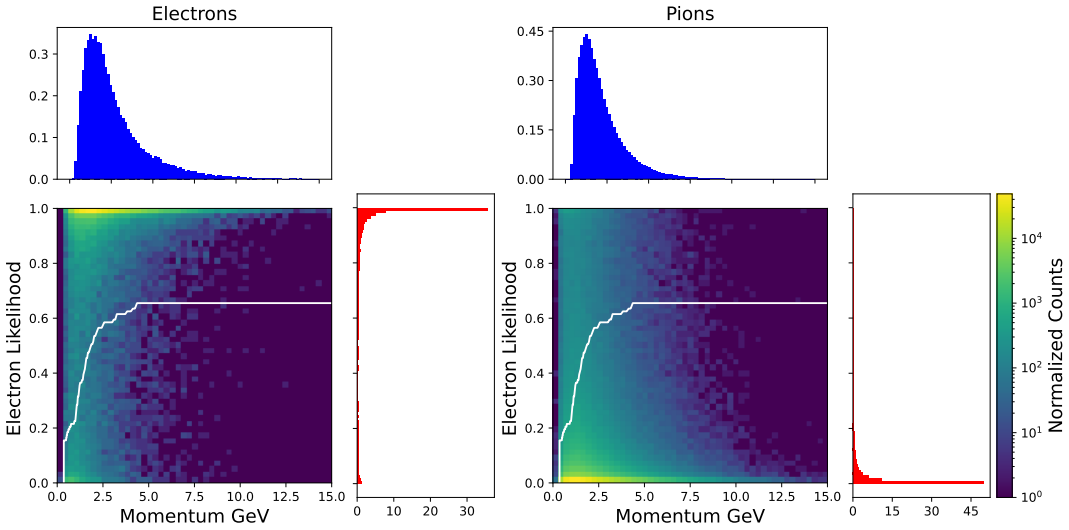


Figure 4.4: Two-dimensional histograms of TRD likelihood versus momentum. The left panel shows contributions for electrons and right panel for pions. The projections on the one-dimensional histograms for momentum and likelihood are added at the sides. The white line shows the two-dimensional cut. Entries above the line are classified as electrons. Entries below the line are classified as pions.

Most electrons in Figure 4.4 are assigned to the value 1, which can be seen from the projection of the 2D histogram onto the y-axis. Pions, on the other hand, are very sharply assigned to the value 0. The white cut boundary was calculated so that the TRD has an electron efficiency of 80%. This means that 1 out of 5 electrons is wrongly classified as a pion.

4.3 Conventional Classification with TOF

The time of flight of a particle is the only measurable variable that can be used for particle identification with the TOF. Together with the track length, this can be calculated into a particle speed. The length of the track is extracted from the fit to the trajectory of the particle, which in general can be curved. For each particle the β-value (velocity by speed of light) can be calculated from the momentum and the mass of the particle. Assuming the mass of the particle is 511 keV (mass of an electron), the $\Delta\beta_{\text{Electron}}$ value can be calculated with formula 4.5. This gives the difference between the β-value measured by the TOF and the β-value predicted by the tracking if it is an electron. P and E are the momentum and the energy of the

particle. The energy is calculated from the momentum with the assumption that the particle has the mass of an electron.

$$\Delta\beta_{\text{Electron}} = \frac{\text{Track Length}}{\text{Time of Flight}}\frac{1}{c} - \frac{c \cdot P}{E} \tag{4.5}$$

It is expected that electrons have $\Delta\beta_{\text{Electron}}$ around 0. The heavier the particle, the lower the expected $\Delta\beta_{\text{Electron}}$ value. Particles heavier than electrons should have on average $\Delta\beta_{\text{Electron}}$ values smaller than 0. Figure 4.5 shows histograms of $\Delta\beta_{\text{Electron}}$ for electrons and various hadrons, which are the main components of the background tracks.

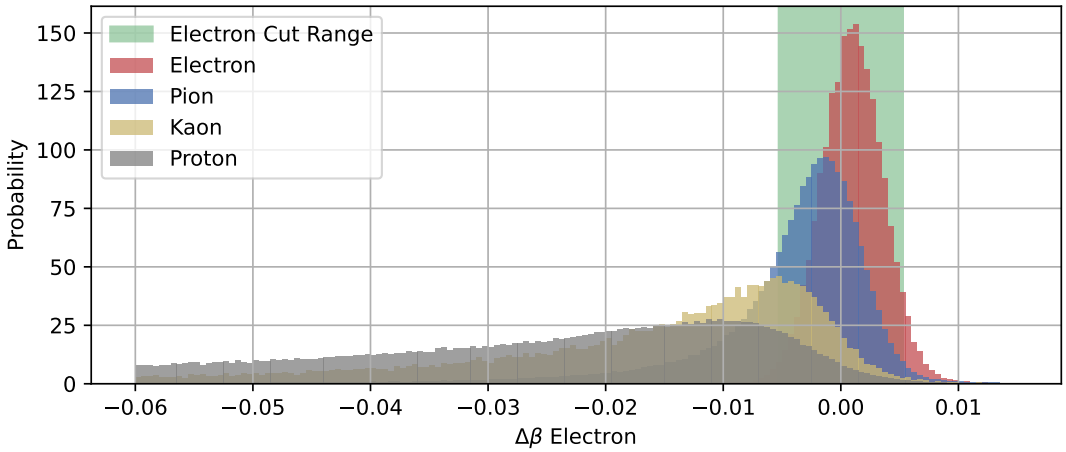


Figure 4.5: Normalized histograms of the $\Delta\beta_{\text{Electron}}$ value (see Equation 4.5) for simulated tracks of electrons, kaons, pions and protons. Tracks with values within the cut range (green) are classified as electrons by TOF. For electrons a $\Delta\beta_{\text{Electron}}$ value of 0 is expected.

In Figure 4.5 it can be seen that the histograms are ordered from heavy to light. The green area shows which tracks the conventional methods assign to electrons. This area is symmetric around 0.

4.4 Combination of the PID Cuts

The decision mechanisms of the PID detectors are momentum dependent, so that in the following only the decision mechanisms for momenta between 1.95 and 2 GeV are investigated. From Figure 4.1 and 4.4 a cut value for a momentum of 2 GeV/c can be read from the white line. For example, for 2 GeV/c a cut value of 0.61 can be found for the RICH ANN. The TOF classifies particles with $|\Delta\beta_{\text{Electron}}| < 0.00528$ as electrons, independent of the momentum (see green area Figure 4.5). Figure 4.6 a) shows a scatter plot of electrons and pions with momenta between 1.95 and 2 GeV/c distributed within the two-dimensional parameter space of the RICH-ANN

and $\Delta\beta_{\text{Electron}}$ value of the TOF. In the two-dimensional space electrons cluster in the upper right and pions cluster in the lower left. The red lines show the cut boundaries described above. Particles inside the green square are classified as electrons. Particles outside the square are classified as pions. Note that the cuts refer to the projections of the two-dimensional histogram on the x and y axes. For momenta of 2 GeV/c the cut value for the TRD likelihood is 0.5. In Figure 4.6 right three classification variables of the PID detectors TOF, RICH and TRD are shown as a three-dimensional space. Electrons and pions are distributed in different corners of this space. The green cube shows the region classified as electrons by the conventional methods. Particles outside the cube are classified as pions. For different momenta, the size of this cube would change, but also the distribution of electrons and pions.

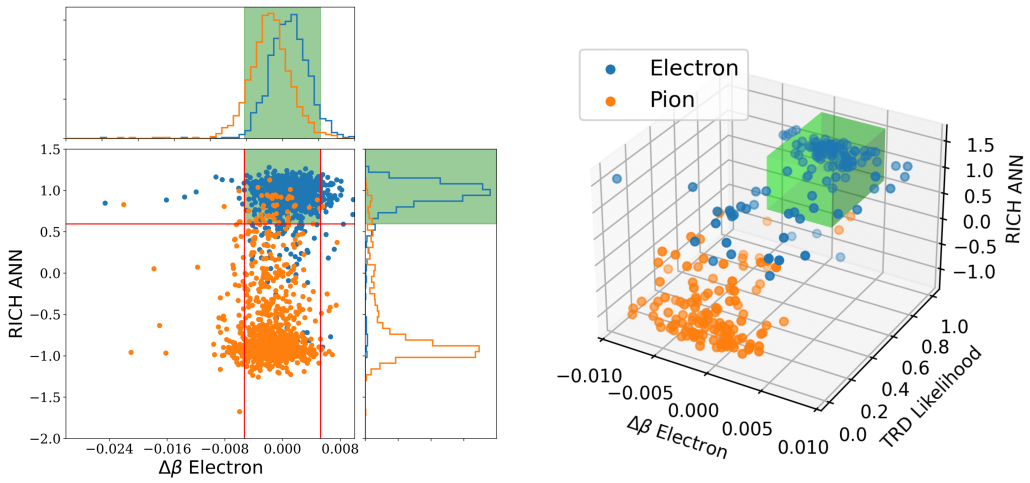


Figure 4.6: Left: Scatter-plot of pions and electrons in two-dimensional detector space. On the x-axis the $\Delta\beta_{\text{Electron}}$ value is plotted and on the y-axis the RICH ANN output of the tracks. On the sides are histogram projections of the data. The plotted data have a momentum between 1.95 and 2 GeV. The red lines show the conventional cut values. Values inside the green square in the upper right are identified as electrons. Right: Scatter-plot for electrons and pions in three-dimensional PID detector variable space. Particles inside the green cube are classified as electrons. The left plot is a projection of the right 3D plot onto the wall at the back right.

4.5 Computing Methods

The so called "**PairAnalysisPAckage**" or "PAPA" framework is a C++ written framework for dielectron analysis. It was implemented in CBMROOT in 2005 by Julian Book based on the dielectron analysis framework of the ALICE collaboration and has been improved and extended since then [4]. It is programmed in a very object-oriented way and allows a fast processing of the data. It is possible to access the Montecarlo information directly, which allows to test the efficiency of the analysis methods and the cuts. Since there are only a few libraries implemented in ROOT

for machine learning, the data was converted into a Python readable format at the beginning of the research presented in this thesis. This posed a particular difficulty due to CBMROOT's own data types in the storage formats. Instructions for the developed procedure can be found under [23]. Python has a large number of libraries that can be used quite easily for machine learning (e.g. NumPy, SciPy, ScikitLearn, PyBrain etc.). The large community involved in the research of machine learning with Python and the simple syntax of the language make for developer friendly programming.

5 Significance of PID for Pair Analysis

In the conventional methods for pair analysis, for example, an electron efficiency of 80% is required for all momenta for the TRD. This means that 4 out of 5 electrons are classified as electrons. Under these conditions the hadron impurity is minimized to a large extent. In this chapter another method is proposed than maximizing the pion suppression for a given electron efficiency. The new method maximizes the significance of the measurement for the dielectron analysis and is introduced visually in the following. In Figure 5.1 a) electrons and pions are shown which were detected by two different detectors. The measurements of the individual particles are plotted on the X and Y axes respectively. It can be seen, that the electrons are clustered in the upper right of the graph and the pions are clustered in the lower left. The pions and the electrons are not completely separated from each other and the "electron cloud" reaches into the pion cloud. With machine learning methods like the Random Forest classifier a probability of being an electron in the 2D space of the two detectors can be established. In Figure 5.1 b) darker areas are assigned to a very high electron probability (black: P(e) = 1) and lighter areas to a very low electron probability (white: P(e)= 0). In the black areas all trees of the Random Forest vote for an electron, in white all trees vote for a pion. The Random Forest from Figure 5.1 was trained with 40000 pions and 40000 electrons, and tested with 2000 data points which are represented as blue and orange markers.

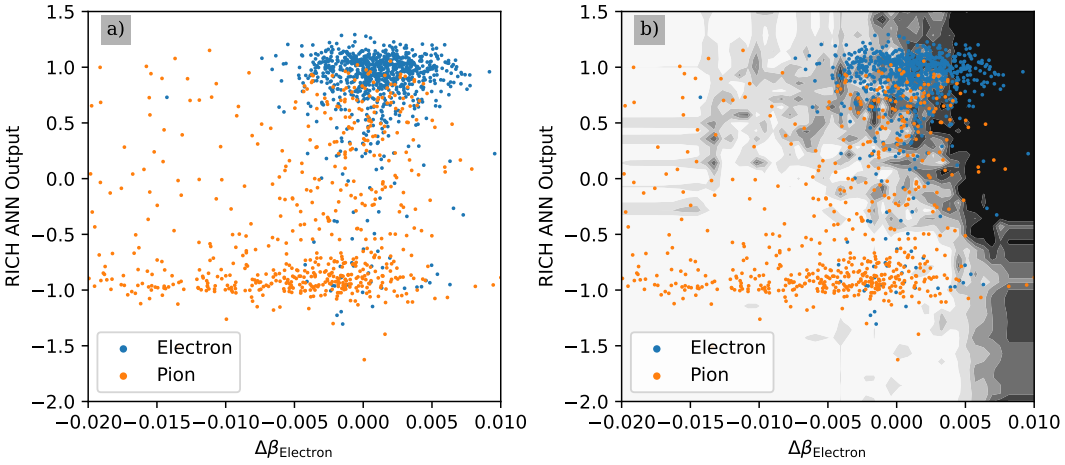


Figure 5.1: a) Simulated electron and pion data of two detectors assigned by a Random Forest classifier to their Electronness b). Black corresponds to a high probability of being an electron.

A comparison of the Figure 5.1 with the conventional methods shows that no square is cut out of the parameter space. Furthermore, also particles with large $\Delta\beta_{\text{Electron}}$ values are assigned to electrons. $\Delta\beta_{\text{Electron}}$ values larger than 0.00528 are not assigned to electrons but to pions in the conventional method. A first suggestion for the improvement of the conventional method would be not to use the $\Delta\beta_{\text{Electron}} > 0.00528$ cut. This can be considered in principle already from Figure 4.5, since there are no background particles lighter than electrons. For the pair analysis it must now be decided from which Random Forest output (gray value in Figure 5.1) the particles are classified as electrons. With the help of this threshold (cut) the Random Forest output which is between 0 and 1 can be mapped to a binary classification (electron or pion). Figure 5.2 a) shows the binary decision map for a high threshold. In contrast, Figure 5.2 shows the result for a lowhigh threshold.

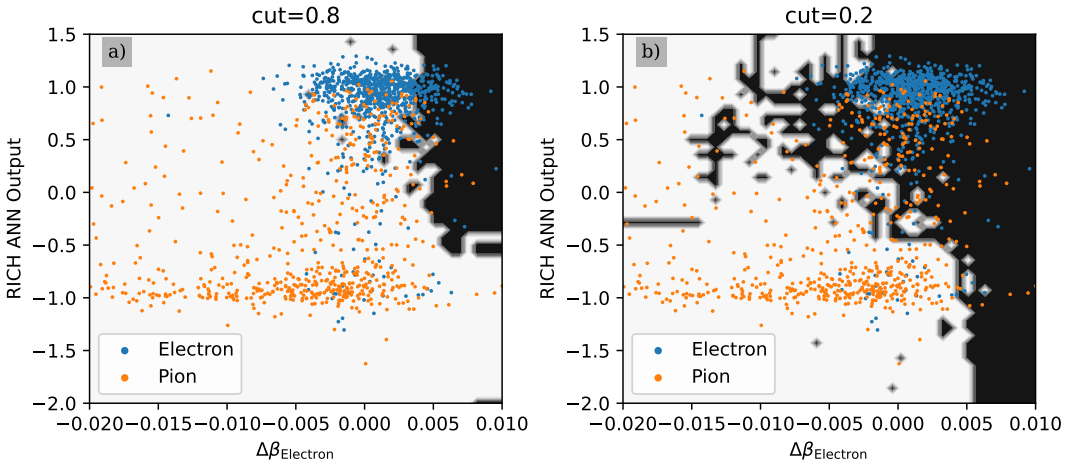


Figure 5.2: Binary classifications landscape for electrons (black) and pions (white). For higher threshold values (cut) of the classifier, a larger area is classified as electron.

It can be seen that for a low threshold value more electrons are classified as electrons, but also more pions are falsely classified as electrons. How well the classifier distinguishes electrons and pions can be investigated with the ROC which is shown in Figure 6.5 on the left. For each threshold (cut) the fraction of electrons (blue) which are in the black area (electron efficiency) and the fraction of pions which are wrongly in the black area (pion impurity) are calculated.

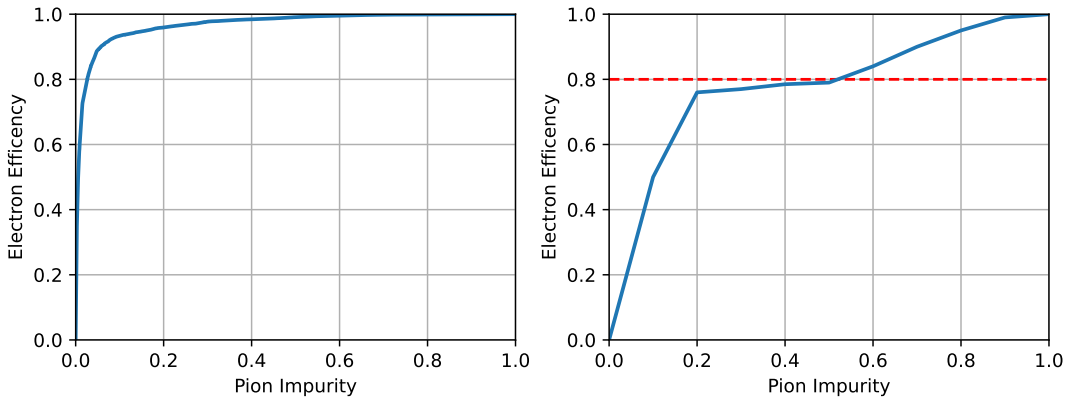


Figure 5.3: Left: ROC of the Random Forest classifier which distinguishes electrons and pions from each other using the parameters $\Delta\beta_{\text{Electron}}$ and RICH ANN output. Right: ROC of a hypothetical classifier which is supposed to distinguish electrons and pions from each other at an electron efficiency of 80% (Red Line).

Conventional PID methods require an electron efficiency of 80% for the TRD without knowing the ROC of the classifier. Each PID detector has its own ROC for each momentum which is specific to the detector. Assuming that the ROC of the TRD looks like the hypothetical ROC from Figure 6.5 on the right, a much better pion impurity could be achieved by accepting a minimally worse electron efficiency. Thus, it should be investigated to what extent the significance of the analysis changes when varying the required electron efficiency. The following considerations should lead to the calculation of an optimal compromise between electron efficiency and pion impurity. The significance of the measurement depends on the electron efficiency e and the pion impurity p. By maximizing the significance the optimal cut value can be found. The significance of a measurement is given by formula 5.1 where S is the number of signal points and B is the number of background points.

$$\text{Significance} = \frac{S}{\sqrt{S+B}} \tag{5.1}$$

By using a high cut value like the one in Figure 5.2 on the left, one gets a particularly good signal to background ratio, since nearly all pions are sorted out. However, a lot of signal is lost so that the measurement would have to run longer to get as much statistics as with a lower threshold. The significance calculates exactly this compromise. Figure 5.4 shows a simplified histogram of the pair analysis, where the signal (in blue) follows a Gaussian distribution and should be subtracted from large combinatorial background (orange). The distribution functions belonging to signal and background are shown in red. In the dielectron measurement signal and background can only be measured together as shown in Figure 5.4 b). By subtracting the red distribution curve of the background the histogram of the signal can be obtained (see Figure 5.4 c)). In the pair analysis, the distribution curve of the

background is obtained by pairing positive tracks with negative tracks from another event. In this way the tracks cannot come from the same parent particle.

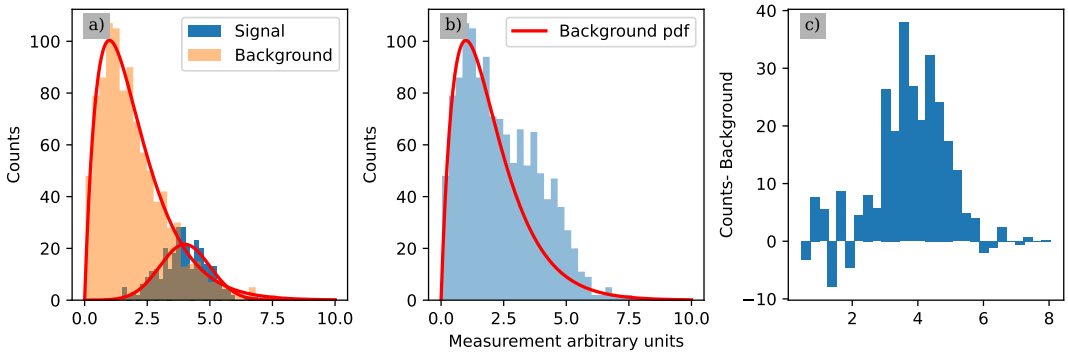


Figure 5.4: Illustration of the procedure for subtracting the background (orange) from the observed signal+background (light blue) to measure the signal (dark blue). The red curves show the distribution functions for signals and background. The distribution function of the signal is unknown and should be measured.

Note that the variance in the individual bins is significantly larger than that of the original signal from Figure 5.4 a) (blue histogram). The larger variance in the bins is explained by the additional inclusion of the variance of the background measurements. In order to evaluate the measurement from Figure 5.4 c) an **as small as possible variance of the mean in each bin** should be achieved. Figure 5.5 shows a histogram with uncertainties in the bins. The goal of the measurement is to capture the signal as accurately as possible, so the relative uncertainty $\Delta S/S$ should be minimized. It turns out that the inverse of this relative uncertainty $S/\Delta S$ is the significance of the measurement within a bin, which will be explained in the following.

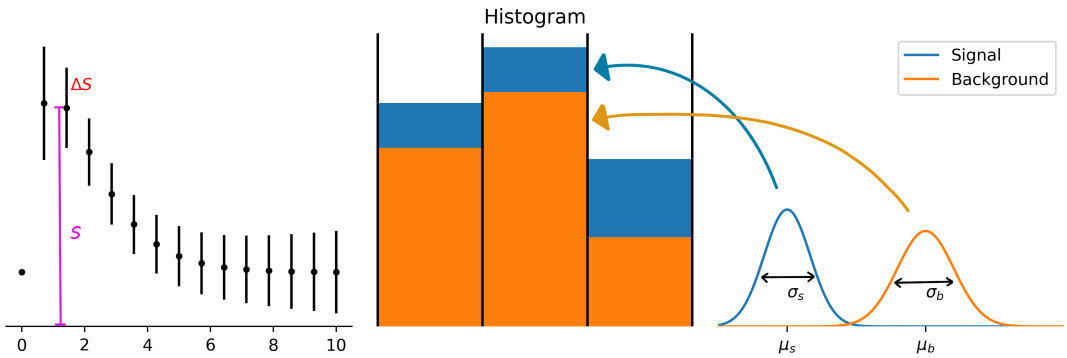


Figure 5.5: Left: Example of histogram measurement with uncertainties. The ratio ΔS to S represents the relative measurement uncertainty. Center: Zoom into the bins of the histogram, which are filled with signal and background measurement points. Right: Number of measurement points in the bin is Gaussian distributed for signal and background.

The number of signal measurements per bin is binomial distributed, with the expected

value $\mu_s = S \cdot p$ and the variance $\sigma_s^2 = S \cdot p_s(1-p_s)$. p_s is the probability of counting an entry in a bin and can be calculated as 1 by the area of the probability density function of the signal over the bin. S is the total number of signal measurement points. For a large number of measurement points, the binomial distribution approaches a Gaussian distribution. If the number of bin points is very small, the probability distribution becomes so small that $p(1-p) \approx p + \mathcal{O}(p^2)$. Similarly, for the distribution of the background, the expected value is $\mu_b = B \cdot p$ and the variance $\sigma_b^2 = B \cdot p_b(1-p_b)$. B is the number of total background measurements. Because of the large number of measurements taken in the pair analysis, the number of counts per bin for the signal s and the background b approaches a Gaussian distribution (see equation 5.2). Figure 5.5 illustrates this. The sum of normal distributions gives again a normal distribution in which the expectation values are added and the standard deviations are added (compare equation 5.3). In equation 5.4 the distribution of the histogram in Figure 5.4 c) is computed. The expected value for the background is subtracted from the signal + background. The variance of the bin content for a bin in the histogram from Figure 5.4 c) is obtained with the standard deviations σ_s^2 and σ_b^2 introduced above to equation 5.5.

$$s \sim \mathcal{N}(\mu_s, \sigma_s^2), \quad b \sim \mathcal{N}(\mu_b, \sigma_b^2) \tag{5.2}$$

$$s + b \sim \mathcal{N}(\mu_s + \mu_b, \sigma_s^2 + \sigma_b^2) \tag{5.3}$$

$$s + b - \mu_b \sim \mathcal{N}(\mu_s, \sigma_s^2 + \sigma_b^2) \tag{5.4}$$

$$\Delta S \hat{=} \mathrm{var}(s + b - \mu_b) = \sqrt{S \cdot p_s + B \cdot p_b} \tag{5.5}$$

$S \cdot p_s$ is the expected value of signal points per bin since p_s is the probability that a count will end up in a given bin. Likewise, $B \cdot p_b$ is the expected value for the background counts in a bin. The relative uncertainty of the signal measurement in a bin is thus calculated to:

$$\frac{S}{\Delta S} = \frac{S}{\sqrt{S \cdot p_s + B \cdot p_b}} \tag{5.6}$$

Actually, this relative uncertainty is a quantity that can be calculated for each bin. Since the information to be taken from the measurement is not based on a single bin, it is useful to define the overall significance as $S/\sqrt{S+B}$. This minimizes the relative uncertainties in the mean for all bins. The use of more exact mathematical models, which include e.g. the Boltzman fit function of the spectrum for the determination of the temperature, leads to a more exact but also more complicated significance equation. In the previous discussion on classification, electrons were treated as signals and pions as backgrounds. However, since pairs of electrons and positrons are considered as a signal, some considerations must first be made on how to calculate the S and $S+B$ quantities. In the pair analysis all positive tracks are combined with

all negative tracks (see Figure 5.6). Here only e^+e^- pairs can form as signal pairs (in green). Combinations that contain a pion always form background (in red).

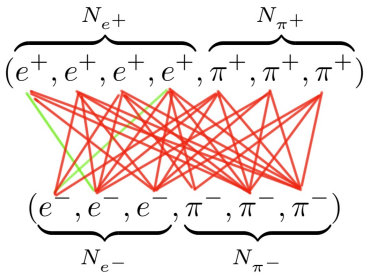


Figure 5.6: Illustration of the combination of positive and negative tracks for pair analysis. Signal pairs are connected with green lines. Background pairs are connected with red lines. The symbols above the curly brackets indicate the number of signal particles (electrons) and background particles in the positive and negative tracks.

If an electron efficiency of 50% is given, 50% of the electron tracks and 50% of the positron tracks are eliminated. But in total 75% of the signal is lost, because both, the electron and the positron of the pair, must be in the not removed group. This can be compared to a row of mixed shoes where half of the shoes are discarded recklessly. The probability that someone will still be able to wear their favorite pair of yellow shoes is then $50\% \cdot 50\% = 25\%$ and not 50%. What is the point of one shoe if the second one is missing?

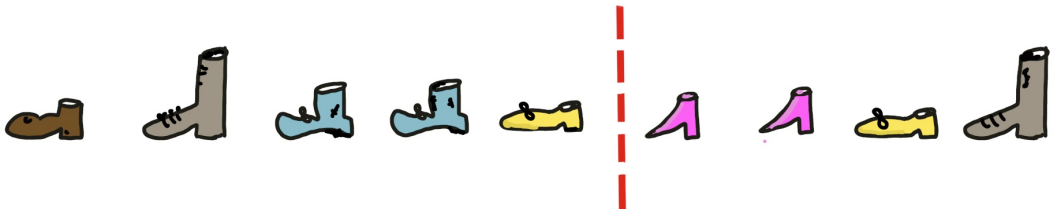


Figure 5.7: Drawing of a few shoes to illustrate the significance calculation, shoes after the red line are removed from the randomly shuffled sample.

The number of signal measurements is therefore proportional to the electron efficiency squared, which is a crucial relation in the course of the thesis. More precisely, the signal is calculated as shown in equation 5.7, assuming that the total number of electrons and positrons detected is about equal.

$$\text{Signal} = N_{e^+} \cdot e \cdot N_{e^-} \cdot e \approx N_e^2 \cdot e^2 \tag{5.7}$$

The number of background + signal is exactly the number of all combinations of positive tracks. So the sum of the red and the green combinations in the Figure 5.6.

If p is the pion impurity then $N_\pi \cdot p$ pions remain after classification. Equation 5.8 calculates the sum of signal and background, again assuming that the number of π^+ is approximately equal to the number of π^-.

$$\text{Background} + \text{Signal} = (N_{\pi^+} \cdot p + N_{e^+} \cdot e) \cdot (N_{\pi^-} \cdot p + N_{e^-} \cdot e) \approx (N_\pi \cdot p + N_e \cdot e)^2 \tag{5.8}$$

Overall, the significance of the pair analysis depends on the electron efficiency e and the pion impurity p of the equation 5.9.

$$\text{Significance}(e, p) \approx \frac{N_e^2 \cdot e^2}{N_\pi \cdot p + N_e \cdot e} \propto \frac{e^2}{\frac{N_\pi}{N_e} \cdot p + e} \tag{5.9}$$

Since the pion impurity depends on the electron efficiency via the ROC of a classifier, the function can be maximized and thus the optimal electron efficiency can be estimated. The ratio of electrons and pions can be obtained from the simulations. Figure 5.8 on the left shows the ratios of the dominant detected tracks arising in the collision. The ratio of pions to electrons is 22.19 for a UrQMD simulation of AuAu 12 AGeV collisions. When in the course of the thesis the significance is mentioned, it refers to the actual significance divided by the number of electrons N_e. This normalization has the advantage that an optimal classification i.e. 100% electron efficiency 0% pion impurity has the value 1.

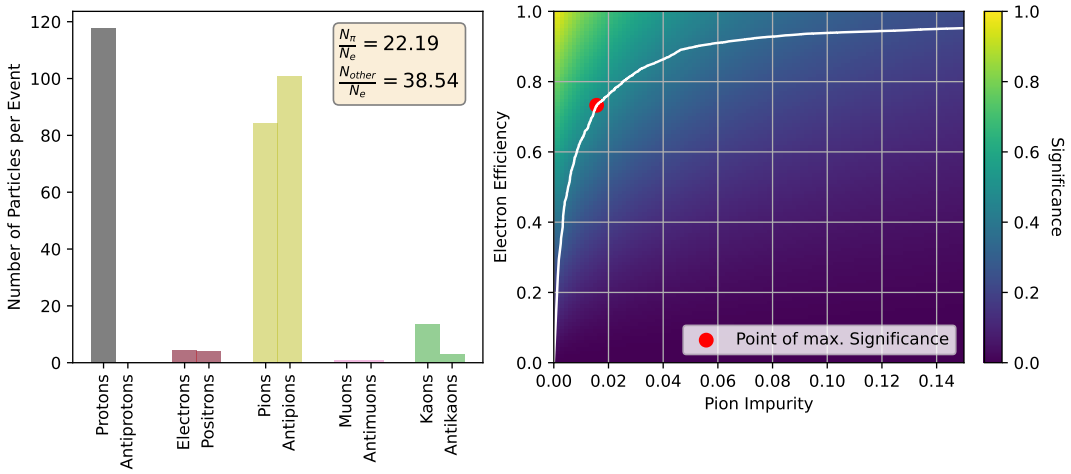


Figure 5.8: Left: Histogram of the number of different particle tracks created in the collision (UrQMD simulation). The ratio of pions to electrons as well as pions, protons, muons and kaons (N_{other}) to electrons (N_e) is shown in the upper right. Right : ROC of a Random Forest method to distinguish between pions and electrons. The color code shows the significance for the dielectron analysis resulting from equation 5.9. The point of maximum significance along the ROC is shown in red.

Figure 5.8 right shows the ROC from Figure 6.5 left in the interval of pion impurities from 0-15%. The color code shows the significance normalized to 1. The red marker

indicates the point of maximum significance along the ROC. In the example in Figure 5.8, an electron efficiency of 75% would maximize the significance of the pair analysis. This is not a relevant result but only serves to illustrate the procedure for determining the optimal electron efficiency of a classifier. In the course of the thesis not only methods for the distinction of electrons and pions are investigated but also methods for the distinction of electrons from the background particles which contain besides pions for example also protons and kaons. For this purpose, equation 5.9 for the significance calculation can be derived completely analogously. The difference is that the ratio of pions to electrons changes into the ratio of background particles (N_{other}) to electrons (N_e). For the simulated 12 AGeV AuAu collisions used in this thesis, the ratio is 38.54.

6 ML Applied to each Detector Separately

In this chapter, classification methods are presented that improve particle identification by the RICH detector. Since the PID methods for the RICH detector already use a neural network (see chapter 4), it is especially illustrative to compare its results with more modern machine learning methods. The neural network used for the PID was already implemented in the software of the experiment in 2008, which is why there is now a great potential for improvement due to the new knowledge in the field of machine learning. In the second part of the chapter, machine learning methods are applied to the TRD data and tested to what extent machine learning methods can keep up with the likelihood method.

6.1 Classification for RICH Data

The ANN implemented in CBMROOT has an architecture of 9 input neurons, a hidden layer with 18 neurons and 1 output neuron. As input to the neural network 9 parameters were used, which also serve as the basis for the classification methods used in this thesis. Of the 9 variables, 8 come from the information of the detected photon ellipse. The remaining input variable is the reconstructed momentum p of the particle. After the photon rings generated by the Cherenkov light are projected onto the photon plane, ellipses are formed on a two-dimensional pixel surface. Each of these ellipses can be uniquely described by 5 parameters, which are shown again in Figure 6.1.

In Figure 6.1 the orange dots represent the actual photon hits. Their number of hits (N_{hits}) assigned to an ellipse is one of the most important measurements of the PID of the RICH, since electrons generate on average more hits. The detected photons are fitted by an ellipse (in red) with the following 5 free parameters: The major semi-axis A and minor semi-axis B (in blue), the position of the center in space in polar coordinates (radial position r and the radial angle Θ), the orientation of the major axis with respect to the x-axis angle ϕ and the χ^2/NDF value of the fit. The χ^2/NDF value is given by

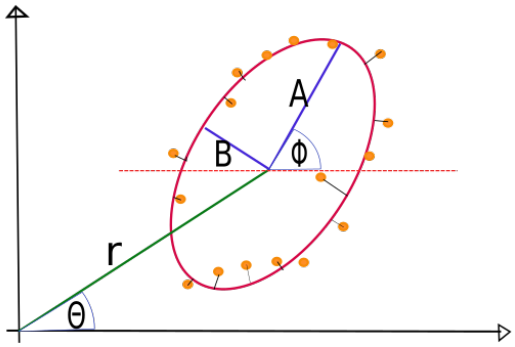


Figure 6.1: Illustration of a photon ellipse detected by the RICH detector and its fit.

the squared distance of the hits to the ellipse and the NDF value, which is given by $N_{\text{hits}} - 5$. The 5 are the number of free parameters of the fit. The χ^2/NDF value describes how well the ellipse fits the actual photon hits. Furthermore, each detected ellipse can be assigned another variable "Distance to Track" that cannot be read from the Figure 6.1. From the information of the ellipse the origin of the photon ring can be determined by back projection. This three-dimensional information can then be mapped to a track. The distance of the detected origin of the Cherenkov radiation to the track is called "Distance to Track" in the following and is an important variable for the distinction of pions and electrons. Figure 6.2 shows histograms of the 8 detector variables for electrons and pions.

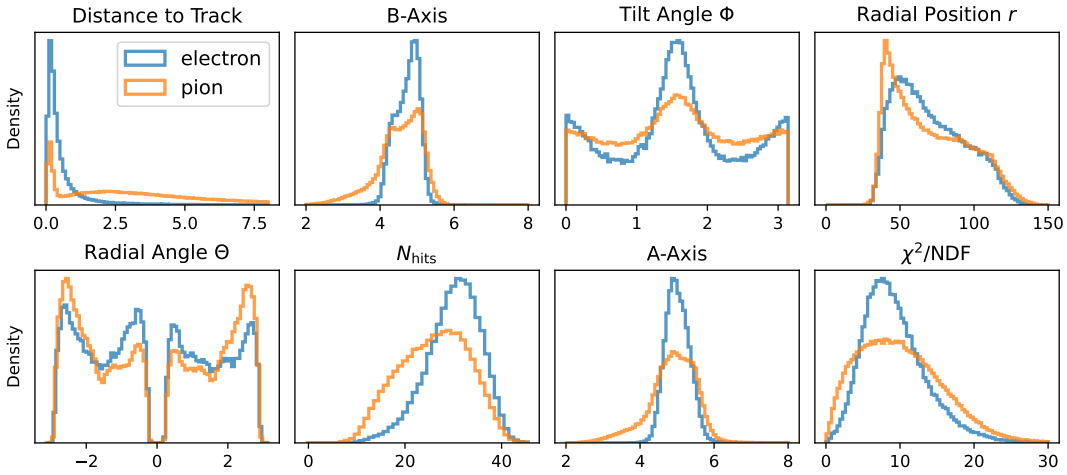


Figure 6.2: Histograms of the 8 effective measurements of the RICH detector, for electrons and pions. Different distributions for pions and electrons lead to a better distinction of the particles on the basis of the variables.

Figure 6.2 how electrons and pions differ in their distributions of feature values. The information on the distribution of the major semi-axis A and the minor semi-axis B shows that pion ellipses are on the average smaller than those of electrons. From the measurement of the tilt angle ϕ it is evident that the ellipses of the electrons have a more vertical or horizontal orientation than those of the pions, which is due to the detector geometry. Since pions are heavier than electrons, they are deflected more by the magnetic field at the same momentum, so that the detector center is dominated by electron measurements and the edges by pion measurements. For these reasons, the measurements of radial position and angle can contribute to the PID. In Figure 6.2 N_{hits} it can be seen that pions tend to have fewer photon hits than electrons. The histogram of the track distance show that pions tend to have a larger distance to the particle track than electrons.
To further examine the input data to the neural network, the correlation matrix was investigated. Since the data is in a higher 9-dimensional space and therefore mostly the projections of the data on the respective axes are considered (Figure 6.2), a correlation matrix is suitable to observe the relationships of the data.

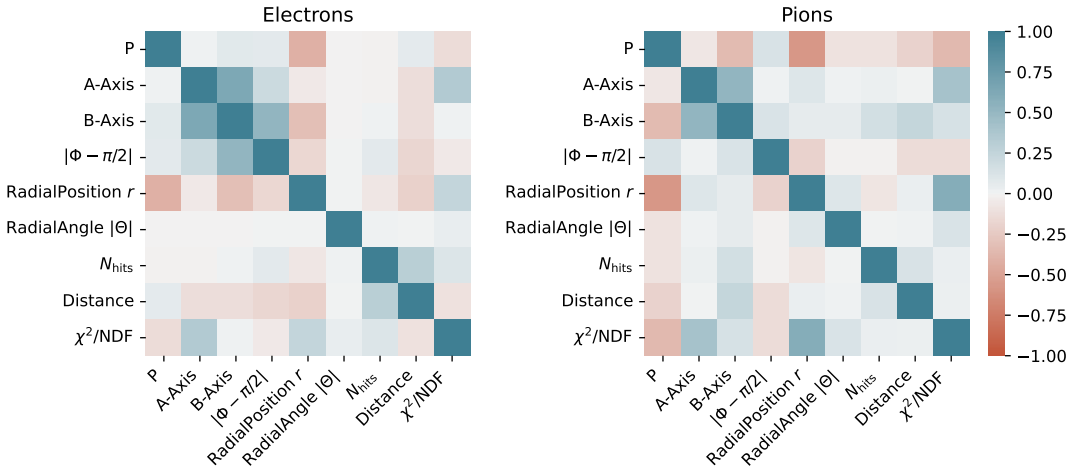


Figure 6.3: Correlations matrices for the input data in the classification methods, for electrons and pions separately. Blue values belong to a positive correlation negative to a red correlation.

From Figure 6.3 an idea of the distribution of the measured values in the 9-dimensional input space can be obtained. The entry in the matrix field for the correlation of B-axis and A-axis is blue for both electrons and pions, which belongs to a positive correlation. This means that ellipses which have a larger semi-axis A also tend to have a larger semi-axis B, so that the ellipses tend to resemble circles. The information about the major semi-axis A therefore contains a lot of information about the semi-axis B of the ellipse. In Figure 6.2 it can be seen that the radial angle Θ is exactly the same for negative angles up to -180° and for positive angles up to +180°. The symmetry of the data in a linear correlation matrix would give the value 0, because the negative correlation would balance the positive. Therefore it is suitable to examine the absolute value of the angle. The same is true for the inclination of the ellipse θ, where first 90° are subtracted and then the absolute value is taken. From the correlation matrix for electrons it can be observed that the radial angle does not correlate with the other data. This is not the case for pions, however, which may be due to the fact that electrons tend to fly through the center of the detector, while pions are deflected stronger by the magnetic field. Thus, the reason of this discrepancy lies in the detector geometry. If a variable (e.g. θ in this case) in both the electron data and the pion data did not correlate with the rest of the data, this feature could be taken out of the classification method and the total probability calculates as follows:

$$P(p,\ldots,\theta,\ldots,\frac{\chi^2}{\mathrm{NDF}}) = \underbrace{P(p,\ldots,\frac{\chi^2}{\mathrm{NDF}})}_{\text{classifier output}} \cdot \underbrace{P(\theta)}_{\text{Likelihood}} \tag{6.1}$$

The reduction of a feature can increase the speed of the evaluation of the classifier as well as provide a more transparent classification. This reduction cannot be done

because of the correlation of all features for the pions in the correlation matrix (see Figure 6.3).

Using the data for 25000 electrons and 25000 pions, different classifiers were trained and their hyperparameters optimized. An ANN, a Nearest Neighbors classifier, Gaussian NB, MLP, QDA, XGBoost and Random Forest were tested. In the chapter introducing the different classifiers, the classifiers that best classify the test data (another 25000 electrons and 25000 pions) were discussed in detail. Figure 6.4 illustrates the classifiers by their ROC. The yellow curve of the Random Forest lies exactly behind the green curve of the XGBoost classifier. These two classifiers perform best and are examined in more detail in Figure 6.5. Figure 6.4 shows the expected result that the Random Forest and XGBoost classifier outperform the other methods, which is why other methods like QDA will not be investigated in detail in the course of the thesis. It is widely known that Random Forest and XGBoost usually outperform the other methods, which is why these methods are the most common particle classification methods in high energy physics, based on different measurement variables.

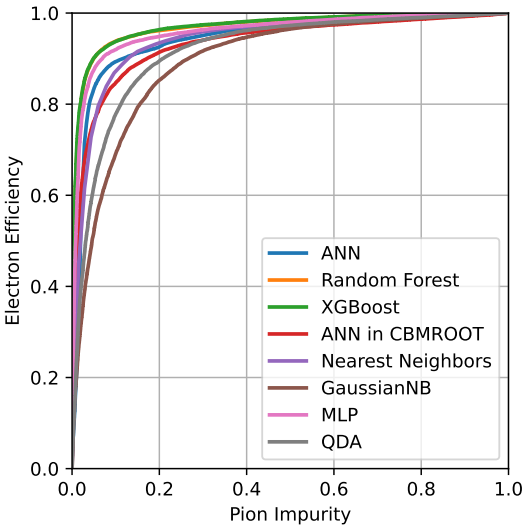


Figure 6.4: Comparison of the ROC for different classifiers that distinguish between pions and electrons based on the RICH data. The orange ROC of the Random Forest classifier is exactly below the ROC of the XGBoost.

In Figure 6.5 it can be seen that the new classification methods beat the ANN implemented in CBMROOT. This means that for the same electron efficiency, fewer pions are misclassified as electrons. The result with the highest significance is obtained by the XGBoost classifier with 903 estimators, a maximum depth of 3 and a learning rate of 0.074. In addition to the XGBoost and the Random Forest, an ANN with a similar architecture to the ANN implemented in CBMROOT was compared to this one. This ANN performs similarly well to the ANN trained in 2008. This shows that the possibilities of classification with neural networks are limited. Theoretically, neural networks could also be useful for the classification of RICH data, since the raw data of the detector are the photon hits on a 2D pixel plane. These data are images, which can also be evaluated by neural networks. The result of these networks could have as output the input parameters for the PID presented here. Accordingly, a network could also be trained, which has as input the 2D image with the ellipses and as output a probability for the classification of the particle as well as the origin of the photon ring. The development of such a combined network could be a point of later research. Compared to the classifier implemented in CBMROOT, XGBoost and Random Forest have the further advantage that the output of the methods is between 0 and 1 so that they can also be interpreted as a probability. The ANN implemented

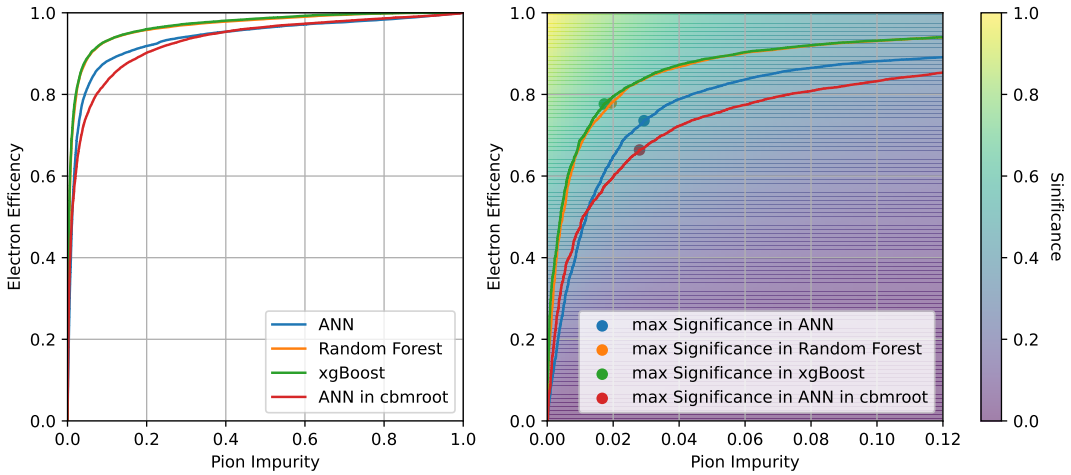


Figure 6.5: ROC for the three new methods Random Forest Classifier, Artificial Neural Network (ANN) and XGBoost, which distinguish electrons and pions based on RICH data. The ANN implemented in CBMROOT in 2008 (in red) is shown for comparison. The ROC of the Random Forest classifier is exactly below the ROC of the XGBoost classifier. Right: Zoom of the ROCs for small pion impurities. The color code shows the normalized significance of the pair analysis. The markers show the points of maximum significance along the ROC.

in CBMROOT generally maps the input data to the real numbers. In more modern methods, a classifiable neural network is assigned a "softmax" function as the last layer, which converts the neural network outputs into probabilities for the respective classes.

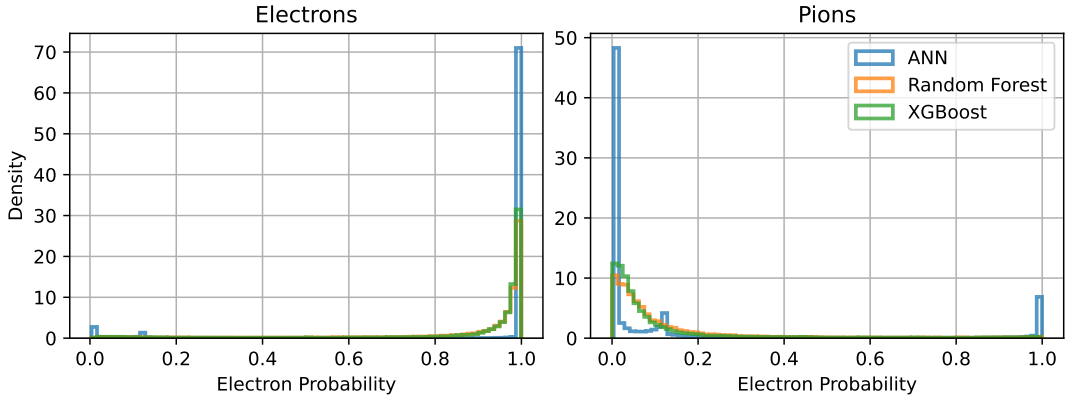


Figure 6.6: Output of the classifier (ANN), Random Forest and XGBoost for electrons (left) and pions (right). Output 1 represents the classification of the particle as an electron. The output 0 stands for the classification as a pion.

In Figure 6.6 the output of the different modern classification methods can be compared. Figure 6.6 a) shows the output of the classifier for electrons and Figure 6.6 b) that for pions. The output value 1 belongs to the classification of a particle as an electron and the output value 0 to the classification as a pion. In Figure 6.6 on the left

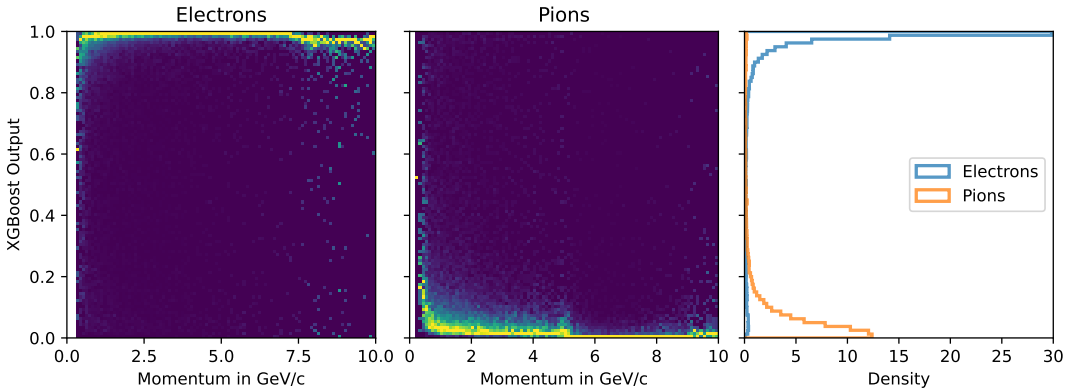


Figure 6.7: Electron identification value assigned to electron and pion tracks by a XGBoost Classifier as a function of the momentum. The assigned rings for both distributions require at least 6 hits on the photon plane. Both distributions are normalized to unity for each momentum bin individually

it can be seen that for the electrons, as expected, the maximum for the three methods is close to 1. In Figure 6.6 on the right, the expected case occurs that the maxima of the histograms are close to zero, which belongs to the classification as pion. The neural network tends to output outputs close to 0 or 1 and thus does not resemble a typical probability distribution like that of the random forest classifier. The reason for this behavior is the architecture of the neural network and in particular the "softmax" function in the last layer. In the chapter presenting the conventional PID methods, the output of the conventional ANN for different momenta was investigated, and it was found that pions for high momenta are assigned to electrons. Figure 6.7 shows the same plot as Figure 4.2 but this time for the XGBoost classifier, which performed best in the test. It can be observed that the XGBoost classifier is able to distinguish the particles also for higher momenta, so that pions are much less likely assigned to electrons as in Figure 4.2 for high momenta. Furthermore, in Figure 6.7 the momenta integrated output of the classifier can be seen, which in comparison resembles clearly more a probability function.

So far, the classifiers have been trained and tested only on the data for pions and electrons, but not for the discrimination of electrons from the whole background of particles, which besides pions also includes protons, kaons, muons and a small fraction of other particles. Since the main part of the background particles are hadrons, the impurity of particles which are not electrons is called hadron impurity in the course of this thesis. The classification of electrons and pions is idealized and should be replaced by the classification of electrons and hadrons. Of course, one difficulty is that the percentages of the different components of the hadron cocktail are calculated by simulations (see Figure 5.8 on the left). Figure 6.8 shows the ROCs for classification methods ANN, Random Forest and XGBoost again in comparison with the conventional ANN. This time, however, 25000 electrons and 25000 hadrons were used to train and test the modern methods. The conventional ANN was tested on the same 25000 electrons and 25000 hadrons.

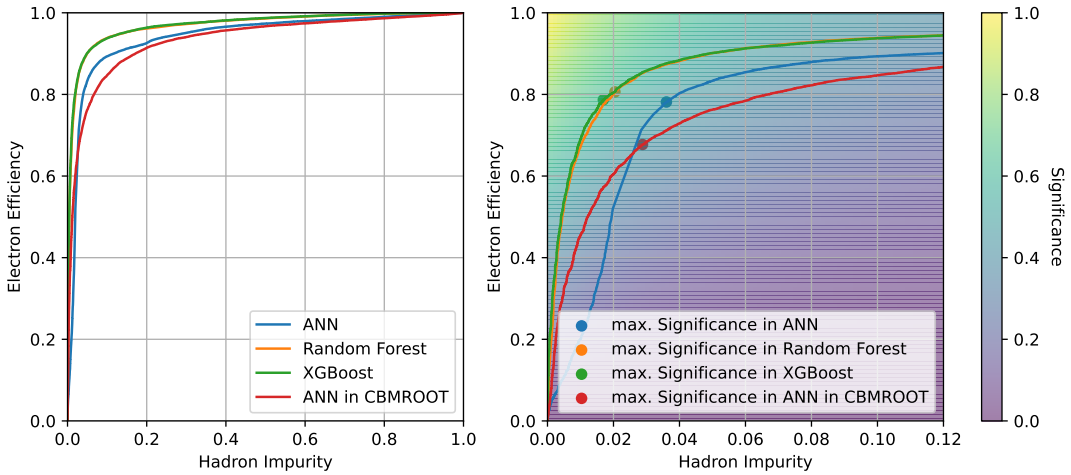


Figure 6.8: ROCs of the classification methods ANN, Random Forest and XGBoost as well as the ANN implemented in CBM for the classification of two electrons and hadrons based on the RICH data. The conventional ANN is trained only to distinguish between pions and electrons. Right: Zoom of the ROCs for smaller hadron impurities. The color-code shows the normalized significance for the pair analysis.

The ROCs in Figure 6.8 show similar results as for testing on electrons and pions. Again, XGBoost performs as well as the Random Forest classifier. Both classifiers manage to obtain 20% more electron efficiency at a hadron impurity of 2% than the conventional ANN (see Figure 6.8 on the left). Another important remark is that the calculation of the optimal cut for significance maximization, which results for the Random Forest to 80% electron efficiency, is calculated here only for the case that the RICH detector is the only classifier. The significance calculation depends on the results of the other detectors. If for the PID in the experiment the conventional PID methods are used and the neural network of the RICH is updated, an optimal cut value must be computed. Furthermore the significance equation 5.9 has to be replaced by the new electron hadron ratio after the TRD and TOF PID cuts were applied.

Note: The Machine Learning algorithm presented in this thesis has been improved by optimizing the hyperparameters of the method using a grid-search. This optimization is done with python packages as they are for example implied in hipe4ml [24].

6.2 Classification for TRD Data

The perfect classification method is to know the probability distribution for electrons and pions in the whole n-dimensional classification space. The n-dimensional variable space of the TRD is five-dimensional and includes the energy entries in each of the four TRD layers and the number of TRD hits. Usually, it is not possible to calculate a probability distribution for many dimensions because there is often too little data. However, in the introduction to the likelihood method it was shown that the energy depositions in the different TRD layers are uncorrelated (see Figure 4.3), so it is possible to define a probability at each point of the input space of the TRD classifier. Figure 6.9 shows the input datasets of the TRD into the classification method. The four energy entries of the particles in the respective layers are shown for electrons and pions as well as a comparison of the number of hits for electrons and pions.

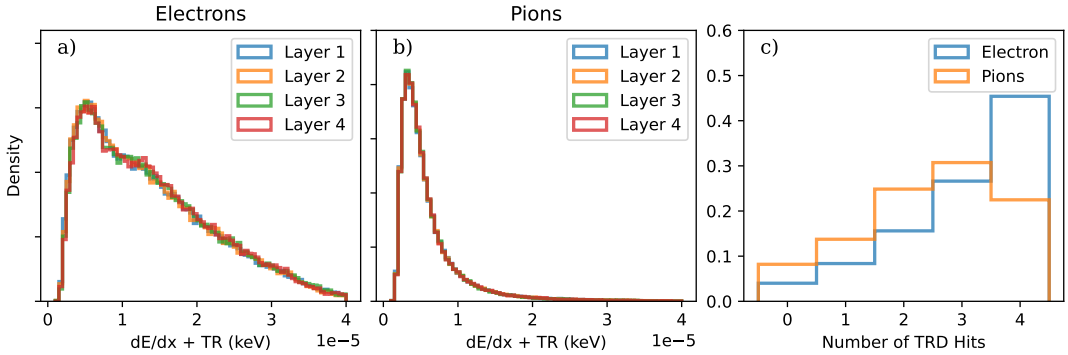


Figure 6.9: a) shows histograms of the energy deposition of electrons at the TRD in the different layers. b) shows histograms of the energy deposition of pions. The histograms in a) and b) differ, but the histograms of the different layers do not. c) shows histograms of the number of TRD hits for electrons and pions. Pions tend to have fewer hits.

In Figure 6.9 a) it can be seen that the electrons in the respective TRD layers have the same energy distribution. The same is true for pions (see Figure 6.9 b). From this it follows that the input space has a symmetry for the four TRD layers. In Figure 6.9 c) it can be seen that pions tend to hit less TRD layers. The reason for this is that pions are more strongly deflected by the magnetic field due to their higher mass and therefore have a higher probability to hit fewer TRD layers. Electrons, on the other hand, are less deflected and move through the center of the detectors, so they are more likely to be seen by all four TRD layers. Therefore, for a fair comparison of the likelihood method with the machine learning methods, only the energy entries of the four detector layers and not the number of hits should be input for the classifier. Furthermore, only particle tracks hitting all four TRD layers should be tested. Otherwise, the method could identify the zeros at the no hit locations and use the geometric information to distinguish between pions and electrons.

Figure 6.10 shows the ROCs of the three classifiers Likelihood method, Random Forest and XGBoost. 25000 electrons and 25000 pions were used for the training. The methods were also tested on 25000 electrons and 25000 pions, different to those of the training.

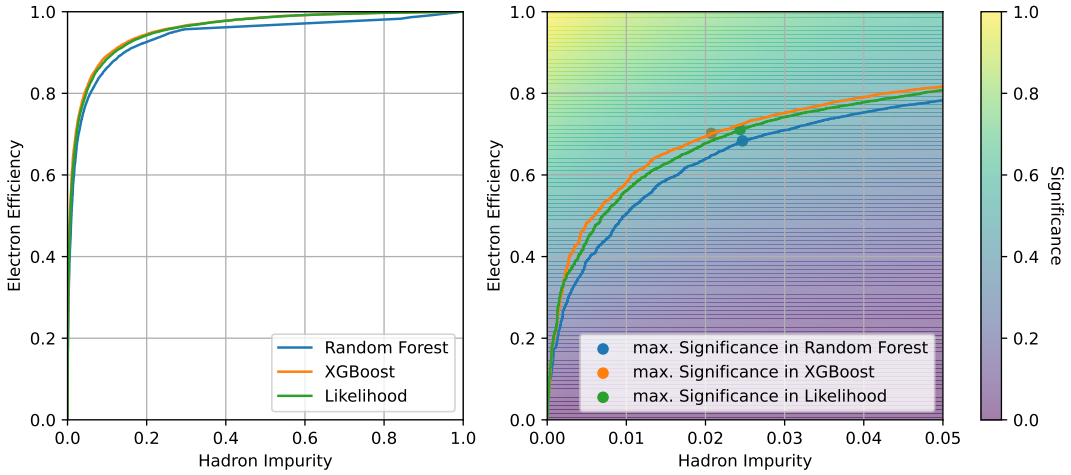


Figure 6.10: Left: ROCs of the three classifiers likelihood method, Random Forest and XGBoost to distinguish pions and electrons with the TRD. Input variables are the four energy depositions in the detector layers. All test tracks hit all four layers of the TRD. Right: Zoom of the ROCs for small pion impurities up to 5%. The color-code shows the normalized significance for the pair analysis.

In Figure 6.10 it can be seen that the three classifiers can all distinguish pions and electors similarly well. Notably, the XGBoost classifier minimally outperforms the likelihood method as can be seen in Figure 6.10 on the right. This can be explained by the fact that the single detector layers correlate minimally, so that the XGBoost achieves a 1% better electron efficiency at a hadron impurity of 2%. The Random Forest classifier performs worst in the comparison. How close the machine learning methods operate to the ROC of the likelihood method shows the quality of the classification algorithms. To compare whether the outputs of the classifier are similar to the actual likelihood function, the outputs of the Random Forest and the XGBoost are compared with the output of the likelihood method in Figure 6.11. The output of the likelihood method is by definition, if the layers of the TRD do not correlate, equal to the probability of detecting an electron.

In Figure 6.11 it can be seen that the output of the XGBoost classifier for electrons most closely resembles the idealized probability function of the likelihood method. This means that the output of the XGBoost classifier can be interpreted as a kind of "probability". For pions, the Random Forest classifier does not assign a value of 0 to pions, but 0.2, which is for unknown reasons and could be improved. Furthermore, the XGBoost classifier tends to assign larger values to pions than the likelihood method. Overall, the likelihood method can distinguish electrons from pions based on the simulated data minimally worse than the XGBoost, but it is more transparent and

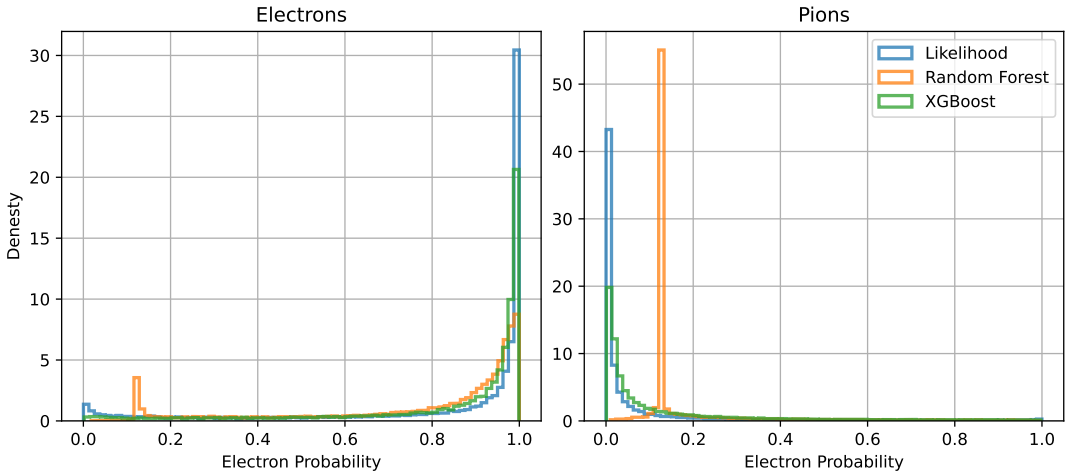


Figure 6.11: a) shows the output of the methods for electrons. The output 1 belongs to the classification of the particle as an electron, the output 0 to the classification as a pion. b) shows the outputs of the different methods for pions. The likelihood method assigns the values 0 and 1 to the electrons and pions most sharply.

less of a black box than the XGBoost classifier. Furthermore, it has a much faster computation time and depends less on simulated data and more on well understood concepts. It is therefore not worthwhile to replace the likelihood method in the conventional methods by a machine learning method. It was shown how well the XGBoost classifier works and that its output can also be interpreted as probability.

7 ML for all PID Detectors

In this chapter, classification methods are presented, which distinguish electrons from hadrons by using all PID detector data of the CBM detector. The 19 input variables have already been presented in detail in the course of the thesis and are summarized in the grey box. It is possible to use the 4 energy depositions in the detector layers of the TRD as one likelihood variable. This reduces the size of the input vector by three, which could result in a faster and more accurate calculation, and will be investigated later.

Global Parameters : $\underbrace{P, \Phi, \Theta}_{\text{Momentum}}$, χ^2 to Vertex ,Charge

---

TRD Parameters : $\underbrace{E_1, E_2, E_3, E_4}_{\text{Likelihood}}$, No. of Hits

---

RICH Parameters : No. of Hits , A-Axis , B-Axis , Radial Position , Radial Angle, χ^2/NDF, Tilt Angle Φ, Track Distance

---

TOF Parameters : $\Delta\beta_{\text{Electron}}$

- P (global variable): momentum of the particle in MeV/c
- Φ (global variable): azimuth reconstructed momentum
- Θ (global variable): polar angle of the reconstructed momentum
- χ^2 to Vertex (global variable): square distance to variance of the collision point to the reconstructed track
- Charge (global variable): charge of the particle value, +1 or -1
- E_1- E_4 : energy deposition in the individual detector layers of the TRD
- No. of Hits (TRD parameters): number of TRD hits from 1 to 4
- No. of Hits. (RICH parameters): number of hits for the reconstruction of the ellipse
- A-Axis: major half axis of the ellipse
- B-Axis: minor half axis of the ellipse
- Radial Position & Radial Angle: center of the ellipse in spherical coordinates
- Tilt Angle Φ: rotation of the ellipse in the 2D plane around the angle Φ
- Track Distance: distance of the fitted track to the center of the ellipse.
- χ^2/NDF: quality of the fit of the ellipse.
- $\Delta\beta_{\text{Electron}}$: Difference of beta value measured by TOF to the expected electron beta value for electrons.

7.1 Missing Detector Data

A peculiarity of the input data is that they have "Not a Number (NaN)" entries in case of missing detector data. For example, if the track does not hit the TRD, this does not mean that the energy depositions in the detector are all 0. This would cause the track to be classified as a pion. If the track does not hit the TOF and therefore the $\Delta\beta_{\text{Electron}}$ value is set to 0, the particle would always be classified as an electron. Figure 7.1 illustrates that the data with missing information is sent to a subspace of the data space. In Figure 7.1, the detector information TRD likelihood, RICH ANN and $\Delta\beta_{\text{Electron}}$ from TOF is clustered in different corners of the 3D space. Assuming that the track has no RICH hit, all this data is projected onto one plane (see Figure 7.1 right). The problem now is that the projected values from Figure 7.1 left overlap with the non-projected Figure 7.1 right. Therefore, special care must be taken, to assure that zero is suitable for projection. A projection to 1 would wrongly classify the particles as electrons and a projection to -1 would classify them as pions. So the output of the classification method depends strongly on which real number the NaN values are set to.

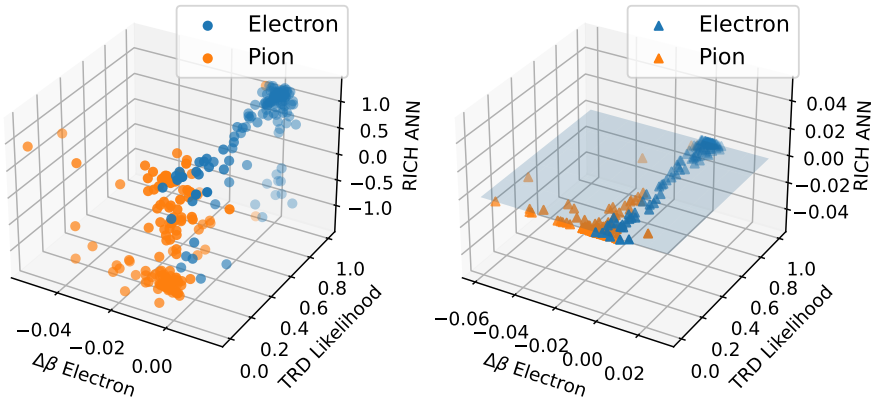


Figure 7.1: a) Figure of the equal data sets for the three detectors TRD, RICH, and TOF in three-dimensional space. b) The RICH ANN values of the data were projected to 0. Note that the pions are located in the front cluster and the electrons are located in the back at the top. If the RICH information is removed, these data are in a plane that is here set to 0. In the CBM data, the points from the right and left image overlap.

If all the data are to be classified by a single classifier as electron or pion (or hadron) it is therefore suitable to set the missing detector data to values like -99 which are clearly distinguishable from the usual measurement data. Another way to avoid the problem of missing detector data is to divide the data into different classes. For example, if there is no TOF hit then the $\Delta\beta_{\text{Electron}}$ value can be taken from the input vector so that it has one dimension fewer. As a result there is no NaN entry in

the vector anymore. All these classes without NaN values must then get their own classifier, which was trained exactly on this class. In order to avoid having too many different classes, it is suitable to use the likelihood instead of the energy depositions in the detector layers. Then it is not necessary to train and use different classifiers for 1, 2, 3, or 4 hits in the TRD. When using the likelihood method, the number of NaN entries in the vector is reduced since 1 hit and thus 3 NaN entries in the TRD are also mapped to a likelihood. Different classification methods perform differently on data sets with missing information. For example, the XGBoost classifier is known for accurately predicting data despite missing information.

7.2 AddROC Method

To avoid the problem of missing data in the training and test datasets, the data can be split into different classes, depending on which detector data is missing for any given track. Different classifiers can then be trained specifically for each data class. The splitting of the data into the different classes and the subsequent training of the classification methods is not particularly difficult. Non-trivial is only the combination of the classifiers to a complete classifier and the related analysis of the ROC of the complete classifier. Since the ROCs of the individual subclassifiers are added for this purpose, this method is called **AddROC**. AddROC was developed in the course of this thesis specifically for this problem. The method is illustrated in Figure 7.2 and explained step by step. The subfigures shown in Figure 7.2 are only examples and not relevant results.

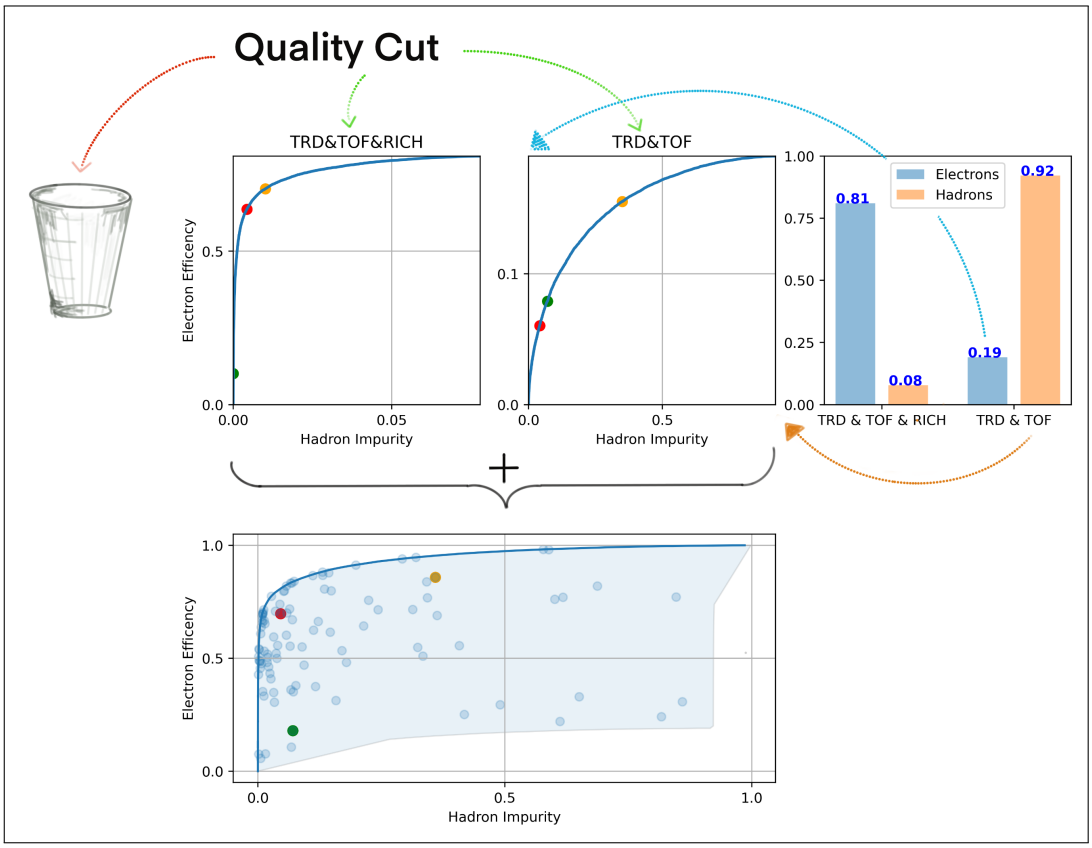


Figure 7.2: Illustration of the AddROC method which was developed during the thesis for combining different classifiers for subcalsses of data.

1. **Quality cuts** are applied to the data. The quality cuts used in this thesis are that the detector has at least three hits in the tracking detectors. In addition, χ^2 to vertex must not be greater than 3. Furthermore, the track must be seen by at least one particle identification detector. If not, the track cannot be evaluated for the particle determination and cannot be identified. If the condition is met,

further steps follow.

2. All **possible classes** are **determined** in a way that no NaN entries occur in the training and test data. For example, the tracks seen by TRD & TOF form one class, the tracks seen by all PID detectors form a second class, etc. In Figure 7.2 not all possible classes are shown. Theoretically, there are 10 possibilities to combine the information of the detectors. 3x the detectors individually, 6x two detectors combined, and 1x the combination of all detectors. For this thesis only 5 classes are of importance.

3. The **ratio of the data in each class** is calculated for the target tracks (electron) and background tracks (hadron). (See example histogram in Figure 7.2)

4. For each class, classifiers are trained and tested. The classifier types can differ from each other, for example, the class TRD & RICH & TOF uses Random Forest and the other classes use XGBoost. After testing the best classifiers, the ROC is calculated based on the test data.

5. The electron efficiency and the hadron impurity of the ROC of the sub-classifier must be rescaled. If the sub-classifier designates all particles as target particles (electrons), it can reach at most an electron efficiency of the percentage in which electrons occur in its class. In the same way, only the hadron impurity can be reached in which hadrons occur in the class of the classifier.

6. The ROCs must be added together to obtain an ROC for the entire method. To do this, arbitrary points on the respective ROC of the classes are selected and their electron efficiency and hadron impurity is summed. In Figure 7.2, for example the values of the green points are added and the values of the violet points are added. The new point can then be plotted in a diagram where the axes indicate the hadron impurity up to 100% and the electron efficiency up to 100%. All points then end up in the sky-blue colored area, which is also called alpha shape. The upper boundary of the calculated points (convex hull of the data) is the new ROC.

7. Each point of the new ROC can be mapped to two cut values of the single classifier in the example of Figure 7.2. Only the combination of these cut values is suitable. For example, the red point in Figure 7.2 lies within the sky blue area and thus does not separate as well as the values on the convex hull.

The method of selecting random points on all ROCs and then adding the electron efficiency and hadron impurities is not the most efficient one. This is especially not the case if there are many sub-classes and therefore many ROCs, since the number of points exponentially increases with each ROC. A more efficient method is to take the first two ROCs, calculate all the coordinate combinations and then calculate the convex hull of the obtained (electron efficiency, hadron impurity) pairs. The newly obtained ROC can then be added to the 3rd ROC. This means that all combinations are calculated and the convex hull is calculated. This procedure can be continued iteratively and goes linearly with the number of ROCs in the calculation time. The

quality cuts used for the AddROC method are the minimum quality cuts that can be made. In previous studies of pair analysis, higher quality cuts have always been used, such as requiring 2 TRD hits. Since the electron efficiency changes when the quality cuts change, only the minimum quality cuts (ground cuts) are used in all following investigations.

Ground Cuts: Only particles that meet the ground cuts can be used for the PID in the dielectron analysis.

- STS hits + MVD hits $=> 3$
- χ^2 to vertex < 3
- TOF hit > 0 or TRD hit > 0 or RICH hit >5

7.3 Application of the AddROC Method

As a first step in the analysis of all detector data for classification, the fractions of electrons and hadrons detected by the different combinations of PID detectors are considered. Figure 7.3 shows the histogram of the distinct classes whose data no longer have NaN entries. Figure 7.3 is the real analog to the histogram from the schematic illustration of the AddROC procedure. All classes, for example, RICH & TOF that do not appear in the histogram of Figure 7.3 are represented very sparsely or not at all. This mainly has geometrical reasons. For example, the TRD lies between RICH and TOF so that the class RICH & TOF is almost not represented.

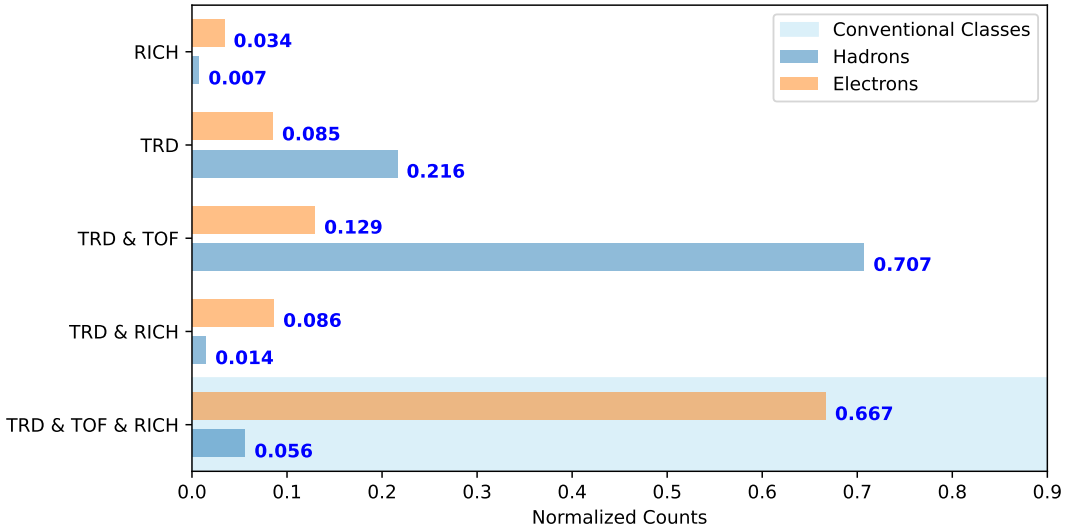


Figure 7.3: Histogram for electrons and hadrons detected by the detector combination shown on the left. Hadrons produce on average fewer RICH hits.

Figure 7.3 shows that only few hadrons fall into the class TRD & RICH & TOF. This is because hadrons tend to have fewer RICH hits than electrons. In the conventional PID methods, it is required that all tracks are in the class TRD & RICH & TOF, and thus have hits in all PID detectors. This requirement leads to the fact that many of the hadron tracks are already filtered out, but also to the fact that 33% of the electrons and thus of the signal are lost. The point of discussion in this thesis will be whether tracks that fall into a class other than TRD & RICH & TOF are still suitable for pair analysis. Next, classifiers must be created for each subclass. The ROC of the test data of the best classifiers in each subclass is shown in Figure 7.4. In the figure, the axes are also scaled so that the maximum electron efficiency and hadron impurity correspond to the proportions of electrons and hadrons in the classes (see Figure 7.3). The blue solid curve shows the ROC for the data in the class. The green dashed ROC was created with the same classifications but tested with the tracks that have TRD & RICH & TOF information. The tracks that have TRD & RICH & TOF information are the least noisy, so the difference between the ROCs is only due to the fact that the tracks that have only TRD information and not TOF and RICH information have a significantly lower quality.

In Figure 7.4 it can be seen that the ROC runs further along the diagonal when less information of the PID detectors is available. Tracks that have only TRD information can be classified worst in the 5 seen classes. The best tracks to classify are those that are seen by all PID detectors. The red dots in Figure 7.4 show the cut values at which the significance for the pair analysis is maximized. The calculation of these points is not trivial because first, the ROC of the whole classifier has to be examined (see Figure 7.4). Each point on the whole ROC belongs to an exact combination of 5 cut values in the subclassifiers. The red points are now exactly these 5 cut values for the point on the whole ROC that maximizes the significance. This is an ideal tool to understand if data that does not have TRD,TOF, and RICH information can be classified at all. In the conventional methods, all tracks that are not seen by all three PID detectors are classified as background. Figure 7.4 shows that by additional classification of the TRD & TOF data, another 5% of the collected electrons should not be discarded to maximize the significance. For the tracks that are only seen by TRD and RICH, 7% of them are not discarded and 2% of the tracks with only RICH information contribute to the total electron efficiency. Since the cut value for the tracks with only TRD information is so close to the origin of the coordinate system, a classification of this data is not very useful. Overall, 14% of the total electrons can be classified as such by additionally classifying the tracks that have only TRD & RICH, or TRD & TOF information to maximize significance. If the underlying simulation is close to reality, it is worthwhile to classify the tracks with missing PID information if sharper cuts are made to the data. The conventional methods could also be improved by requiring a higher electron likelihood and a smaller $\Delta\beta_{\text{Electron}}$ value for the TRD & TOF data for classification.

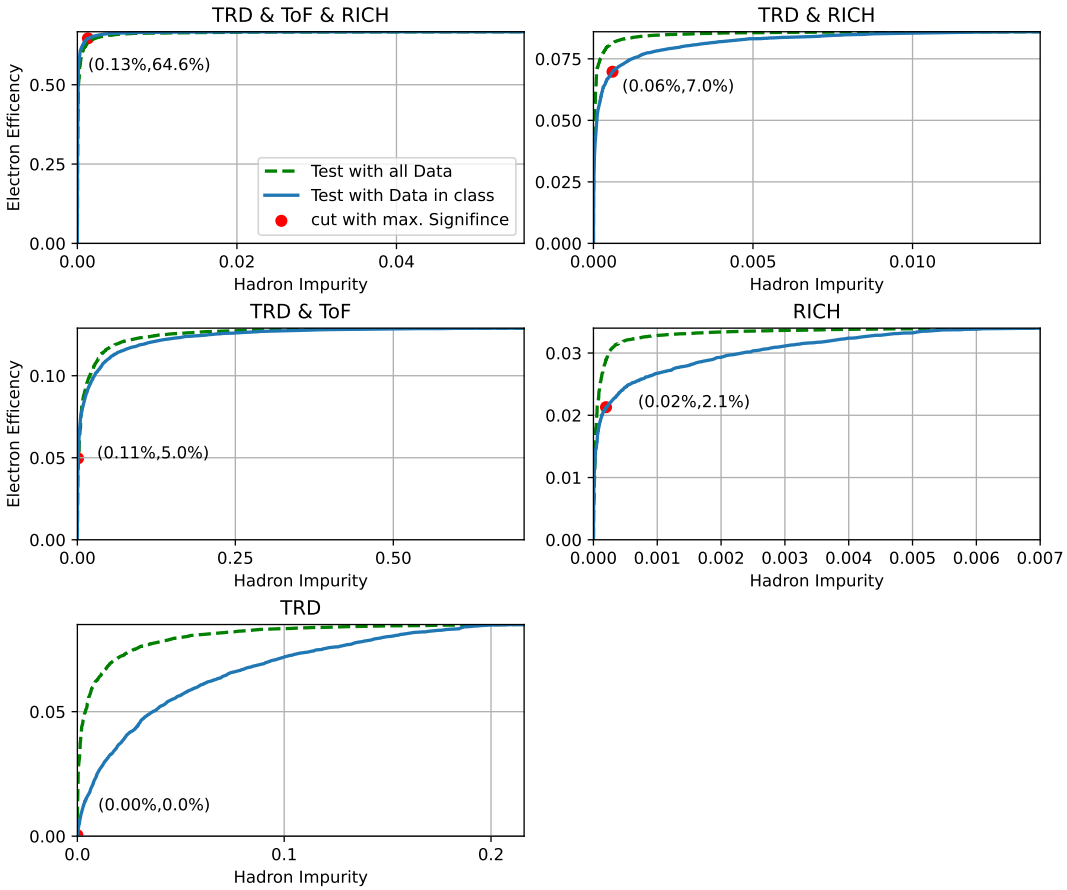


Figure 7.4: ROC is the classifier for the individual subclasses (see title). The data was tested exclusively with class specific tracks, that means all tracks from the class TRD & TOF have no RICH hit. The red dots indicate the cuts that maximize the significance for the pair analysis. The dashed green line was tested with the least noisy tracks. The difference between the green and the blue ROCs is due to the noise in the data.

The red ROC in Figure 7.5 on the right shows the performance of the AddROC classifier, which is made up of 5 sub-classifiers that are each Random Forest classifiers. Figure 7.5 on the left shows a zoom of the ROC for small hadron impurities. The figure also shows the ROC of an XGBoost classifier which distinguishes electrons from hadrons without splitting the data into sub-classes. As a reminder that this is a single classifier, it will be referred to as "single clf. XGBoost". In the same way a single Random Forest classifier (in blue) was investigated. Figure 7.5 on the left shows that the ROC of the ML methods are very similar and there are only minimal differences for small and large hadron efficiencies, respectively. For small hadron impurities, the AddROC classifier wins the race whereas the XGBoost performs better for large hadron impurities. The blue ROC of the single Random Forest is minimally below the other ROCs.

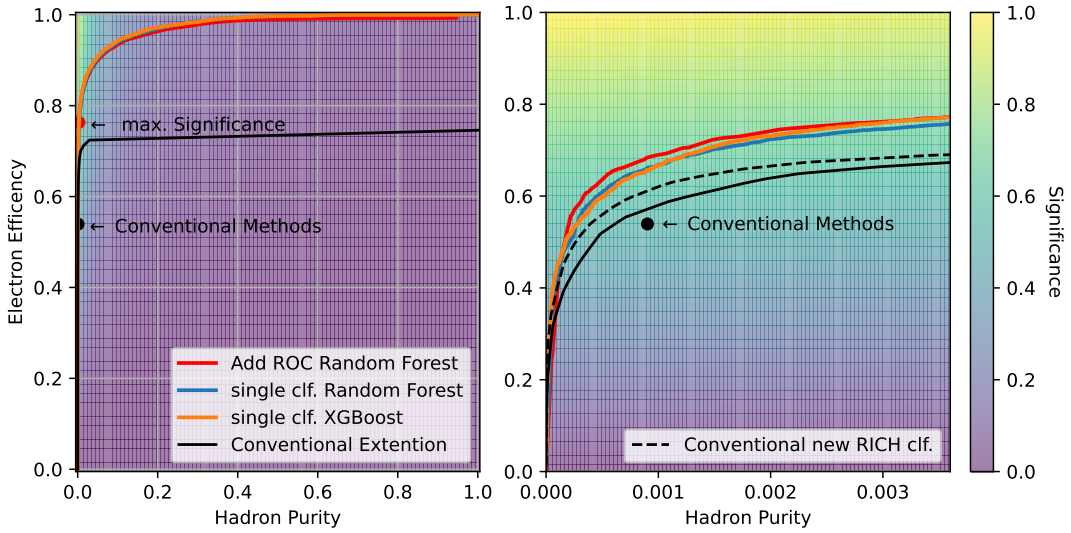


Figure 7.5: Comparison of the performance of different classification methods using data from all PID detectors. The figure on the right shows a zoom in hadron impurities. Modern machine learning methods (in color) improve the PID for the data analyzed here. The best result is obtained with the AddROC classifier, which is based on 5 Random Forest subclassification. The dashed black ROC shows the ROC of the conventional method with the RICH classifier presented in chapter 6. The red dot on the left shows that the maximum significance of the pair-analysis for electron efficiencies is reached around 75%.

In Figure 7.5 the results of the three machine learning methods and the results of the conventional methods are shown. The black dot in Figure 7.5 shows the setting that 80% electron efficiency of the TRD is required and 90% electron efficiency of the RICH detector as well as a cut value of $|\Delta\beta_{\text{Electron}}| > 0.00528$. Furthermore the ground cuts were applied. In principle, there is not one setting for the conventional methods but a lot of fine tuning parameters which result in different black points (see Figure 7.5). Conventional methods with other parameters will be examined later. The black curve in Figure 7.5 shows the calculated ROC of the conventional method. If different settings are chosen for the electron efficiency in TRD and RICH and different cut values for the TOF, then for each of the three settings a point is obtained in the

electron efficiency vs hadron purity diagram. The black curve shows the convex hull of these points. In principle, the conventional methods combine different classifiers similar to the AddROC method, so it is not trivial to find optimal cut values that lie on the convex hull. In Figure 7.5 it can be seen that there are settings for the three cut values of the detectors which have a better performance (higher electron efficiency) than the currently implemented settings. If conventional methods are actually used in the CBM experiment due to their plasticity in decision making, it is useful to perform an analysis of the ROC of the methods to determine the best possible cut values. The ROC (Conventional Extension) in Figure 7.5 converges to an electron efficiency of 75.3%, since 75.3% of the electrons fall into the class RICH & TRD & TOF or the class RICH & TRD and these are the only classes used by the conventional methods for PID. Figure 7.5 also shows the Electron Efficiency for the Maximum Significance (in red). This is clearly above the Electron efficiency of the conventional methods. The analysis of the significance therefore shows that the significance of the experiment increases if no such strong cuts are applied as those implemented in the conventional methods. In order to test whether the conventional methods are a suitable classification method within the class TRD & TOF & RICH, the Machine Learning classifier of the subclass TRD &TOF & RICH (in blue) was compared with the ROC of the conventional methods, as shown in Figure 7.6. Furthermore, Figure 7.6 shows how the ROC improves as more and more classes are added.

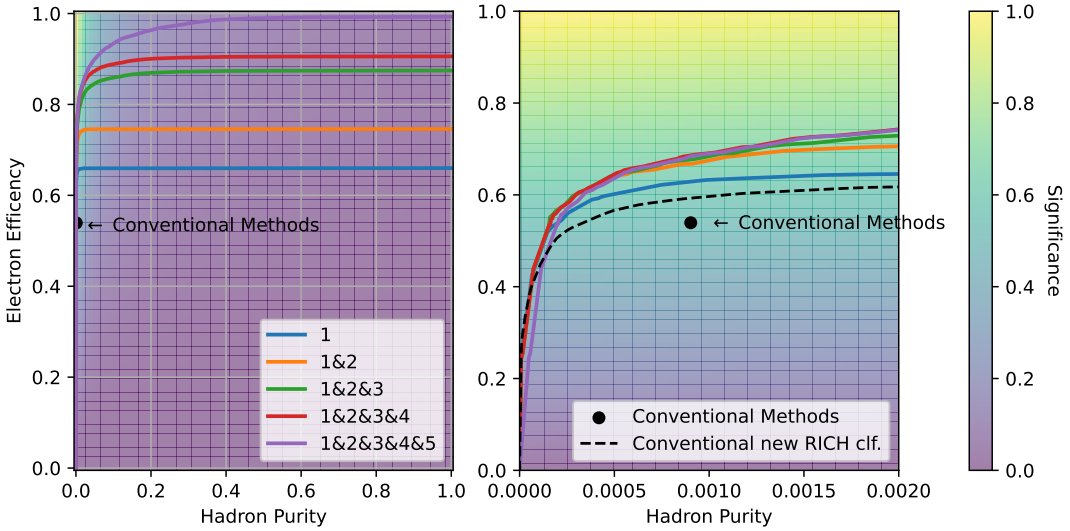


Figure 7.6: The numbers represent the classes 1: TRD & TOF & RICH, 2: RICH & TRD, 3: TRD & TOF, 4: TRD and 5:RICH. The performance is improved by adding another class. The blue curve shows the ROC of a classifier, which only evaluates data of the class TRD & TOF & RICH, just like the conventional methods do. The ROC of the conventional methods with the updated RICH classifier (dashed black) is lower than the ROC of the machine learning classifier.

The result of the Figure 7.6 is that Machine learning classifiers in the data subclass TRD & TOF & RICH perform slightly better than the conventional methods with

the subdata RICH classifiers presented in chapter 6. This is mainly due to the fact that no cube is cut out of the three-dimensional PID data space as illustrated in chapter 4. Such a cube is not adapted to the more complicated data distributions, which leads to losses in electron efficiency. Nevertheless, it should be noted that by gaining a much larger decision making transparency that the updated conventional methods have, only about 3% less electron efficiency is achieved within the class TRD & TOF & RICH. As mentioned above, there is not exactly one conventional method but a lot of cuts on the variables that provide a fine tuning. Cuts like TRD Hits > 2 fall into the class of quality cuts and can be varied arbitrarily. For the most important combinations of quality cuts the electron efficiencies and hadron impurities were calculated (see Figure 7.7).

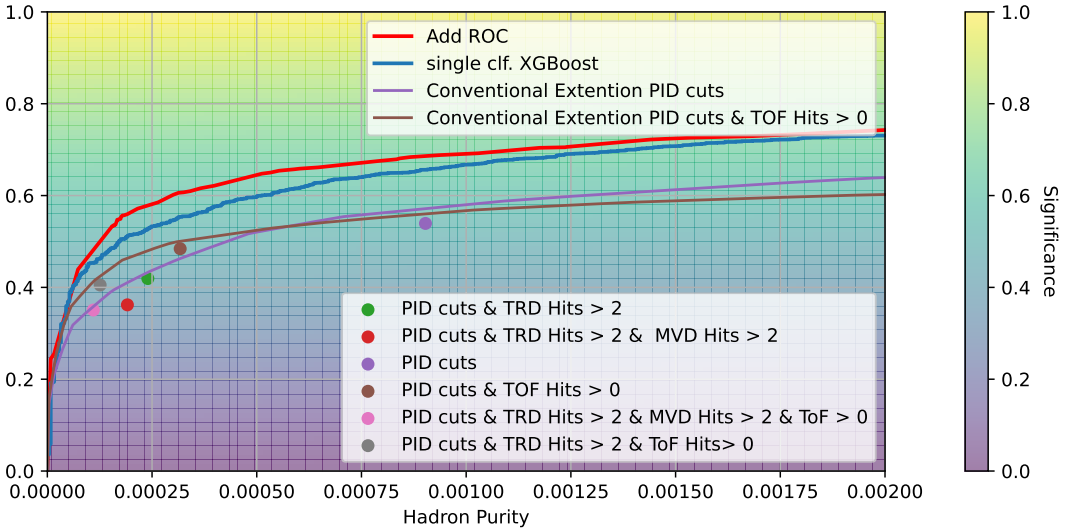


Figure 7.7: ROC of different machine learning methods compared to different quality cut settings of the conventional methods. PID cuts correspond to the cuts on the Likelihood the ANN output and the TOF $\Delta\beta_{\text{Electron}}$ value. The ROCs of the conventional methods for two different quality cut settings were also shown. The ROCs for the cut settings purple and brown are above the points. All points belong to electron efficiency TRD = 80%, electron efficiency RICH = 90% and $|\Delta\beta_{\text{Electron}}| < 0.00528$. All points and ROC of the conventional methods are below those of the machine learning methods.

In Figure 7.7 it can be seen that for all tested quality cut combinations, the conventional classifications method performs worse than the Machine Learning methods. However, it is important to note that such an analysis can be used to compare different quality cuts so that the significance of the pair analysis, which is shown as color coding in Figure 7.7, can be maximized for the conventional methods as well.

Figure 7.8 on the left shows a test data set containing the approximate ratio of expected particle tracks in the CBM PID detectors. On this data set the conventional methods were tested. It can be seen that they successfully filter electrons out of the particle tracks. After testing the conventional methods, the AddROC filter was applied to the data set. It is important that a cut value is chosen so that the hadron

impurity is equal to that of the conventional methods. This is important for a fair comparison, which in this case shows that the AddROC filter is able to classify more electrons from the data correctly and filters out as many hadrons as the conventional methods.

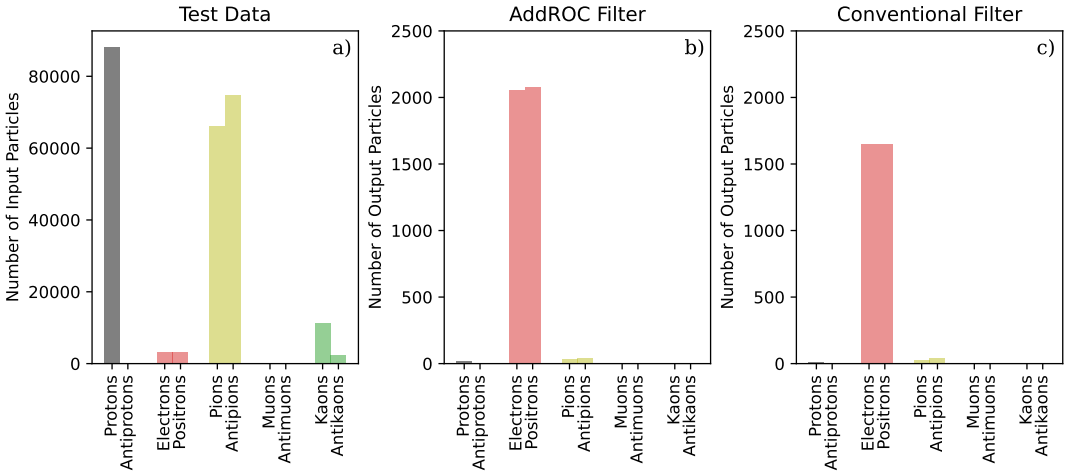


Figure 7.8: Left: Number of test data used as input for the classification of electrons for the AddROC method and the conventional methods. The input corresponds approximately to the expected particle ratios to be classified in the experiment. The AddROC filter (center) can classify electrons as such with the same hadron impurity as the conventional methods (right).

So far, it seems that nothing will stop the application of the AddROC method, which is the best performing method for small hadron impurities. However, there are some arguments against its use (in red), which will be compared with its advantages (in green) in the following.

- AddROC classifier performs best among all classifiers tested on simulated data.
- It is possible to find out which classes are not suitable for PID (e.g. class with only TRD info).
- It can be analyzed how much classification power is lost by tracks with low quality. (green and blue ROC in Figure 7.4)
- It is possible to calculate the cut combinations for the classifiers belonging to electron efficiency, hadron impurity pairs on the convex hull. Not all cut combinations are suitable for classification.

→ AddROC method gains knowledge that becomes interesting for the whole analysis and thus also for conventional methods.

- The AddROC method depends very much on the calculated proportions of the classes. Although these can be calculated accurately by the simulation, they cannot be estimated in the experiment.
- Especially the classes with more noise in the data can deviate even more in the experiment. The calculation of the optimal cut value in the class with complete PID information depends on the other classes.

→ Inaccuracies in the simulation lead to strong inaccuracies in the classification.

If the simulation data have a very high accuracy, the AddROC method is also suitable as a classifier for the experiment. A single XGBoost classifier should still give good results for low hadron purity even with real data different from the simulated data.

7.3.1 Test on Pair Analysis

For the quality cuts TRD Hits >2, MVD Hits >2 and TOF Hit > 0 (compare pink point Figure 7.7) the classification methods were tested directly on the pair analysis. In Figure 7.9 it can be seen that the single XGBoost classifier has more than 30% as high electron efficiency at the same hadron purity. However, this information can already be obtained from Figure 7.7, since the ROC of the XGBoost runs above the red dot.

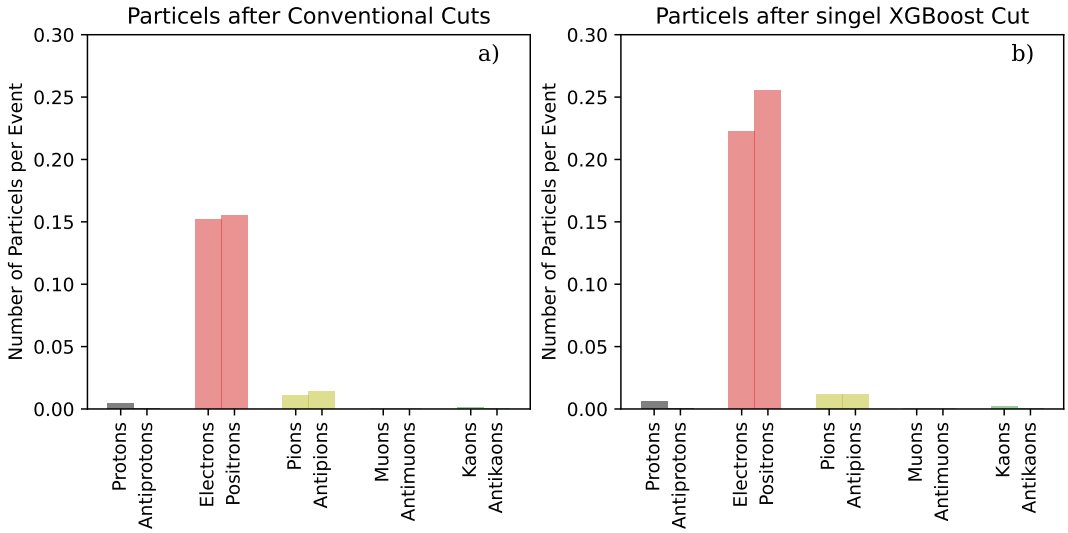


Figure 7.9: Histogram of the number of particles per event after filtering electrons from the total tracks by two different classification methods. On the right are the results for the convention cuts with quality cuts of TRD Hits >2, MVD Hits >2 and TOF Hit > 0. The left panel shows the results of the XGBoost classifier. The XGBoost classifier achieves a higher electron efficiency with the same hadron suppression.

Figure 7.10 shows pair analysis spectra resulting from using the conventional filters and the XGBoost filter. For this 25000 events were simulated and their particle tracks filtered by the PID methods. The black curve shows the counts of signal and background. The signal part is also colored for the different decay processes. The most prominent points are the $\omega \to e^+e^-$ decay (in red) and the $\phi \to e^+e^-$ decay (in blue). The centers of the peaks are the respective masses of the mesons. It is also noted that the signal part is several orders of magnitude smaller than the background, which is why a much higher statistic realized in the experiment is required. Figure 7.10 below

shows the difference of the two upper spectra. Furthermore, the significances for the two upper pair analysis spectra were calculated in Figure 7.10.

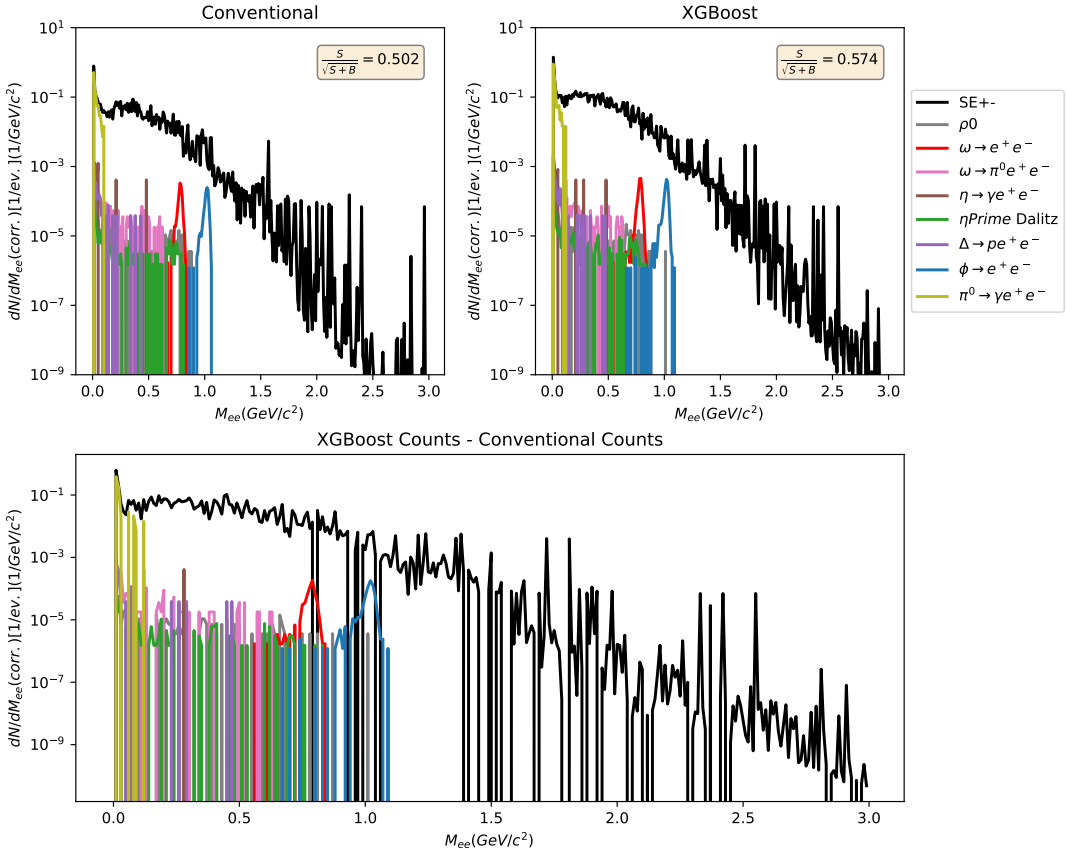


Figure 7.10: Above: compared the invariant mass spectra after pair analysis with the conventional methods on the left and the XGBoost classifier on the right. The significance in the boxes was calculated by the sum of all signal points and signal + background points. The XGBoost classifier achieves a higher significance. Below is the difference of the spectra. The application of the XGBoost classifier provides more signal in the pair analysis. The increase of the background is inevitable because by more electrons there are also more combination possibilities for the background.

The results from Figure 7.10 confirm that the XGBoost classifier can increase the significance of the experiment. This can be observed when looking at the peaks of the ω and ϕ mesons. These have values in the order of magnitude of the peaks for the difference of the spectra (see Figure 7.10 below). It is important to note that the XGBoost classifier was not set to maximize significance but to achieve the same hadron impurity as the conventional methods, otherwise a fair comparison is not possible. The significance was calculated by taking the signal plus background values (S+B) from the integral of all bins of the black curve and the signal measurements from the integral of the sum of the colored signal histograms. For the conventional methods a significance of 0.502 and for the machine learning methods a significance of 0.574 was achieved. However, this significance can be improved in both cases if

other cuts are made.

7.4 Analyse and Optimization of the Final Classifier

The first step in optimizing a classifier is to analyze its performance for different momenta. Figure 7.11 shows the ROC of the XGBoost classifier in a three-dimensional diagram for different momenta. Figure 7.11 on the left shows two-dimensional slices of the blue plane in three-dimensional space for different electron efficiencies. The Conventional methods are set so that theoretically for all momenta an equal electron efficiency is reached. This works well for a single PID detector, but when the PID detectors classify together, the conventional methods do not achieve a uniform electron efficiency for all momenta. In the case of machine learning methods, a different threshold can be required for each momentum so that the same electron efficiency is always obtained.

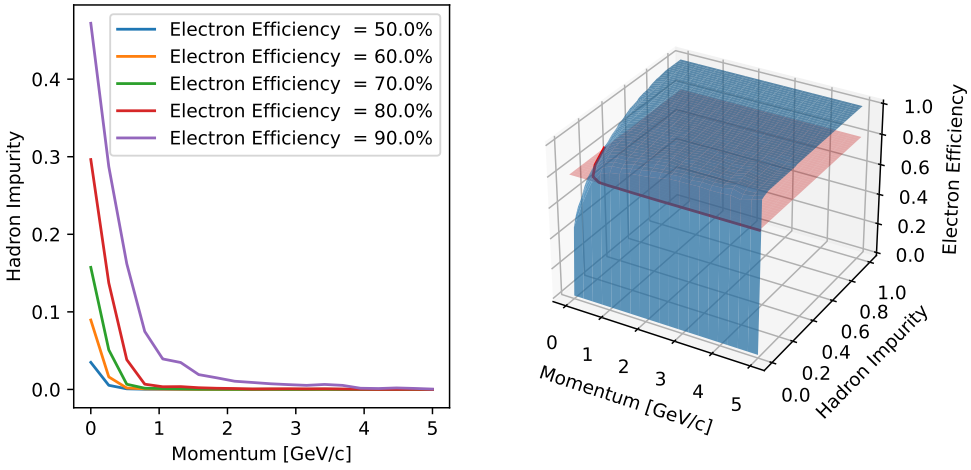


Figure 7.11: Right: ROC of the XGBoost classifier for different momenta. For smaller momenta, the ROCs run further along the diagonal. For higher momenta, the ROC is very close to the optimum. Left: a section of the three-dimensional curve for different electron efficiencies (see red plane).

In Figure 7.11 it can be seen that the PID for small momenta does not work as well as for large momenta. This can be seen from the fact that for large momenta the ROC runs very close to the optimal point at the top of the corner and even forms almost a sharp ridge. For small moment, the ROC runs further along the diagonal. That the ROCs evolve over the momenta is not due to the poor classification power of the machine learning methods but to the ability of the CBM detector to distinguish electrons from hadrons. The plot in Figure 7.11 is therefore an ideal analysis tool for evaluating the quality of the PID of the CBM detector, since it should, if possible, achieve a good PID across all momenta.

7.4.1 SHAP Values

SHAP values (SHapley Additive exPlanations) is a method based on cooperative game theory and used to increase transparency and interpretability of machine learning models. The SHAP values provide information about the proportion of a certain feature in the entire XGBoost classifier that was involved in the decision. Figure 7.12 shows the information of the mean SHAP values. The exact mathematical calculation of the SHAP values can be found in [25].

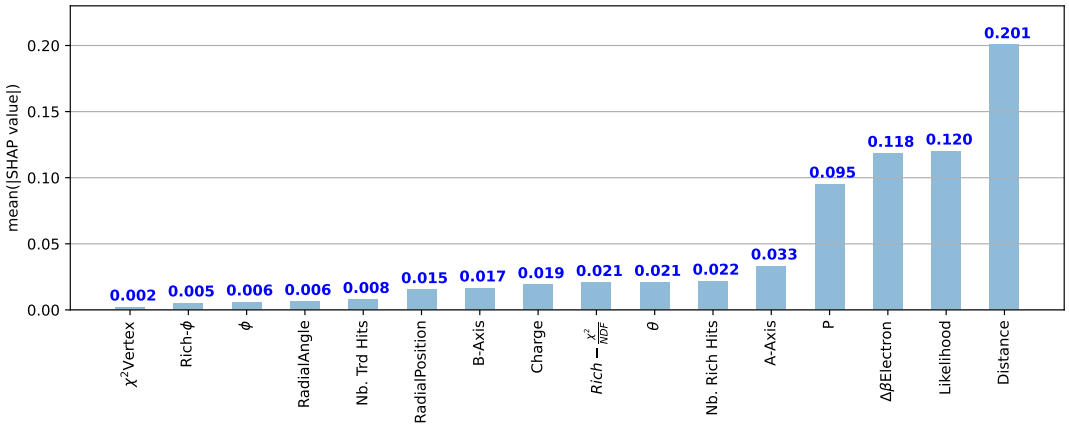


Figure 7.12: Average SHAP values of the individual features in the single XGBoost classifier. The values give the proportion of the total decision of a feature. The distance of the track to the RICH ring ellipse is the largest contributor to the classification between electrons and hadrons.

In Figure 7.12 it can be seen that the distance of the track to the origin of the photon ellipse of the RICH detector is mostly involved in the decisions. The reason for this is that larger distances to the tracks belong to pion tracks which were erroneously assigned to other photon ellipses. Therefore, tracks with a higher distance to the photon ellipse probably did not produce any Cherenkov radiation and are therefore hadrons. It is important to note that for Figure 7.12 only tracks with information from all PID detectors were used. As expected, the likelihood and $\Delta\beta_{\text{Electron}}$ value of the TOF contribute most to the decision. These are the most important information of the TRD and TOF. The χ^2 to vertex value contributes least to the PID and could theoretically be removed from the classifier without much loss. Figure 7.12 gives information about which values are least important for the classification and therefore are sorted out first, if a fast (computing time) classifier is needed for the experiment. However, there are features that could be added to the network: The χ^2/NDF-TRD value and the χ^2/NDF-STS. Since their computation in CBM-ROOT could not be reproduced exactly in the course of this thesis, it is a point to test later if it is worth implementing these values.

7.4.2 Correlation Matrices

Correlational matrices are useful to investigate whether it is possible to extract some data from the classifier using the formula 7.1 for independent probability variables. The application of this formula has already been explained in chapter 6.

$$P(P,\ldots,\mathrm{RICH}-\phi,\ldots,\frac{\chi^2}{\mathrm{NDF}}) = \underbrace{P(P,\ldots,\frac{\chi^2}{\mathrm{NDF}})}_{\text{classifier output}} \cdot \underbrace{P(\mathrm{RICH}-\phi)}_{\text{Likelihood}} \tag{7.1}$$

Figure 7.13 shows the correlation matrix for all features of the XGBoost classifier. The data used are the hadron tracks which are to be distinguished from the electron tracks. The correlation matrix for electron can be found in the appendix of Figure .2. From the correlation matrices, it can be seen that the RICH tilt angle ϕ correlates only weakly with the other data, so that this feature can theoretically be extracted from the classifier. In addition, the likelihood data for hadrons and electrons correlate only weakly with the nb. of TRD hits, so the full classification of TRD is uncorrelated with the other data. The data of the TOF ($\Delta\beta_{\mathrm{Electron}}$) correlate with those of the TOF.

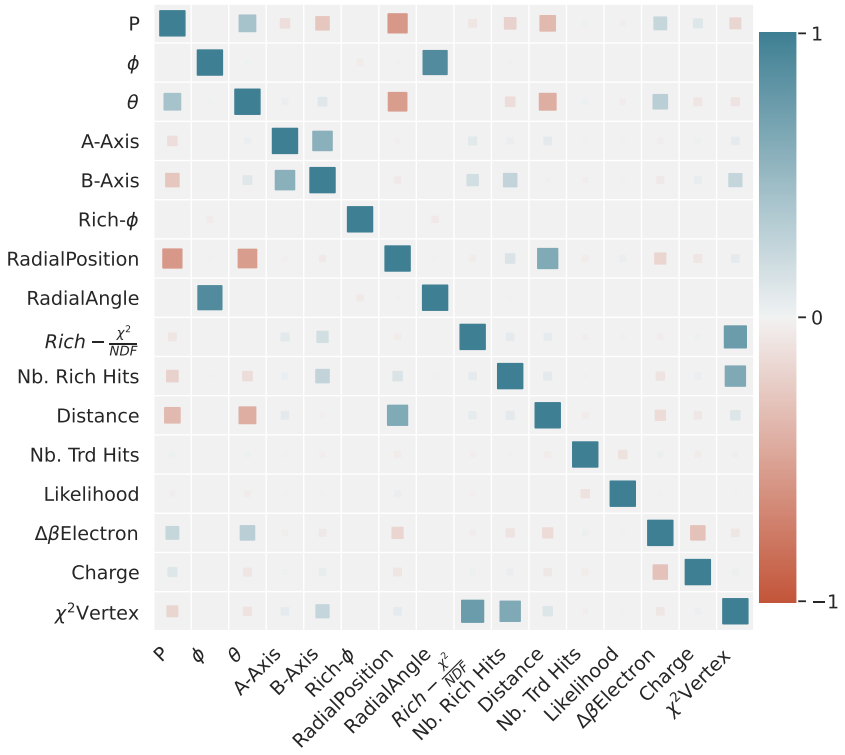


Figure 7.13: Correlations matrix of all input variables of hadrons for a classifications method. Strong negative correlations are shown as larger boxes. The representation allows a quick reading of uncorrelated variables.

The possibilities proposed here to extract the RICH-ϕ or the TRD data from the whole classifier are suitable if the algorithm should become more efficient in computing time. Due to the simplicity of having all data evaluated by a single classification function, the suggestions have not yet been integrated.

8 Outlook on Further Research Questions

In the following, two main research topics are outlined, which have been investigated in the course of this work, but have not yet led to applicable results. Both topics offer open research questions which are briefly listed.

8.1 Pair Classifier

Up to now there are only tracks classifiers for the dielectron pair analysis, which assign a track to signal (as electron) or background. In the pair analysis, a pair of tracks is only a signal if the two e^+, e^- tracks come from the same decay. It is also possible to use a pair classifier instead of a track classifier, which has always a pair of tracks as input and therefore decides if it is a signal pair or background. Figure 8.1 illustrates this approach and the associated problem.

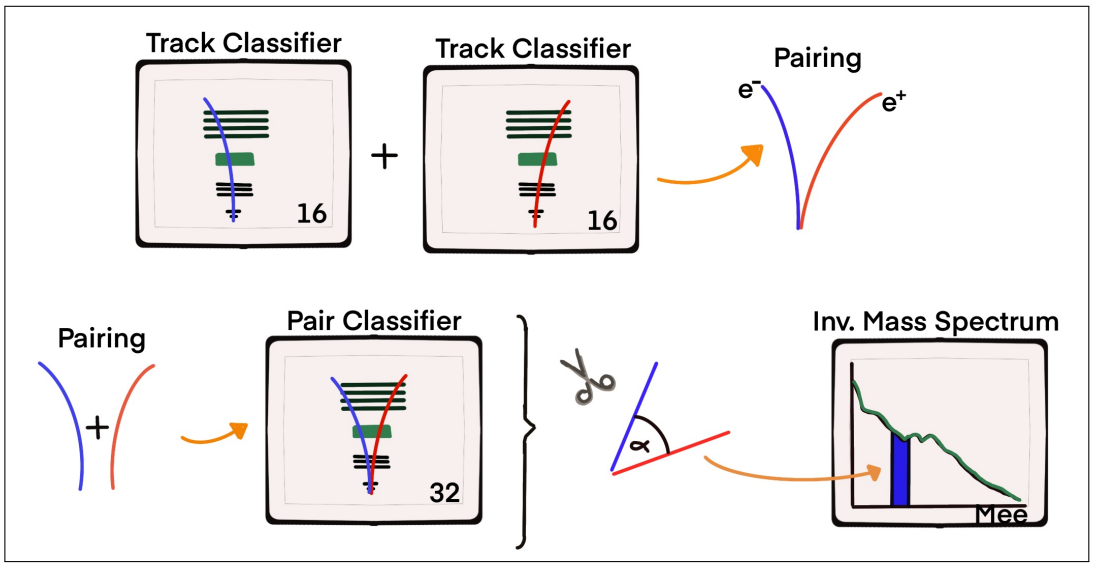


Figure 8.1: Illustration of the possibility of a pair classifier. After pairing the tracks, the pair is classified as signal or background. Problematic are topological cuts that such a pair classifier could perform.

In the following the advantages and disadvantages of such an approach are discussed. The main disadvantage is the topological cut which is illustrated in Figure 8.1.

- (Topological cut) Since the data for a pair also contains the information about the invariant mass indirectly, it can be cut on this, so that the cut has a direct effect on the measured spectrum. This can make the reconstruction of the signal very difficult.
- Since all possible pairs must be classified and there are $n \cdot m$ pairs, with n positive and m negative tracks, there are $n \cdot m/(n+m)$ times as many data to be classified which corresponds to larger computation time.
- The classifier input is doubled, which means that a longer evaluation time of each pair is to be expected.
- It is probably possible to obtain a higher significance, since the pair classifier can more accurately separate background from signal pairs.

8.2 Rejection of Conversion Electrons

The right panel of Figure 8.2 shows the fraction of conversion electrons in the particle spectrum after the application of the XGBoost filter. This fraction is significant and can not be filtered out by the XGBoost filter, because conversion electrons can not be distinguished from signal electrons by the track information. All possible combinations of electrons and conversion electrons contribute the most to the background and therefore the rejection of the conversion electrons will be one of the main tasks of modern pair analysis.

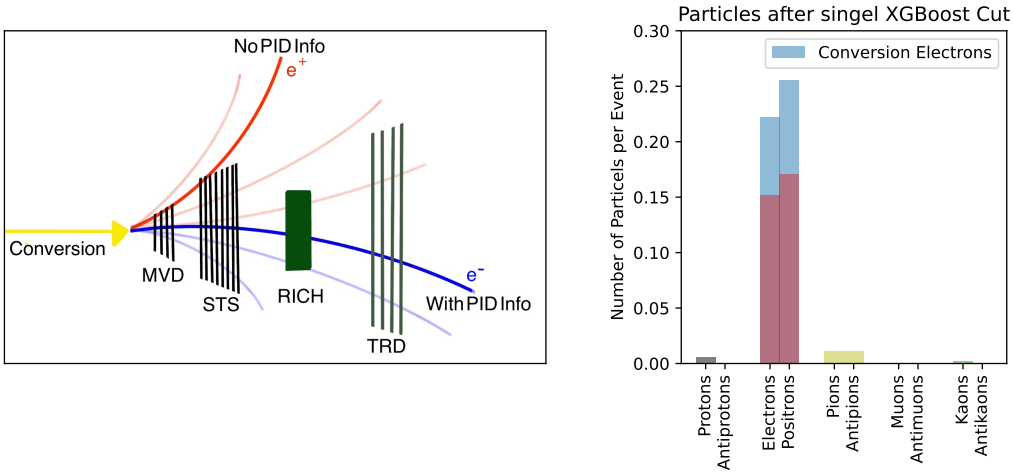


Figure 8.2: Left: Illustration of a method for identifying conversion electrons explained in [26]. Right: Filtered electron tracks according to the XGBoost classification algorithm.

Since there is no way to filter out the conversion electrons based on the track information, topological properties of the conversion have to be used. Since the parent particle of the electron positron pair for conversion electrons is a photon, which is known to be massless, an invariant mass around 0 is expected for the conversion

electron positron pair. If it is naively assumed that the electrons are massless, the equation 8.1 for the invariant mass of the pair is obtained. Here θ is the opening angle and p_{e^+}, p_{e^-} the momenta of the individual particles.

$$M^2 = p_{e^+} \cdot p_{e^-}(1 - \cos(\Theta)) \tag{8.1}$$

Since the momenta are not negative, it can be directly observed that a smaller opening angle is to be expected for the conversion pair. Now, there is the possibility to cut directly on small invariant masses so that the contribution from the conversion electron pairs is reduced. Here again, the problem illustrated in Figure 8.1 arises that cutting on the invariant mass spectrum should be avoided at first. However, there are also a lot of conversion electron pairs, which can be filtered out despite topological cut without cut to the spectrum. These pairs are characterized by the fact that one partner has PID information and the other partner is only detected by the tracking detectors, as illustrated in the left panel of Figure 8.2. A cut applied to the conversion electron pair in the convention methods assigns an electron with full PID information to the conversion electron which has the smallest opening angle to the full PID information electron track. If the opening angle is small enough, both tracks can be removed. The track with the smallest opening angle is without question the most likely partner for a conversion electron track. Mathematically, however, it is only an approximation to consider only the closest partner. The exact mathematical relationship was investigated in the course of this thesis and is illustrated in Figure 8.3. It need to be decided about an electron track (in blue), whether it is a conversion electron. There are several possible partners (in red). For each of these partners, there is a probability that the pair is a conversion electron. This probability can be determined by calculating the invariant mass and reading the probability from the Figure 8.3. Figure 8.3 is illustrating an analogy for calculating this probability. Paul is currently dating 3 girls, Eve, Susi and Clara. If he would meet only Eve, the probability that they become a couple would be 80%, with Susi it would be 50% and with Clara 60%. What is the probability that Paul will have a girlfriend after dating? The calculations for this, are analogous to those for the conversion electrons. Since Paul lives monogamously, just as a conversion electron has only one partner, conditional probabilities must be considered. The probability that Paul will actually get together with Clara is the probability that they will be a couple multiplied by the probability that he will not get together with the other two girls.

$$P(\text{Paul is taken}) = \frac{A}{A + \underbrace{(1-0.5)(1-0.8)(1-0.6)}_{\text{Paul is single}}} = 86.66\% \tag{8.2}$$

$$A = \underbrace{0.8(1-0.5)(1-0.6)}_{\text{Paul goes with Eve}} + \underbrace{0.5(1-0.8)(1-0.6)}_{\text{Paul goes with Susi}} + \underbrace{0.6(1-0.5)(1-0.8)}_{\text{Paul goes with Clara}} \tag{8.3}$$

With a few modifications, the problem can be solved more efficiently from a math-

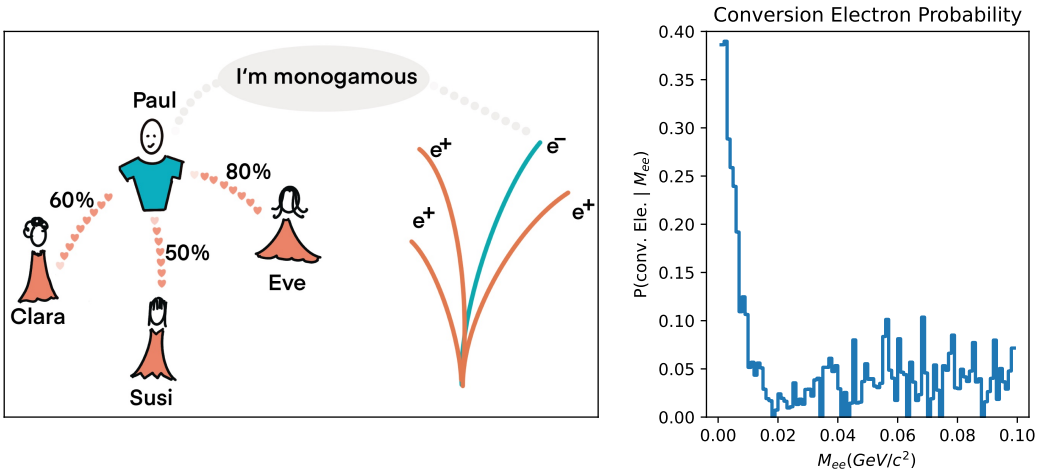


Figure 8.3: Left: Illustration of the calculation of the probability of a conversion electron track. Right: probability that an electron-positron pair with a given invariant mass is a conversion electron. Calculation by simulated pair analysis in which the histogram of the measured Conversion Electron pairs is divided by the number of conv. Ele. pairs and Ele. pairs.

ematical point of view. Formula 8.5 reduces the example to an arbitrary number of possible parameters for the conversion. The p_i are in the case of Paul the single probabilities (80%, 60%, 50%), in the case of the conversion electrons they are the probabilities that the i-th track is a suitable candidate for the conversion.

$$P = \frac{\sum_{i=1}^{n} \frac{a \cdot p_i}{(1-p_i)}}{\sum_{i=1}^{n} \frac{a \cdot p_i}{(1-p_i)} + a} = \frac{k}{k+1} \quad \text{with } k = \sum_{i=1}^{n} \frac{p_i}{1-p_i} \tag{8.4}$$

$$a = \prod_{i}^{n} (1 - p_i) \tag{8.5}$$

Such a mathematical analysis can, if it is programmed efficiently, output probabilities for each track whether it is a conversion electron. A cut can then be applied to this probability. The application of the formula 8.5 in the pair analysis has not yet been completed in the course of this thesis, so that there are still many open questions in the research focus of the conversion electron suppression:

1. What proportion of the conversion electrons fall into the case shown in 8.2?
2. Is it worth calculating the probability of a track being a conversion electron track or is it better to assign it to the nearest neighbor?
3. How computationally intensive is the calculation of the probability for a conversion track?

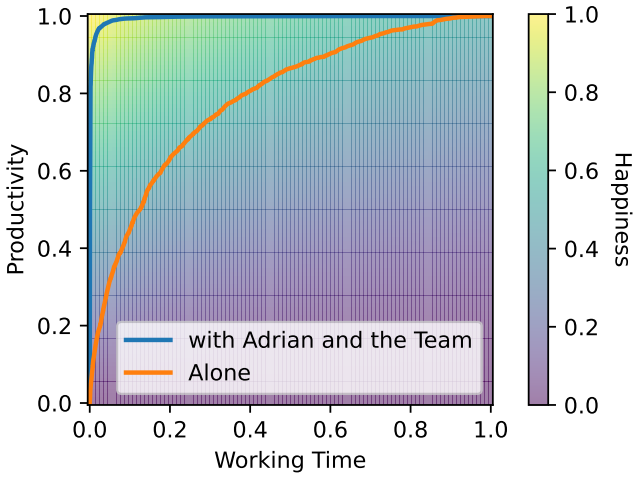
9 Conclusion and Outlook

In this thesis different machine learning methods for particle identification were tested in the CBM experiment. The main goal is to distinguish electrons and positrons from other particle tracks detected by the experiment, such as pions. The identification of electrons from the total detected particles is especially important for the dielectron analysis. This will provide new insights into the phase diagram of matter at the CBM experiment and thus extend the understanding of QCD matter. The CBM detector consists of several sub-detectors, three of which are serving for for particle identification. In conventional methods of electron identification, each of the sub-detectors decides for itself whether a corresponding track is an electron or not. For this purpose, the TRD uses the likelihood method which directly outputs the probability for an electron or another particle type. The likelihood method can be improved only minimally or not at all by machine learning methods, since the knowledge about the probability density in the parameter space is already the optimal classifier. Due to the good transparency of the likelihood method it can be seen as superior to modern machine learning methods. The RICH detector decides in the conventional method by a neural network trained in 2008 with a comparatively simple architecture. The neural network was improved at the beginning of the thesis by using an XGBoost classifier. An XGBoost classifier has already been successfully used in high energy physics for the discovery of the Higgs boson, and was also able to give the best results in the case of the RICH detector. The research focus of this thesis was to train a classifier that uses all PID information at once for the electron PID. As input for this classifier global track information like the 3D momentum of the particle or the distance of the track to the vertex as well as information from the particle identification detectors are used. By combining the data collected by the subdetectors TOF, TRD and RICH, a single classifier can be trained, which has as input a total of 16 detector variables and as output the binary classification as an electron. In the course of this thesis, the AddROC method was developed specifically for simulated data from the CBM experiment, which combines classifiers from different subdata classes. Since the tracks in the CBM experiment are not always seen by all detectors, it is particularly useful to define sub-data classes that contain, for example, only TRD & TOF data and no RICH information. This has the advantage that no data with missing information is fed into the classifier within the sub-data classes.
The result of the AddROC method is that the combination of different classifiers performs best and can greatly improve the conventional methods. For the same electron efficiency, the AddROC classifier achieves more than 30% higher electron efficiency. Results of a single XGBoost classifier applied to all data achieve similarly good results and beat the AddROC classifier for large hadron impurities. The AddROC method is particularly useful for determining whether data that are only

detected by the RICH detector, for example, are classifiable. In conventional methods, data not seen by TRD, TOF and RICH (full PID info) are not evaluated further. The result is that by using the data in TRD & RICH, TRD & TOF and (only) RICH classes, the electron efficiency can be improved. For this it is important that sharper cuts are applied to the classes with less PID info so that only the best tracks are classified as electrons. In the course of this work, a method was developed to compute these best cut values by estimating the significance for the pair-analysis. Furthermore, based on the AddROC method, a method was developed to optimize the cut settings for the conventional methods. This is based on calculating the convex hull of the electron efficiency and hadron impurity pairs of all possible PID cut settings and then maximizing the significance along this curve. To confirm the results of the track classification by the machine learning methods, they were directly tested on the pair analysis. The result is that a single XGBoost classifier can increase the significance of the measurement on the CBM experiment. Although the AddROC method gives the best results in simulation, the method is very sensitive to changes in the fractions of electrons and hadrons that fall into the single subclases such as TRD & TOF. Since these values are not exactly predictable, it is recommended to implement a single XGBoost classifier in the software of the experiment. Furthermore, it is recommended to adjust the cut to achieve the same electron efficiency for each momentum range. For this purpose, the ROC was calculated for different momenta and the cut value was determined. At the end of this thesis some ideas and prospects were presented which could make the rejection of the conversion electrons more efficient and a possible pair classifier was discussed. Overall, it was shown that machine learning methods have high potential to improve pair analysis in the CBM experiment. Particularly good results were achieved with an XGBoost classifier applied to all detector data.

Danksagung

Als ich meine Masterarbeit im Sommer 2021 begonnen habe, kam ich gerade aus meinem Auslandsstudium in Sevilla in Spanien zurück. Meine Zeit dort war wunderschön und intensiv, dennoch ist mir das Studieren auf einer Sprache, die ich zu Beginn gar nicht beherrschte und die vielen Vormittage, die ich wegen Corona alleine vor dem PC verbringen musste schwergefallen. Ich beschloss durch die Gänge des Instituts in Münster zu laufen und mir die Arbeitsgruppe auszusuchen in der ich mich wohl am wohlsten fühlen werde. In einem der Büros, die ich besucht habe, saß Adrian, der sich direkt viel Zeit genommen hat, um mir seine Arbeit zu erklären. Für mich war das einer der wichtigsten Momente, weil ich wusste, dass ich von nun an nicht mehr ganz alleine durch das Studium gehen werde. Mein Dank in diesen Zeilen geht an diesen jungen Doktoranden, der mich seitdem in allem unterstützt hat und niemals gesagt hat, dass er sich für etwas keine Zeit nehmen möchte. In meinem Leben habe ich oft gemerkt, dass man mit 40% der investierten Arbeitszeit 80% der Arbeit erledigen kann. Eigentlich beschreibt dieser Satz eine ROC, die Kurve, die ich die letzten Monate täglich gesehen habe. Doch, wenn man ein tolles Team wie die AG Andronic an der Seite hat, dann verläuft die ROC noch näher an der y-Achse. Danken möchte ich auch Prof. Dr. Anton Andronic und Prof. Dr. Tetyana Galatyuk, die mir viele meiner Fragen schnell beantworten konnten und deren unglaubliche Begeisterung für das Forschungsprojekt auf mich abgefärbt haben.
Eine weitere Person, die nicht Teil des Experiments ist, der ich aber dennoch danken möchte, ist Prof. Dr. Eleonora Viezzer, da sie sich bereit erklärt hat, zweite Gutachterin für diese Arbeit zu sein und somit zu einer tollen Kooperation zwischen Münster und Sevilla beiträgt.



Appendices

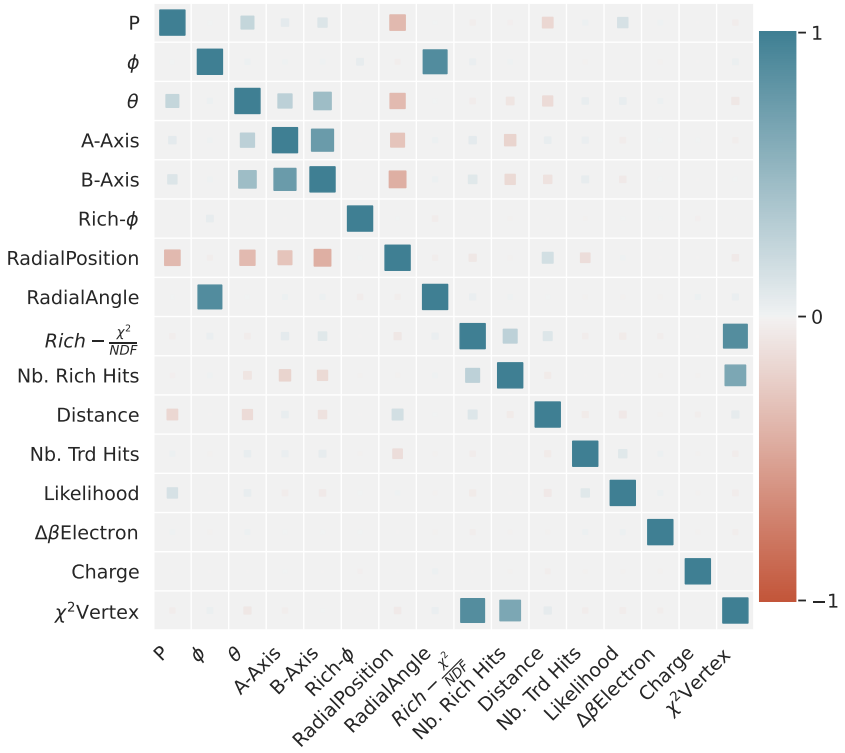


Figure .1: Correlations matrix of all input variables of electrons for a classifications method. Strong negative correlations are shown as larger boxes. The representation allows a quick reading of uncorrelated variables.

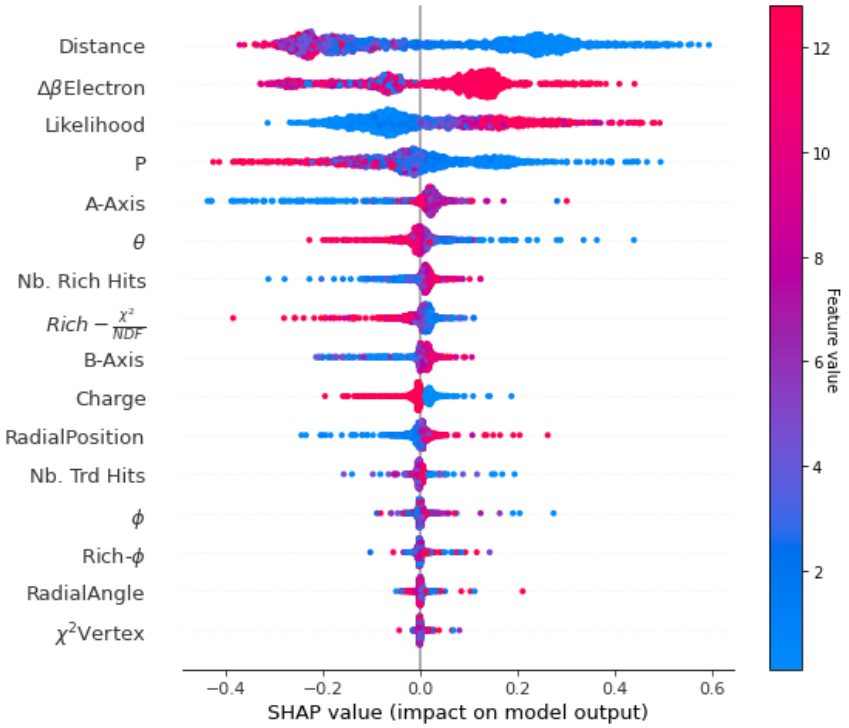


Figure .2: Beeswarm plot, designed to display an information-dense summary of how the top features in a dataset impact the single XGBoost model's output. Each instance the given explanation is represented by a single dot on each feature fow. The x position of the dot is determined by the SHAP value.

Bibliography

[1] *Cbm-homepage*, `https://www.gsi.de/work/forschung/cbmnqm/cbm`, accessed: 2022-06-24.

[2] T. Ablyazimov et al., *Challenges in QCD matter physics –the scientific programme of the compressed baryonic matter experiment at FAIR*, The European Physical Journal A, 53 (2017).

[3] V. Friese, *The high-rate data challenge: computing for the CBM experiment*, Journal of Physics: Conference Series, 898, 112003 (2017).

[4] E. Bechtel, *Development of a detector simulation and reconstruction for the cbm-trd and the analysis of thermal dielectron pairs in 12 a gev au+au collisions*, Doktorarbeit (2020).

[5] Z. Fodor und S. Katz, *Critical point of QCD at finite T and μ, lattice results for physical quark masses*, Journal of High Energy Physics, 2004, 050 (2004).

[6] B.-J. Schaefer, J. M. Pawlowski, und J. Wambach, *Phase structure of the polyakov-quark-meson model*, Phys. Rev. D, 76, 074023 (2007).

[7] I. Froehlich, L. C. Boado, T. Galatyuk, V. Hejny, R. Holzmann, M. Kagarlis, W. Kuehn, J. G. Messchendorp, V. Metag, M. A. Pleier, W. Przygoda, B. Ramstein, J. Ritman, P. Salabura, J. Stroth, und M. Sudol, *Pluto: A monte carlo simulation tool for hadronic physics* (2007).

[8] P. P. Bhaduri et al., Phys. Rev., C, 89 (2014).

[9] R. Rapp und J. Wambach, *Low-mass dileptons at the CERN-SpS: evidence for chiral restoration?*, The European Physical Journal A, 6, 415 (1999).

[10] H. J. Specht, *Thermal dileptons from hot and dense strongly interacting matter*, AIP Conference Proceedings, 1322, 1 (2010).

[11] R. Rapp und H. van Hees, *Thermal dileptons as fireball thermometer and chronometer*, Physics Letters B, 753, 586 (2016).

[12] A. Puntke, *First mtrd performance studies in the mcbm 2020 campaign*, Masters thesis (2021).

[13] C. Sturm und H. Stöcker, *The facility for antiproton and ion research fair*, Physics of Particles and Nuclei Letters, 8, 865 (2011).

[14] *Mega bauprojekt fair*, `https://www.gsi.de/forschungbeschleuniger/fair/bau_von_fair/mega_bauprojekt_fair`, accessed: 2022-06-20.

[15] *The Transition Radiation Detector of the CBM Experiment at FAIR : Technical Design Report for the CBM Transition Radiation Detector (TRD)*, Technischer Bericht FAIR Technical Design Report, Darmstadt (2018).

[16] M. Durante, P. Indelicato, B. Jonson, V. Koch, K. Langanke, U.-G. Meißner, E. Nappi, T. Nilsson, T. Stöhlker, E. Widmann, und M. Wiescher, *All the fun of the FAIR: fundamental physics at the facility for antiproton and ion research*, Physica Scripta, 94, 033001 (2019).

[17] A. A. Weber, *Development of readout electronics for the rich detector in the hades and cbm experiments - hades rich upgrade, mrich detector construction and analysis* -, Doktorarbeit (2021).

[18] J. Heuser, W. Müller, V. Pugatch, P. Senger, C. J. Schmidt, C. Sturm, und U. Frankenfeld (Herausgeber) *[GSI Report 2013-4] Technical Design Report for the CBM Silicon Tracking System (STS)*, GSI, Darmstadt, 2013.

[19] *The detector and data processing system*, `https://www.cbm.gsi.de/detectors`, accessed: 2022-08-13.

[20] *Cbm collaboration, common plot collection*, `https://docs.google.com/presentation/d/1HiQglfqf3mMmJQ-F_ibG2P-S25vZoogEBNs8fBI3p0U/edit#slide=id.gc35a9bbe60_2_445`, accessed: 2022-06-23.

[21] N. Memon, S. B. Patel, und D. P. Patel, *Comparative analysis of artificial neural network and xgboost algorithm for polsar image classification*, in: *Pattern Recognition and Machine Intelligence*, (Herausgegeben von B. Deka, P. Maji, S. Mitra, D. K. Bhattacharyya, P. K. Bora, und S. K. Pal), 452–460, Springer International Publishing, Cham, 2019.

[22] cmglee und MartinThoma, *ROC curve*, `https://en.wikipedia.org/wiki/Receiver_operating_characteristic`, accessed: 2022-08-13.

[23] *Pair analysis programs with python*, `https://drive.google.com/drive/folders/1Ve0VA_3jnKqsCddokyh5GD99l570uu4c?usp=sharing`.

[24] *hipe4ml documentation*, `https://hipe4ml.github.io/`, accessed: 2022-08-15.

[25] *Shap documentation*, `https://shap.readthedocs.io/en/latest/index.html`, accessed: 2022-07-20.

[26] T. Galatyuk, *Di-electron spectroscopy in hades and cbm: from p + p and n + p collisions at gsi to au + au collisions at fair*, Doktorarbeit (2009).

[27] H. H Gutbrod et al., *FAIR - baseline technical report. executive summary.*

Eidesstattliche Erklärung

Hiermit versichere ich, Henrik Schiller, dass die vorliegende Arbeit mit dem Titel *Application of Machine Learning to Particle Identification for Dielectron Analysis in CBM* selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommenen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Münster, den 24. August 2022

.....
Henrik Schiller