JOINT INSTITUTE FOR NUCLEAR RESEARCH
Veksler and Baldin laboratory of High Energy Physics

Final Report of INTEREST Program

Improving electron identification performance of MPD detector using various Machine Learning techniques in ROOT framework.

Supervisor:
Dr Sudhir Pandurang Rode
Veksler and Baldin Laboratory of High Energy Physics
sudhirrode11@gmail.com

Student:
Thalia Rodríguez Martínez, Cuba
Higher Institute of Applied Sciences and Technologies, University of Havana
thaliarodmar@gmail.com

Participation period:
March 03 – April 12, Wave 12

Dubna, 2025

Index

Abstract

The Multi-Purpose Detector (MPD) experiment at the NICA accelerator complex aims to explore the properties of quark-gluon plasma (QGP), a state of matter believed to have existed shortly after the Big Bang. This study focuses on enhancing electron identification capabilities within heavy-ion collisions using machine learning techniques. To achieve this, we utilized Monte Carlo simulations to generate and reconstruct collision events, employing sub-detectors such as the Time Projection Chamber (TPC), Time of Flight (TOF), and Electromagnetic Calorimeter (ECal) to gather critical data. Various classification algorithms, including Boosted Decision Trees and Multilayer Perceptrons, were applied to distinguish between electron signals and hadronic background.The results demonstrated that the Boosted Decision Tree (BDTG) algorithm, with a cut value of 0.85, provided the best balance between efficiency and purity in electron identification. Most methods showed high efficiency and purity across a wide momentum range, although performance varied at higher momenta. These findings underscore the importance of optimizing particle identification techniques, which are vital for accurate data analysis in high-energy nuclear physics.

Keywords

Quark-gluon plasma, Multi-Purpose Detector, electron identification, machine learning, heavy-ion collisions, Monte Carlo simulations.

1. Introduction

1.1 Introduction to High Energy and Heavy-Ion Physics

High energy and heavy-ion physics represents a crucial area of study within the broader field of physics, concentrating on the behavior of matter under extreme conditions. By simulating environments of exceptionally high temperatures and densities, akin to those found in the early universe shortly after the Big Bang, scientists aspire to gain insights into Quantum Chromodynamics (QCD). This theoretical framework describes the strong interactions that dictate the dynamics of quarks and gluons.

In these extreme scenarios, conventional hadronic matter transitions into a unique state known as quark-gluon plasma (QGP). In this state, quarks and gluons are liberated from their confinement within hadrons, allowing for interactions that reveal fundamental truths about matter and the forces shaping our universe. Investigating QGP and its associated phenomena is essential for understanding not only the fundamental forces of nature but also the structure of matter itself and the processes that have led to the evolution of the universe.

Research in this domain also focuses on critical aspects such as QCD phase transitions, the collective behavior of particles in dense nuclear environments, and the intricate mechanisms of particle production and decay resulting from high-energy collisions. These goals are pursued using sophisticated large-scale particle accelerators, which facilitate collisions at ultra-relativistic speeds, generating the extreme conditions necessary for observing these fascinating phenomena.

By exploring these areas, high energy and heavy-ion physics aims to deepen our comprehension of the universe’s fundamental properties, while also shedding light on the intricate processes that underlie the fabric of matter.

1.2 Overview of the Multi-Purpose Detector (MPD) Experiment

The Multi-Purpose Detector (MPD) experiment is a pivotal initiative in the field of high-energy physics, specifically designed to explore the properties of quark-gluon plasma (QGP) and the fundamental interactions that govern matter under extreme conditions. Located at the NICA (Nuclotron-based Ion Collider fAcility) facility in Dubna, Russia, the MPD aims to investigate the phase transitions of nuclear matter and to enhance our understanding of Quantum Chromodynamics (QCD).

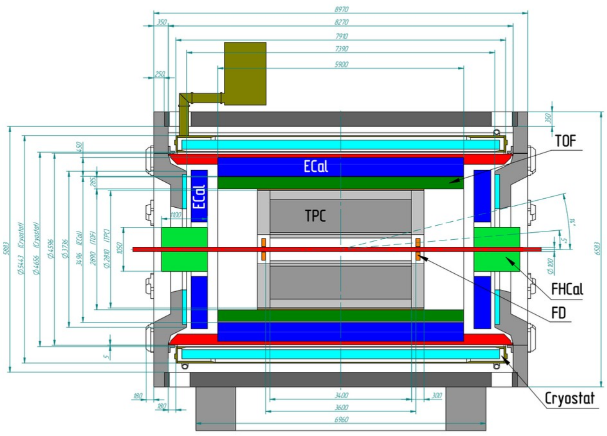


Figure 1. The overall schematic of the MPD subsystems in the first stage of operation (Stage 1) - cross-section by the vertical plane.

At the core of the MPD's mission is the ability to study heavy-ion collisions at unprecedented energies, which allow researchers to recreate the high-density environment similar to that of the early universe. These collisions provide a unique opportunity to observe the behavior of subatomic particles in the QGP state, where quarks and gluons are no longer confined within hadrons. By analyzing the resulting particle interactions, the MPD seeks to elucidate the properties of this exotic state of matter, including its temperature, density, and collective behavior.

Through its multifaceted approach, the MPD experiment aspires to answer fundamental questions regarding the nature of QGP, the mechanisms of particle production, and the underlying dynamics of strong interactions. The insights gained from this research will significantly contribute to our understanding of the universe's evolution and the fundamental forces that govern its behavior, positioning the MPD as a key player in the ongoing exploration of high-energy nuclear physics.

The MPD focuses on several key objectives related to heavy-ion collision studies. First, it aims to investigate hadrochemistry by analyzing production ratios of various hadrons to understand QCD phase transitions, particularly deconfinement and chiral symmetry restoration, and to map the QCD phase diagram in the baryon-rich region.

Another significant goal is to measure anisotropic flow in heavy-ion collisions, which will provide insights into the collective dynamics of the quark-gluon plasma (QGP) and its transport properties, informing the equation of state (EoS) of the QGP under varying energy densities. Additionally, intensity interferometry, or femtoscopy, will be employed to extract spatial and temporal characteristics of the particle-emitting source, crucial for understanding the system's evolution.

The MPD will also examine fluctuations of conserved quantities like baryon number and strangeness to identify critical phenomena and potentially locate a critical end point (CEP) in the QCD phase diagram. Furthermore, studying short-lived resonances and electromagnetic probes, such as dileptons and direct photons, will enable the investigation of hadronic medium properties and thermal radiation from the QGP, enriching the understanding of early collision dynamics and QGP formation.

One more physics topic that will be studied in MPD, is dileptons, particularly dielectrons (e^+e^- pairs), represent crucial electromagnetic probes as they do not strongly interact with the medium and thus carry unaltered information from all stages of the collision. The precise identification of electrons is fundamental for these analyses, as it allows for the reconstruction of the invariant mass spectrum of dielectrons, revealing modifications of vector resonances in the medium and thermal radiation from the QGP. Studies of low-mass dielectrons provide information about chiral symmetry restoration, while those in the intermediate mass range offer direct access to the effective temperature of the plasma during its evolution

By concentrating on these objectives, the MPD experiment seeks to advance knowledge in heavy-ion physics and explore the QCD phase diagram, particularly regarding the onset of deconfinement and chiral symmetry restoration.

1.3 MPD Sub-Detectors

The Multi-Purpose Detector (MPD) is equipped with a sophisticated array of sub-detectors, including a Time Projection Chamber (TPC), Time of Flight (TOF) detector, and Electromagnetic Calorimeter (ECal). Each of these components plays a crucial role in the accurate identification and measurement of particles produced in heavy-ion collisions, facilitating a comprehensive analysis of the collision dynamics.

1.3.1 Time Projection Chamber (TPC)

The TPC serves as the primary tracking detector within the MPD framework. It is designed to provide three-dimensional tracking of charged particles through the measurement of ionization energy loss (dE/dx) as particles traverse the gas volume. With a length of 340 cm and a large diameter, the TPC enables the reconstruction of particle trajectories with high spatial resolution. The uniform electric field within the chamber facilitates the drift of ionization electrons to the readout chambers, allowing for precise momentum measurements.

1.3.2 Time of Flight (TOF) Detector

The TOF detector complements the TPC by providing critical timing information necessary for particle identification. Utilizing Multi-gap Resistive Plate Chambers (MRPCs), the TOF achieves time resolutions of approximately 80 ps. By measuring the time it takes for a particle to travel from the interaction point to the TOF, alongside its momentum data from the TPC, the system can distinguish between different particle species, particularly within the intermediate momentum range.

1.3.3 Electromagnetic Calorimeter (ECal)

The ECal plays a vital role in identifying electromagnetic probes, including electrons and photons. Constructed from lead-scintillator sandwiches, the ECal is designed to measure the energy of electromagnetic showers with high precision. It operates within the MPD's magnetic field, allowing for the detection of particles over a wide energy range. The ECal enhances the electron identification process through the measurement of the energy-to-momentum ratio (E/p), which is expected to be approximately 1 for electrons.

1.3.4 Integrated Operation for Electron Identification

The integration of these three detectors is key to the successful identification of electrons. The TPC provides detailed tracking and momentum information, while the TOF offers timing data that, when combined, allow for the reliable identification of charged particles. The ECal further refines this identification by measuring the energy of the detected electrons.

Electrons are identified through a multi-step process: first, candidates are selected based on their tracking information in the TPC, followed by timing cuts from the TOF, and finally confirmed through energy measurements in the ECal. This synergistic approach enhances the overall efficiency and purity of electron identification, enabling the MPD to provide valuable insights into the dynamics of heavy-ion collisions.

Moreover, the MPD experiment is at the forefront of integrating advanced data analysis techniques, including machine learning algorithms, to enhance particle identification and extraction of physical observables. Specifically, this work focuses on improving electron identification performance of the MPD detector using various machine learning techniques within the ROOT framework. These innovations not only improve the efficiency and accuracy of data interpretation but also pave the way for new discoveries in the field.

2. Project Objectives

This work aims to enhance the electron identification capabilities of the Multi-Purpose Detector (MPD) experiment at the NICA accelerator complex, with a primary focus on studying quark-gluon plasma (QGP) in heavy-ion collisions. The project seeks to implement advanced machine learning techniques to improve the accuracy and efficiency of distinguishing electron signals from hadronic background. By utilizing Monte Carlo simulations, the work generates and reconstructs collision events, creating a robust dataset for comprehensive analysis. Furthermore, it evaluates the performance of various classification algorithms, including Boosted Decision Trees and Multilayer Perceptrons, to assess their effectiveness in identifying electrons across different momentum ranges. The study also aims to determine optimal cut values that achieve a balance between efficiency and purity in electron identification.

3. Methodology

The investigation begins with the simulation and reconstruction of heavy-ion collision events, this process involves three main stages: event generation, particle transport simulation, and reconstruction of the detector signals. These stages are outlined below, focusing on their physical and computational aspects, which ultimately enable the study of particle interactions within the MPD (Multi-Purpose Detector). The primary goal is to utilize the TPC (Time Projection Chamber), TOF (Time-of-Flight) system, and ECal (Electromagnetic Calorimeter) for reconstructing particle trajectories, identifying particles, and measuring their properties. These steps are crucial for generating the data required for later analysis using machine learning techniques, which directly supports the ultimate objective of this work: to improve electron identification performance in the MPD detector through the application of various machine learning techniques within the ROOT framework.

Event generation is performed using the UrQMD (Ultra-relativistic Quantum Molecular Dynamics) model, which describes hadron interactions in detail based on principles of quantum chromodynamics (QCD). In this stage, fundamental parameters such as the properties of the colliding nuclei (xenon with a mass number of 124 and an atomic number of 54), the center of mass energy (7 GeV), the impact parameter (-14), and the total number of simulated events (1,000) are defined. Additionally, specific configurations are adjusted, such as deactivating certain particle decay channels, allowing for a detailed focus on key physical processes. The generated events provide the initial conditions required for subsequent stages.

In the simulation stage, the particles produced during the generated events are transported through the geometry of the MPD (Multi-Purpose Detector) using the MPDROOT framework. This process employs the Virtual Monte Carlo (VMC) engine with Geant4 as the transport model, considering phenomena such as energy loss, multiple scattering, and secondary particle generation. The geometry includes main components such as the Time Projection Chamber (TPC), the Time-of-Flight (TOF) system, and the Electromagnetic Calorimeter (ECal), which are essential for particle tracking, time measurement, and energy deposition detection. Additionally, a uniform magnetic field of 5 kG is configured to allow the reconstruction of charged particle momenta from the curvature of their trajectories. The results of this stage include detailed data on the detector's response, stored in ROOT files, which are used as input for the reconstruction stage.

The final stage involves reconstructing the raw detector signals into significant physical observables, such as trajectories, momenta, and particle identification. The TPC data is processed with clustering and tracking algorithms, such as the Maximum Likelihood Estimation Method (MLEM) and the Kalman filter, to accurately identify the trajectories and momenta of the particles.

3.1 Data preparation for Machine Learning training

The process of preparing data for machine learning training involves extracting and analyzing specific parameters from reconstructed heavy-ion collision events. These parameters include ***DCAx, DCAy, DCAz***, ***dPhi_TOF***, ***dz_TOF***, ***dPhi_ECal***, *and* ***dz_ECal***, which provide critical physical insights into the particle trajectories and interactions within the Multi-Purpose Detector (MPD).

These variables were represented through histograms to scrutinize their distributions. Gaussian function fits were employed on these distributions to extract the mean (μ) and standard deviation (σ).These parameters were expressed as functions of momentum using analytical approaches, including polynomial forms.

These fitted parameters are crucial for building the trees that capture important information about particle interactions and tracks. This helps extract key features needed to improve the predictive power of machine learning algorithms, aiming to enhance particle identification in the MPD experiment.

3.1.1 Distance of Closest Approach (DCA)

The Figure 2 illustrate the relationship between the Distance of Closest Approach (DCA) in the X, Y, and Z dimensions versus the momentum (p) of particles. The mean DCA values reveal how the average proximity of particle tracks to the primary vertex varies with increasing momentum, showcasing a characteristic trend where the mean DCA diminishes as momentum rises, particularly notable in the Y and Z dimensions. This trend suggests that higher momentum particles tend to have trajectories that are closer to the collision point. Conversely, the sigma DCA plots indicate the spread or uncertainty of the DCA measurements. A decreasing trend is observed, indicating that as particle momentum increases, the precision of the DCA measurements improves.

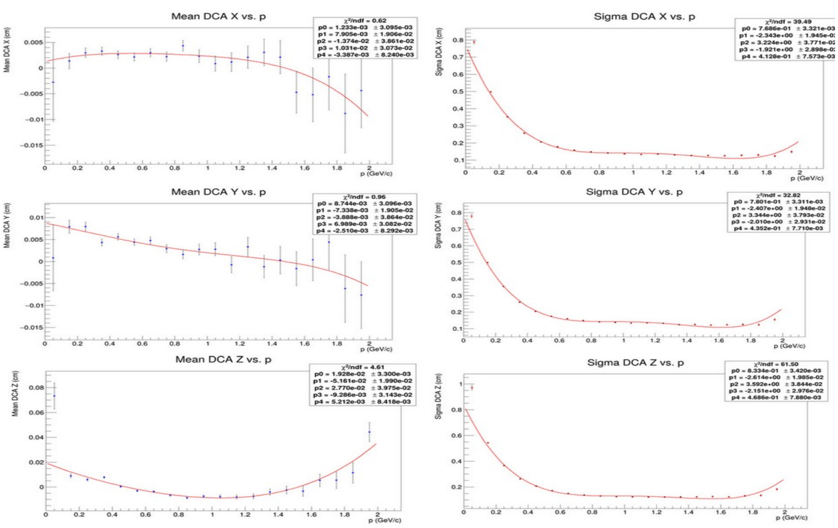


Figure 2. Dependence of Mean and Sigma DCA on Particle Momentum in X, Y, and Z Directions.

3.1.2 Matching in the Time of Flight (TOF) Detector

Figure 3 shows the relationship between the mean and sigma values for matching tracks in the Time of Flight (TOF) Detector as a function of particle momentum (p).

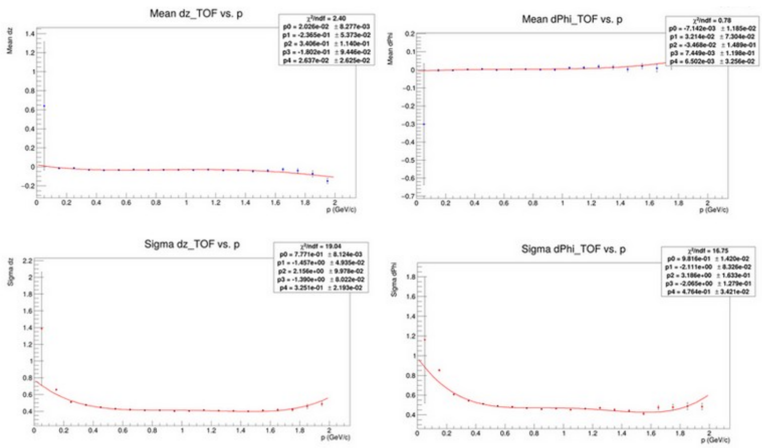


Figure 3. Analysis of Track Matching in TOF.

The mean dz (difference in longitudinal position) remains relatively stable across different momentum levels (p). In contrast, the mean dPhi (difference in azimuthal angle) shows a slight variation with momentum, but it does not exhibit significant fluctuations.

The sigma dz plot demonstrates low uncertainty in the longitudinal position measurements, reinforcing the reliability of the TOF detector's timing data. This low uncertainty is crucial for ensuring that particle trajectories are accurately reconstructed. On the other hand, the sigma dPhi shows a decreasing trend as momentum increases. This indicates that the measurements of the azimuthal angle become more precise at higher momenta, enhancing the overall accuracy of particle identification. The dphi distributions in the TOf matching are expected to exhibit charge dependence in their mean, which is not considered in this study.

3.1.3 Matching in the Electromagnetic Calorimeter (ECal)

Figure 4 shows the relationship between the mean and sigma values for matching tracks in the Electromagnetic Calorimeter (ECal) as a function of particle momentum (p).

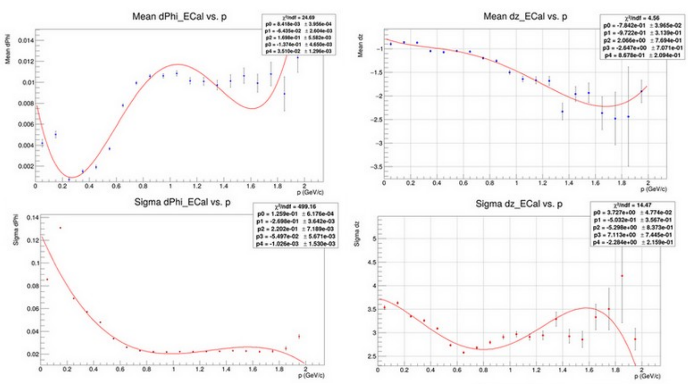


Figure 4. Analysis of Track Matching in Ecal.

The mean dPhi (difference in azimuthal angle) shows a non-linear relationship with momentum (p). This indicates that as the momentum of particles increases, the differences between their expected and actual angles change, reflecting variations in particle behavior at different energies. Additionally, the mean dz (difference in longitudinal position) decreases with increasing momentum, suggesting that higher momentum particles tend to be closer to their expected positions relative to the collision point.

The sigma dPhi plot reveals that the uncertainty in azimuthal matching decreases as momentum increases. This means that the measurements become more reliable at higher momentum levels. In contrast, the sigma dz exhibits fluctuating behavior, indicating that while there is an overall trend of improved consistency in some regions, there are also moments where the uncertainty increases at specific momentum ranges. In the Ecal matching, the dphi distributions are anticipated to show charge dependence in their mean, a factor that is not addressed in this study.

3.2 Machine Learning Implementation for Electron Identification

The implementation of machine learning techniques for electron identification within the MPD (Multi-Purpose Detector) utilized the TMVA (Toolkit for Multivariate Analysis) framework integrated with ROOT.

The training process was designed to configure and simultaneously train multiple classification algorithms, aiming to effectively distinguish between signal (electrons) and background (non-electron particles) based on a variety of detector response parameters. The primary input variables for the classification algorithms included reconstructed track momentum (p), energy loss measured by the Time Projection Chamber (dedx), the total number of hits recorded in the detector (NHits), the ratio of energy to momentum (Ebyp), time-of-flight beta calculated using ECal information (Tofbeta_ECal), measured time-of-flight beta (Tofbeta), pseudorapidity of the reconstructed track (pseudorapid), and azimuthal angle (Phi).

The training process incorporated particle selection criteria that applied quality cuts on track parameters, including Distance of Closest Approach (DCA), matching variables, the number of hits, and relevant kinematic properties. This rigorous selection ensured that only well-reconstructed tracks were utilized for training.

3.2.1 Implemented Methods

- **Probability-Based Classifiers**

The ***Likelihood*** Method approach uses a naive Bayesian estimator to model the probability distributions of input variables for different classes, such as electrons and non-electrons. This method is effective because it captures the unique characteristics of variables like dE/dx and E/p, which are important for identifying electrons. ***K-Nearest Neighbors (KNN)*** classifies particles by comparing them to similar training samples nearby. Using 20 neighbors and a Gaussian kernel is a good choice for particle physics, where the parameters often cluster in the feature space. This helps in making accurate classifications based on proximity.

- **Neural Networks**

The ***Multilayer Perceptron (MLP)*** uses hyperbolic tangent activation functions and hidden layers that match the number of input variables. This configuration is effective for capturing nonlinear relationships between key variables, such as momentum, dE/dx, and E/p. These relationships are crucial for distinguishing electrons from other types of particles.

- **Boosted Decision Trees**

The ***Boosted Decision Tree (BDT)*** method uses an adaptive approach with 850 trees and a maximum depth of 3. This is particularly effective for particle identification, as it can manage complex correlations between variables. The ***Gradient Boosted Decision Tree (BDTG)*** is a variant that employs 1000 trees and focuses on gradually improving classification accuracy. It iteratively reduces classification errors, especially in difficult areas of the particle phase space, making it a powerful tool for enhancing particle identification.

These methods are highly relevant because they effectively capture complex correlations among variables, handle non-Gaussian distributions, and ensure a balanced representation of signal and background through appropriate data preparation, preventing bias in classification.

The presented graphs in Figure 5 illustrate the overfitting check for the classifiers that we used (BDT, BDTG, KNN, Likelihood, and MLP) aimed at distinguishing signals from background.

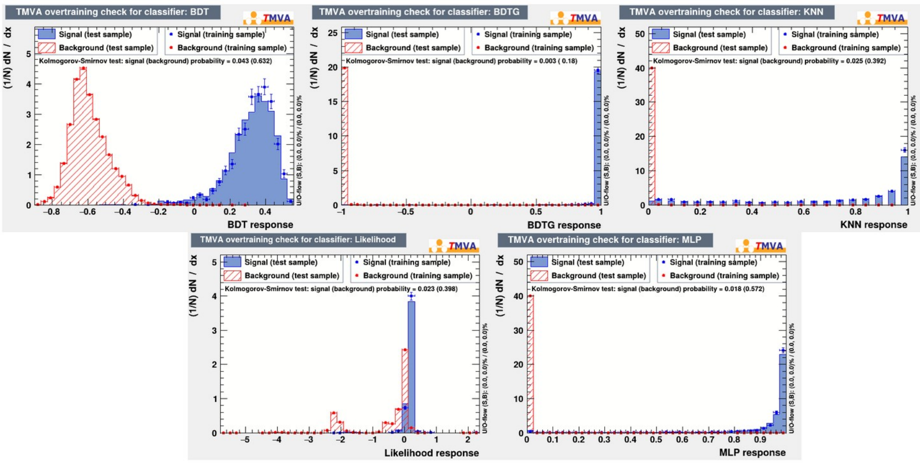


Figure 5. Overfitting check for different classification methods. The graphs display the response distributions of signal and background for BDT, BDTG, KNN, Likelihood, and MLP classifications, along with the results of the Kolmogorov-Smirnov test quantifying the difference between both distributions.

Overtraining, occurs when a model learns the training data too well, including noise and outliers, which can lead to poor performance on new, unseen data. This can result in a model that appears accurate during training but fails to generalize.

The analysis of the classifiers reveals that those with Kolmogorov-Smirnov (K-S) test probability values less than 0.01 may be experiencing overtraining. For example, the BDTG, with a K-S of

0.003, indicates an excessive fit to the training data. Other models, such as KNN and the Likelihood classifier, have K-S values greater than 0.01.

In summary, the BDTG model shows signs of overtraining, which could affect its ability to generalize.

The analysis of the ROC (Receiver Operating Characteristic) curve reveals important insights into the performance of the classification methods in balancing signal efficiency and background rejection, the information obtain form our training it is in Figure 6.

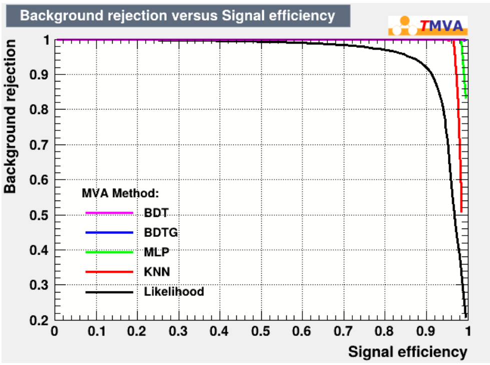


Figure 6. Background rejection versus signal efficiency for various classification methods. The graph illustrates the trade-off between the efficiency of signal detection and the rejection of background events across BDT, BDTG, MLP, KNN, and Likelihood techniques.

The Likelihood classifier (black line) demonstrates the best overall performance, achieving exceptionally high background rejection (greater than 0.9) until a signal efficiency of approximately 0.85, after which its performance declines rapidly. Both the BDTG and BDT classifiers (blue lines) also show excellent performance at high signal efficiencies, though they do not quite match the Likelihood classifier in this critical region. The MLP classifier (green line) performs well but exhibits slightly lower background rejection, particularly in the mid-range of signal efficiencies. In contrast, the KNN classifier (red line) ranks as the weakest, with significantly lower background rejection capabilities across most of the signal efficiency spectrum.

Overall, these findings suggest that while the Likelihood classifier provides the best balance between signal detection and background rejection, it is essential to address the overfitting concerns identified earlier, potentially through regularization techniques, cross-validation, or parameter tuning, before implementation.

Once the data preparation stage and the implementation of machine learning techniques for electron identification are complete, the next step involves analyzing efficiency and purity across different cuts and models. This analysis will generate multipanel plots that illustrate how efficiency and purity behave with each cut and model applied. In the results section, these plots will be described, highlighting the variations in efficiency and purity. From this analysis, the best cuts from each model will be selected and compared, enabling us to identify the most effective configurations for electron identification within the context of the MPD experiment.

4. Results

4.1 Efficiency Estimation

The Figure 7 illustrates the efficiency of electron identification as a function of momentum (p) in GeV/c, employing various classification methods: MLP, BDT, BDTG, KNN, and Likelihood, with different cut values.

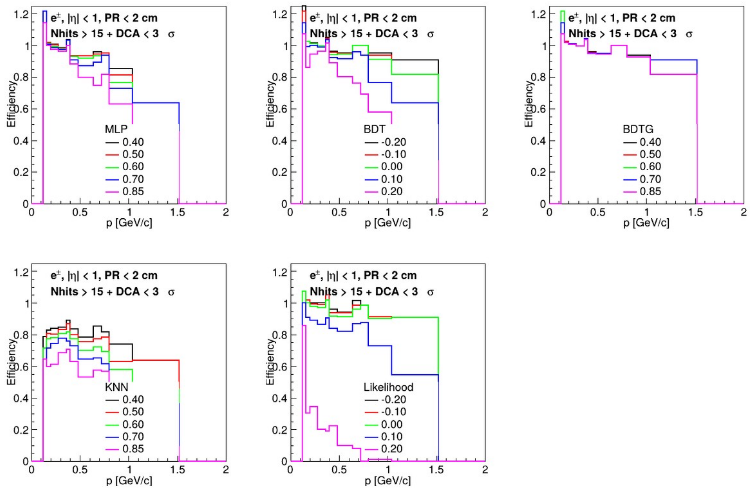


Figure 7. Efficiency of electron identification versus momentum (p) in GeV/c for classification methods: MLP, BDT, BDTG, KNN, and Likelihood. Each plot displays the efficiency trends across various cut values, emphasizing the comparative performance of each method across low and high momentum ranges.

Most methods exhibit high efficiency (0.8-1.0) in low momentum ranges (0-1 GeV/c), with a notable decrease in efficiency at higher momenta (>1 GeV/c). BDTG, particularly with a cut of 0.85, maintains the highest efficiency across nearly the entire momentum range, especially at higher momenta where other methods decline more rapidly. BDT and MLP show similar patterns, demonstrating good efficiency up to approximately 1 GeV/c before starting to decline. KNN exhibits moderate performance, with more pronounced fluctuations throughout the momentum spectrum. In contrast, Likelihood shows significantly poorer performance, especially with cuts ≥0.10, experiencing a steep drop in efficiency at higher momenta.

4.2 Purity Estimation

The Figure 8 illustrates the purity of electron identification across the same momentum range and classification methods. Purity is a critical metric for assessing the quality of electron identification, as it indicates the proportion of correctly identified electrons among all identified particles.

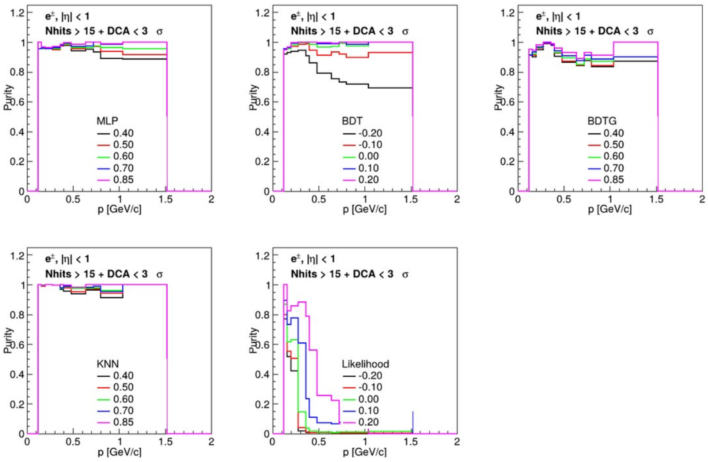


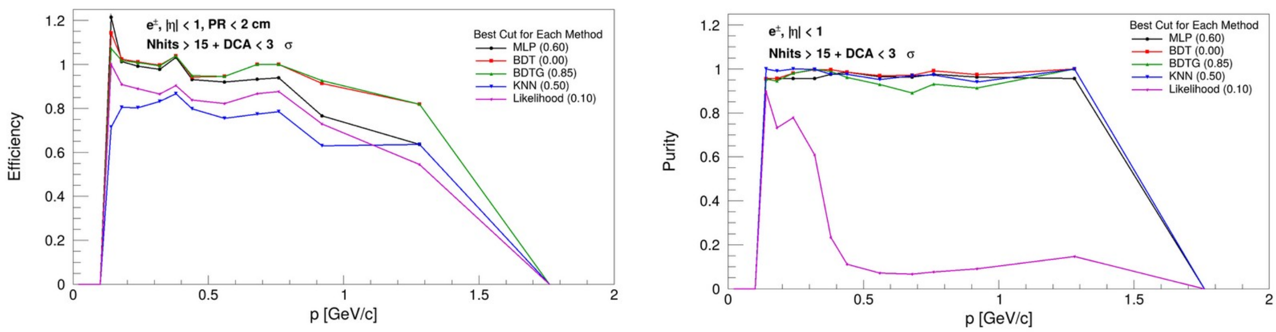
Figure 8. Purity of electron identification versus momentum (p) in GeV/c for classification methods: MLP, BDT, BDTG, KNN, and Likelihood.

Most methods maintain very high purity (>0.9) across a wide momentum range (0.2-1.5 GeV/c). This consistency suggests that the classification methods are effective in distinguishing electrons from other particles within this momentum interval. MLP, BDT, BDTG, and KNN all demonstrate excellent purity, ranging from approximately 0.95 to 1.0 throughout most of the momentum range. In contrast, the Likelihood method exhibits markedly different behavior, achieving high purity only at very low momenta (0.1-0.3 GeV/c) before experiencing a dramatic decline.

Unlike efficiency, purity remains remarkably consistent across different cut values for MLP, BDT, BDTG, and KNN. This indicates that these methods robustly separate electrons from hadronic background, regardless of adjustments to the threshold settings.

4.3 Comparison of Best Cuts

The bottom two graphs provide a comparison of the best-performing cut values for each classification method in terms of efficiency and purity. This analysis highlights the strengths and weaknesses of each method across the momentum spectrum.



(a) (b)

Figure 9. Comparison of the best-performing cut values for each classification method in terms of efficiency (a) and purity (b).

4.3.1 Efficiency Comparison

BDTG with a cut value of 0.85 and BDT with a cut of 0.00 clearly outperform the other methods across the entire momentum range. MLP, with a cut of 0.60, follows closely behind, maintaining good efficiency up to approximately 1 GeV/c. In contrast, KNN (0.50) and Likelihood (0.10) exhibit lower efficiency, particularly at mid to high momenta, indicating their limitations in effective electron identification.

4.3.2 Purity Comparison

In terms of purity, MLP (0.60), BDT (0.00), BDTG (0.85), and KNN (0.50) all sustain excellent purity levels (>0.95) across most of the momentum range. However, Likelihood (0.10) demonstrates dramatically lower purity, except at very low momenta, making it generally unsuitable for effective electron identification despite its strong background rejection capabilities noted in the ROC curve analysis.

The BDTG algorithm, with a cut value of 0.85, appears to be a strong option for electron identification, effectively balancing efficiency and purity across a wide range of momentum. Its consistent performance suggests it is well-suited for the MPD experiment.

In the performance comparison, BDTG is the top performer, followed by BDT, which also maintains a solid balance. The MLP shows good results but has slightly lower efficiency at higher momenta. KNN achieves decent purity but struggles with efficiency, particularly in mid to high momentum ranges. The Likelihood method tends to underperform across various cuts, showing limitations in both purity and efficiency.

All methods demonstrate good performance between 0.2-1.0 GeV/c, but identifying electrons becomes more difficult above 1.5 GeV/c, highlighting the need for further research to improve capabilities in this area. The results illustrate the trade-off between efficiency and purity in particle identification; stricter cut values tend to enhance purity at the expense of efficiency. However, BDTG manages to maintain a commendable level of both, making it a useful tool for electron identification without significant compromises.

5. Conclusion

In this report, we explored how to improve electron identification in the MPD experiment using modern machine learning techniques. Through simulations and the use of different algorithms, the goal was to distinguish electron signals more efficiently and accurately from the many other particles produced in heavy-ion collisions.

The results show that decision tree-based methods, especially the BDTG algorithm, offer a good balance between efficiency (the ability to find as many electrons as possible) and purity (ensuring that the identified electrons are truly electrons and not other particles). However, it was also observed that as the particle momentum increases, correctly identifying electrons becomes more difficult, suggesting that there is still room to improve the techniques used.

Overall, this study confirms that the use of machine learning tools can bring significant improvements to high-energy physics experiments, making it possible to analyze large volumes of data more precisely. Nevertheless, it is important to be cautious and to keep refining the methods, as each algorithm has advantages and limitations that must be considered according to the experimental context.

Acknowledgements

I sincerely thank Professor Sudhir Pandurang for believing in me and guiding me during these seven weeks in a project that I initially did not fully understand. He was able to explain it clearly and help me grasp the concepts. I greatly appreciate his attention and willingness to assist me at any time, despite our time zone differences. He taught me that we can make mistakes; what matters is knowing where the error lies, fixing it, and moving forward.

Additionally, I would like to thank once again the staff at the University Center, especially the INTEREST team, who kindly kept us informed through their emails. On a personal note, I am grateful for the opportunity to once again be close to Dubna, which has taught me so much over the past two years.

References

1. Status and initial physics performance studies of the MPD experiment at NICA.
2. Relativistic Kinematics by R. Sahoo.
3. Three Stages of the NICA Accelerator Complex.
4. MPD ROOT Guide. Available at: MPD ROOT Guide.
5. TMVA Users Guide, Toolkit for Multivariate Data Analysis with ROOT.