

**SEARCH FOR PENTAQUARK STATES IN  
AU+AU COLLISIONS AT  $\sqrt{s_{NN}} = 39$  GeV  
WITH THE STAR EXPERIMENT AT RHIC  
USING MACHINE LEARNING  
TECHNIQUES**

*By*

**Saket Kumar Sahu**

*A dissertation submitted for the partial fulfilment of*

**BS-MS Dual Degree**

in

**Science**



*to*

**INDIAN INSTITUTE OF SCIENCE EDUCATION  
AND RESEARCH  
BERHAMPUR - 760 010, INDIA**

*May 31, 2022*

## DECLARATION

The work presented in this dissertation has been carried out by me under the guidance of Prof. Sonia Kabana at Instituto de Alta Investigación, Universidad de Tarapacá, Arica and co-guidance of Dr. Md Nasim at the Indian Institute of Science Education and Research, Berhampur.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.



**Saket Kumar Sahu**

May 31,2022

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.



**Prof. Sonja Kabana** (supervisor)

Date: May 31, 2022

In my capacity as the co-supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

**Dr. Md. Nasim** (co-supervisor)

Date: May 31, 2022

## **UNDERTAKING**

I, Saket Kumar Sahu, hereby declare that the thesis titled, **SEARCH FOR PENTAQUARK STATES IN AU+AU COLLISIONS AT  $\sqrt{s_{NN}} = 39 \text{ GeV}$  WITH THE STAR EXPERIMENT AT RHIC USING MACHINE LEARNING TECHNIQUES**, submitted as a partial requirement for the completion of the BS-MS dual degree programme at IISER Berhampur, is based on my original research/data/experiments and has not been copied/taken verbatim from anyone or from any other sources. I further certify that this thesis has been checked for plagiarism through a plagiarism detection tool i.e. Ouriginal, approved by the institute. I am aware of the regulations on prevention of plagiarism of University Grant Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions Regulation, 2018). I also undertake that if at any stage the institute notices any plagiarism or any other discrepancies in this thesis of mine, I will take full responsibility and abide by whatsoever action taken against me by the institute, as deemed necessary.



**Signature of the Student**

Name: Saket Kumar Sahu

Roll No. 17086

Date May 31,2022

## CERTIFICATE

This is to certify that the work contained in this project report entitled "**SEARCH FOR PENTAQUARK STATES IN AU+AU COLLISIONS AT  $\sqrt{s_{NN}} = 39 \text{ GeV}$  WITH THE STAR EXPERIMENT AT RHIC USING MACHINE LEARNING TECHNIQUES**" submitted by **Saket Kumar Sahu (Roll No: 17086)** to Indian Institute of Science Education and Research Berhampur towards partial requirement of **BS-MS Dual Degree** in Physics has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.



Prof. Sonja Kabana  
(supervisor)

(FIC/HOD)

Dr. Md Nasim  
(co-supervisor)

Dr. Natasha Sharma  
(Examiner)

(Convener,DUGC)

Date: May 31,2022

*Dedicated to my friends and family*

## ACKNOWLEDGEMENT

In my 5 years of BS-MS I have had the privilege of knowing many wonderful people who will always remain an integral part of my life. Here I thank only a few of them. I would like to express my deepest gratitude to my supervisor, Prof. Sonja Kabana for providing me with the opportunity to be part of her group and carry on research in this exciting area of Physics. Her valuable guidance and constant support through online discussions made this journey a smooth sailing one. Her guidance taught me how to carry out collaborative research and how to have the mindset and patience required for research.

I would like to thank Dr. Vipul Bairathi for being patient with me and clearing all my doubts. During this one year I was able to learn a lot from him especially regarding programming and analysis techniques. I would like to thank my co-supervisor Dr. Md Nasim who encouraged me to pursue this project and whose classes helped me clear my concepts and have a better understanding of the physics. I am very grateful to Dr. Ujjal Kumar Dey whose QFT and Particle Physics classes were the main reason for peaking my interest in this area of research. The discussions with him broadened my understanding of Physics. I have never met any professor who answers mails faster than him (even at 4am).

I would like to thank Mr. Haris Avudaiyappan M for his advices regarding this project and my research career. I would like to express my heartfelt gratitude to my friends (The third floor boys) who have become an integral part of my life. I cannot imagine how I would have endured these five years without them.

I would like to thank my father whose constant curiosity about my research

work kept me motivated throughout this journey. I would like to thank my mother who was always there to take care of me and supporting me. I would like to thank my brother who was always ready to have discussion with me on any topic and encouraged me throughout my life. I am forever in debt to them.

Berhampur - 760 010

**Saket Kumar Sahu**

Date: May 31,2022

# ABSTRACT

In this thesis we present the preliminary results of the analysis of Au+Au  $\sqrt{s_{NN}} = 39$  GeV data with the THE STAR detector at RHIC for the search of pentaquark candidates in the  $\Lambda K_s^0$  decay channel. We have used the KF-Particle code to reconstruct the  $V^0$  particles required for this analysis. The KFParticle code is based on Kalman Filter algorithm for reconstructing short lived particles formed during heavy ion collisions. Boosted Decision Trees (BDT) and Multi Layer Perceptron with Bayesian Regulator (MLPBNN) classifiers of TMVA package of ROOT are trained using signal and background trees to help classify signal from background. The invariant mass of  $\Lambda K_s^0$  are plotted after TMVA analysis to identify the signals that correspond to the pentaquark candidate.

**Keywords:** Heavy Ion Collisions, Pentaquark, STAR, KFParticle, TMVA, BDT, MLPBNN, Machine Learning, Neural Network

# Contents

|   |              |
|---|--------------|
| <b>List of Figures</b>                        | <b>xi</b>    |
| <b>List of Tables</b>                         | <b>xviii</b> |
| <b>1 Introduction</b>                         | <b>1</b>     |
| 1.1 Standard Model . . . . .                  | 1            |
| 1.2 Quantum Chromodynamics . . . . .          | 4            |
| 1.3 Quark Gluon Plasma . . . . .              | 5            |
| 1.4 $V^0$ particles and decays . . . . .      | 8            |
| 1.5 Pentaquark . . . . .                      | 9            |
| 1.6 Thesis Motivation . . . . .               | 11           |
| <b>2 Experimental Setup</b>                   | <b>13</b>    |
| 2.1 Relativistic Heavy Ion Collider . . . . . | 13           |

|          |  |           |
|----------|--|-----------|
| 2.2      | Solenoidal Tracker at RHIC . . . . .         | 16        |
| 2.2.1    | Time Projection Chamber . . . . .            | 17        |
| 2.2.2    | Time Of Flight . . . . .                     | 18        |
| <b>3</b> | <b>Kalman Filter Particle Reconstruction</b> | <b>20</b> |
| 3.1      | Kalman Filter Algorithm . . . . .            | 21        |
| 3.1.1    | Steps . . . . .                              | 21        |
| 3.1.2    | KFParticle STAR Package . . . . .            | 23        |
| <b>4</b> | <b>Toolkit for Multivariate Analysis</b>     | <b>26</b> |
| 4.1      | Boosted Decision Tree . . . . .              | 29        |
| 4.2      | Multi Layer Perceptron . . . . .             | 31        |
| <b>5</b> | <b>Experimental Analysis</b>                 | <b>33</b> |
| 5.1      | Dataset Information . . . . .                | 33        |
| 5.1.1    | Event Quality Assurance . . . . .            | 34        |
| 5.1.2    | Centrality Selection . . . . .               | 34        |
| 5.1.3    | Track Quality Assurance . . . . .            | 37        |
| 5.2      | KFParticle Analysis . . . . .                | 41        |

|          |                                      |           |
|----------|--------------------------------------|-----------|
| 5.2.1    | Armenteros Podolanski Plot . . . . . | 42        |
| 5.2.2    | TMVA Analysis . . . . .              | 45        |
| <b>6</b> | <b>Conclusion</b>                    | <b>76</b> |
|          | <b>Bibliography</b>                  | <b>78</b> |

# List of Figures

|     |   |    |
|-----|---|----|
| 1.1 | Particles of SM which have been detected . . . . .  | 3  |
| 1.2 | Schematic representation of history of the Universe . . . . .   | 6  |
| 1.3 | Schematic QCD Phase Diagram for nuclear matter . . . . .  | 7  |
| 1.4 | Example of the production of a $V^0$ particle by a single charged<br>particle . . . . .   | 8  |
| 1.5 | $m_{J/\psi p}$ distribution . . . . .   | 10 |
| 1.6 | Invariant mass distribution $\Lambda K_s^0$ (in GeV) in min. bias Au+Au<br>collisions at $s_{\text{NN}} = 200$ GeV measured with the STAR exper-<br>iment in 2001 . . . . . | 12 |
| 2.1 | Aerial view of the RHIC accelerator complex at Brookhaven<br>National Laboratory . . . . .  | 14 |
| 2.2 | 3-D representation of the STAR Detector System . . . . .  | 17 |

|     |   |    |
|-----|---|----|
| 2.3 | The Time Projection Chamber detector of STAR . . . . .  | 18 |
| 3.1 | Scheme of Kalman Filter Algorithm . . . . .   | 23 |
| 3.2 | List of decay channels currently available in KFParticle . . . . .  | 24 |
| 4.1 | Flowchart showing steps involved in using TMVA for analysis   | 28 |
| 5.1 | (Left)primary vertex position along the longitudinal direction distribution( $V_z$ ) before event cuts. (Right) primary vertex position along the longitudinal direction distribution( $V_z$ ) after event cuts . . . . . | 35 |
| 5.2 | (Left)primary vertex position along radial direction distribution( $V_r$ ) before event cuts. (Right) primary vertex position along radial direction distribution( $V_r$ ) after event cuts . . . . .                     | 35 |
| 5.3 | Uncorrected charged particle multiplicity for $ \eta  < 0.5$ . . . . .  | 37 |
| 5.4 | (Left)Distribution of TPC Hits before track cuts.(Right) Distribution of TPC Hits after track cuts. . . . .   | 38 |
| 5.5 | (Left)Distribution of ratio of TPC Hits and possible Hits before track cuts.(Right) Distribution of ratio of TPC Hits and possible Hits after track cuts. . . . .   | 39 |
| 5.6 | (Left)Distribution of dEdx Hits before track cuts.(Right) )Distribution of dEdx Hits after track cuts . . . . .   | 39 |

|  |    |
|--|----|
| 5.7 (Left) Distribution of Pseudorapidity before track cuts.(Right) Distribution of Pseudorapidity after track cuts . . . . .  | 40 |
| 5.8 (Left) Energy loss of charged particles in TPC vs Momentum before track cuts.(Right) Energy loss of charged particles in TPC vs Momentum after cuts . . . . .  | 40 |
| 5.9 Schematic diagram for PQ reconstruction and decay . . . . .  | 41 |
| 5.10 Calculation of $p_T^{Arm}$ . . . . .  | 43 |
| 5.11 (Left) Armenteros-Podolanski plot for $K_s^0$ for real data after MLPBNN response cut $> 0.03$ (Data as Bkg) . (Right) Armenteros-Podolanski plot for $K_s^0$ for 1PQ/event simulation .  | 43 |
| 5.12 (Left) Armenteros-Podolanski plot for $\Lambda^0$ for real data after MLPBNN response cut $> 0.03$ (Data as Bkg). (Right) Armenteros-Podolanski plot for $\Lambda^0$ for 1PQ/event simulation .   | 44 |
| 5.13 (Left) Armenteros-Podolanski plot for reconstructed PQ-candidates from real tracks after MLPBNN response cut $> 0.03$ (Data as Bkg). (Right) Armenteros-Podolanski plot for reconstructed PQ from 1PQ/event simulation tracks . . . . . | 44 |
| 5.14 Distribution of the input variables used. Here Blue distribution represents the signal and red distribution represents the background . . . . .   | 48 |

|   |    |
|---|----|
| 5.15 Distribution of the input variables used. Here Blue distribution represents the signal and red distribution represents the background . . . . .  | 49 |
| 5.16 (TopLeft)ROC curve for MLPBNN method. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by MLPBNN. (Bottom) Over-training check of the classifier. . . . . | 50 |
| 5.17 $\Lambda^0 K_s^0$ invariant mass for centrality 0-80% after MLPBNN response cuts of (from top left to right) $> (0, 0.1, 0.01 \text{ and } 0.02)$ with 1MeV per bin . . . . .  | 53 |
| 5.18 $\Lambda^0 K_s^0$ invariant mass for centrality 0-80% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$ with 1MeV per bin . . . . .  | 54 |
| 5.19 $\Lambda^0 K_s^0$ invariant mass for centrality 10-80% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0, 0.1, 0.01 \text{ and } 0.02)$ with 1MeV per bin . . . . .   | 55 |
| 5.20 $\Lambda^0 K_s^0$ invariant mass for centrality 10-80% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$ with 1MeV per bin . . . . .   | 56 |

|   |    |
|---|----|
| 5.21 $\Lambda^0 K_s^0$ invariant mass for centrality 10-50% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0,0.1,0.01 \text{ and } 0.02)$ with 1MeV per bin . . . . .   | 57 |
| 5.22 $\Lambda^0 K_s^0$ invariant mass for centrality 10-50% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0.03,0.04,0.05,0.07 \text{ and } 0.09)$ with 1MeV per bin . . . . .  | 58 |
| 5.23 $\Lambda^0 K_s^0$ invariant mass for centrality 50-80% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0,0.1,0.01 \text{ and } 0.02)$ with 1MeV per bin . . . . .   | 59 |
| 5.24 $\Lambda^0 K_s^0$ invariant mass for centrality 50-80% after MLPBNN response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from top left to right) $> (0.03,0.04,0.05,0.07 \text{ and } 0.09)$ with 1MeV per bin . . . . .  | 60 |
| 5.25 (TopLeft)ROC curve for MLPBNN method with real data as background. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by MLPBNN with real data as background. (Bottom) Overtraining check of the classifier with real data as background. | 62 |
| 5.26 $\Lambda^0 K_s^0$ invariant mass for centrality 0-80% after MLPBNN response cuts with real data as background (from top left to right) $> (0,0.1,0.01 \text{ and } 0.02)$ with 1MeV per bin . . . . .  | 63 |

|   |    |
|---|----|
| 5.27 $\Lambda^0 K_s^0$ invariant mass for centrality 0-80% after MLPBNN response cuts with real data as background (from top left to right) > (0.03,0.04,0.05,0.07 and 0.09) with 1MeV per bin . . .  | 64 |
| 5.28 $\Lambda^0 K_s^0$ invariant mass for centrality 10-80% after MLPBNN response cuts with real data as background (from top left to right) > (0,0.1,0.01 and 0.02) with 1MeV per bin . . . . .      | 65 |
| 5.29 $\Lambda^0 K_s^0$ invariant mass for centrality 10-80% after MLPBNN response cuts with real data as background (from top left to right) > (0.03,0.04,0.05,0.07 and 0.09) with 1MeV per bin . . . | 66 |
| 5.30 $\Lambda^0 K_s^0$ invariant mass for centrality 10-50% after MLPBNN response cuts with real data as background (from top left to right) > (0,0.1,0.01 and 0.02) with 1MeV per bin . . . . .      | 67 |
| 5.31 $\Lambda^0 K_s^0$ invariant mass for centrality 10-50% after MLPBNN response cuts with real data as background (from top left to right) > (0.03,0.04,0.05,0.07 and 0.09) with 1MeV per bin . . . | 68 |
| 5.32 $\Lambda^0 K_s^0$ invariant mass for centrality 50-80% after MLPBNN response cuts with real data as background (from top left to right) > (0,0.1,0.01 and 0.02) with 1MeV per bin . . . . .      | 69 |
| 5.33 $\Lambda^0 K_s^0$ invariant mass for centrality 50-80% after MLPBNN response cuts with real data as background (from top left to right) > (0.03,0.04,0.05,0.07 and 0.09) with 1MeV per bin . . . | 70 |

|  |    |
|--|----|
| 5.34 (TopLeft)ROC curve for BDT method $20\Lambda^0$ and $20K_s^0$ as background. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by BDT. (Bottom) Overtraining check of the classifier. . . . . | 72 |
| 5.35 $\Lambda^0 K_s^0$ invariant mass for centrality 0-80% after BDT response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from left to right)<br>$> (-0.4 \text{ and } -0.35)$ with 1MeV per bin . . . . .  | 73 |
| 5.36 $\Lambda^0 K_s^0$ invariant mass for centrality 10-80% after BDT response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from left to right)<br>$> (-0.4 \text{ and } -0.35)$ with 1MeV per bin . . . . .   | 74 |
| 5.37 $\Lambda^0 K_s^0$ invariant mass for centrality 10-50% after BDT response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from left to right)<br>$> (-0.4 \text{ and } -0.35)$ with 1MeV per bin . . . . .   | 74 |
| 5.38 $\Lambda^0 K_s^0$ invariant mass for centrality 50-80% after BDT response cuts with $20\Lambda^0$ and $20K_s^0$ as Background (from left to right)<br>$> (-0.4 \text{ and } -0.35)$ with 1MeV per bin . . . . .   | 75 |

# List of Tables

|     |   |    |
|-----|---|----|
| 1.1 | Properties of $K_s^0$ and $\Lambda$ . . . . .   | 9  |
| 5.1 | Data-set and trigger information for Au+Au collisions at $\sqrt{s_{NN}} = 39$ GeV . . . . . | 34 |
| 5.2 | Event Selection Cuts for Au+Au collisions at $\sqrt{s_{NN}} = 39$ GeV                       | 34 |
| 5.3 | Track Parameter cuts for Au+Au collisions at $\sqrt{s_{NN}} = 39$ GeV                       | 38 |
| 5.4 | KFParticle Cuts for Au+Au collisions at $\sqrt{s_{NN}} = 39$ GeV . . .                      | 42 |
| 5.5 | Dataset used for TMVA Training and Testing . . . . .  | 46 |
| 5.6 | Total events which pass the corresponding MLPBNN response cuts. . . . .                     | 52 |
| 5.7 | Total events which pass the corresponding MLPBNN response cuts. . . . .                     | 61 |
| 5.8 | Cut Set used for analysis using BDT . . . . .   | 72 |

# Chapter 1

## Introduction

The Higgs Boson discovery with the ATLAS and CMS detector at the LHC [1] [2] sent a huge wave of enthusiasm among particle physicists. Among other things, in the QCD sector, in the last few years various pentaquarks states have been discovered at the LHC. This thesis is concerned with the search for evidence of the production of pentaquark states in the  $\Lambda K_s^0$  decay modes using data from the STAR experiment at RHIC and improve the signal using the Multivariate Analysis (TMVA) package of ROOT.

### 1.1 Standard Model

The Standard Model (SM) unifies three of the four known fundamental forces of nature. Currently it is the best description of the subatomic world and

has become established as a well tested physics theory through different experiments. The Standard Model is essentially a local  $SU(3) \times SU(2) \times U(1)$  gauge symmetry and the dynamics depend on 18 parameters [3] whose values are calculated experimentally. They are

1. Fine structure constant ( $\alpha$ )
2. Weinberg angle or weak mixing angle ( $\theta_w$ )
3. The coupling constant  $g_3$  of the strong interaction
4. The electroweak symmetry breaking energy scale (or the Higgs potential vacuum expectation value, (v.e.v.))  $v$
5. The Higgs potential coupling constant  $\lambda$  or alternatively, the Higgs mass  $m_H$
6. The three Cabibbo-Kobayashi-Maskawa (CKM) mixing angles ( $\theta_{12}, \theta_{23}, \theta_{13}$ )
7. The CKM CP violation phase ( $\delta$ )
8. Nine Yukawa coupling constants

These are the parameters which have been found experimentally but there are more parameters which have to be determined experimentally.

All matter around us is made up of elementary particles, the building blocks of matter. These particles occur in two basic types called quarks and leptons (fermions). Each group consists of six "types or flavour" of particles. These

|             |               |           |               |                 |           |                |                  |           |
|-------------|---------------|-----------|---------------|-----------------|-----------|----------------|------------------|-----------|
| u           | u             | u         | c             | c               | c         | t              | t                | t         |
| d           | d             | d         | s             | s               | s         | b              | b                | b         |
| $e^-$       | $\nu_e$       |           | $\mu^-$       | $\nu_\mu$       |           | $\tau^-$       | $\nu_\tau$       |           |
| $\bar{u}$   | $\bar{u}$     | $\bar{u}$ | $\bar{c}$     | $\bar{c}$       | $\bar{c}$ | $\bar{t}$      | $\bar{t}$        | $\bar{t}$ |
| $\bar{d}$   | $\bar{d}$     | $\bar{d}$ | $\bar{s}$     | $\bar{s}$       | $\bar{s}$ | $\bar{b}$      | $\bar{b}$        | $\bar{b}$ |
| $\bar{e}^+$ | $\bar{\nu}_e$ |           | $\bar{\mu}^+$ | $\bar{\nu}_\mu$ |           | $\bar{\tau}^+$ | $\bar{\nu}_\tau$ |           |
| g           | g             | g         | g             | g               | g         | $\gamma$       | $W^-$            | $W^+$     |
|             |               |           |               |                 |           |                | $Z^0$            | H         |

Figure 1.1: Particles of SM which have been detected [5].

particles are paired in three "generations". The lightest and most stable particles make up the first generation, whereas the heavier and less stable particles belong to the second and third generations [4]. The other type of elementary particle are the bosons which are the force carriers. Each fundamental force has its own corresponding boson; the strong force is carried by the gluon, the electromagnetic force is carried by the photon, and the W and Z bosons are responsible for the weak force.

## 1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the most recent developed component of the SM. It describes the strong interaction between quarks and gluons. It is a non-Abelian  $SU(3)$  theory[6]. The gluons are massless and mediate the strong interaction force between quarks. There are six flavours of quarks: up (u), down (d), charm (c), strange (s), top (t), bottom (b).

The quarks and gluons carry color <sup>1</sup> charge quantum number which are of three kind red, blue and green. Similarly anti-quarks carry anti-color charge which are anti-red, anti-blue and anti-green. This can be thought of as electric charge in Quantum Electrodynamics (QED) but there is a major difference between QCD and QED. The gluons unlike photons carry the charge so they can self interact among each other. This results in asymptotic freedom and color confinement [7] [6].

Hadrons are composed of quarks and gluons. Baryons and mesons are together known as hadrons. Baryon contains three quarks and meson contains a pair of quark and anti-quark. Confinement requires that all naturally occurring particles be color singlets (eg. Proton and Neutron)<sup>2</sup>, and this explains why the quarks and gluons never appear as free particles [8]. They are confined within hadrons which are colorless.

This colorless combination can be characterised by Baryon number. Baryons (three quarks) have a baryon number of +1, mesons (one quark, one anti-quark) have a baryon number of 0, and antibaryons (three antiquarks) have a

---

<sup>1</sup>nothing to do with visible color

<sup>2</sup>In other words, if one could measure the color of the state, there would be equal probabilities of it being red-antired, blue-antiblue, or green-antigreen.

baryon number of -1. Baryon number for a quark is  $\frac{1}{3}$  and that of anti-quark is  $-\frac{1}{3}$ . For normal hadrons we have seen in case of mesons, baryons and antibaryons how white color or colorlessness can be achieved. We can also have different combinations of quark and antiquark than those mentioned above provided each pair has matching color/anticolor. This leads to what are called as "exotic hadrons".

### 1.3 Quark Gluon Plasma

Quark Gluon Plasma (QGP) is defined as a state of deconfined strongly interacting matter. QCD's characteristic feature of confinement and asymptotic freedom makes the prediction that at high enough density an entire ensemble of quarks can be mutually deconfined into a phase of matter governed by a QCD equation of state (EoS)[9][10]. Study of such states has application in modelling of universe immediately following the Big Bang in cooler but dense systems, like neutron stars and also in analysing gravitational signals from neutron star merger Figure 1.2[9].

Theoretical predictions show the QGP formation energy density( $\epsilon_c$ ) is order of 0.5 to  $1\text{GeV}/\text{fm}^3(10^{18}\text{kg}/\text{m}^3)$  [12]. QCD phase diagram is shown in Figure 1.3.

The QCD phase diagram is plotted between temperature (T) and the baryon chemical potential ( $\mu_B$ ). The baryon chemical potential ( $\mu_B$ ) is the energy that is absorbed or released due to change in the number of baryons. It measures the imbalance between quarks and anti-quarks. Different phases of

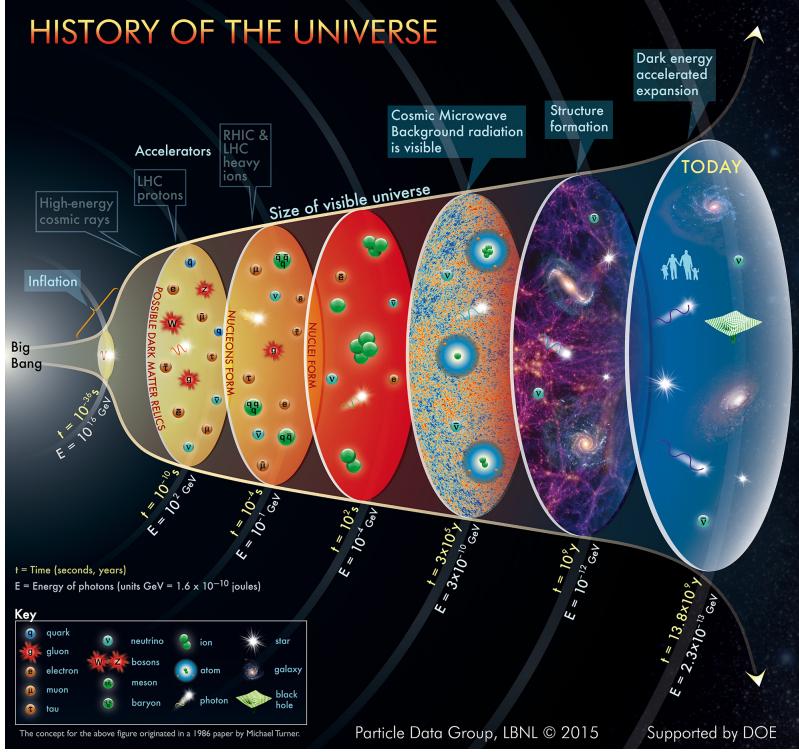


Figure 1.2: Schematic representation of history of the Universe[11].

nuclear matter can be studied by varying  $T$  and  $\mu_B$ . At  $T = 0$  and  $\mu_B \approx 938$  MeV the nuclear state of matter is observed. Deconfined state of quarks and gluons is expected at high  $T$  and high  $\mu_B$  [14]. At low  $T$  and low  $\mu_B$  quarks are confined within hadrons. At high  $T$  and low  $\mu_B$  other interesting phases like color superconductivity can be observed [15].

Theoretically, the phases of QCD are explored through numerical simulations on a space-time lattice in non-perturbative QCD regime. This is known as lattice QCD (lQCD). It predicts a phase transition from hadronic phase to QGP phase of the nuclear matter at a critical temperature ( $T_c$ ) of  $(154 \pm 8\text{MeV})^3$

---

<sup>3</sup>In this analysis all quantities are expressed in terms of natural units i.e  $\hbar = c = k_B = 1$

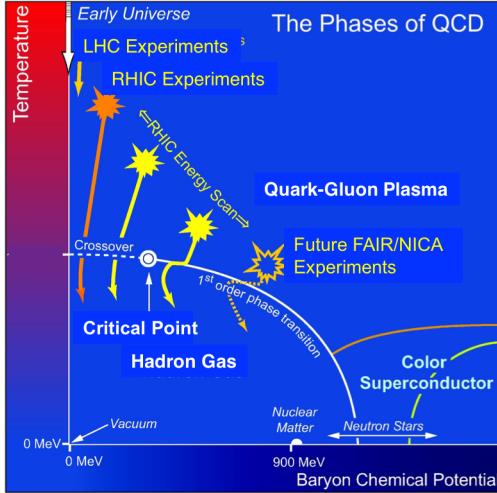


Figure 1.3: Schematic QCD Phase Diagram for nuclear matter [13]

[16]. It also predicts a smooth crossover between phases at large  $T$  and smaller  $\mu_B$  [17] and first order transition may take place between hadron gas and QGP at smaller  $T$  and larger  $\mu_B$  [18]. The point in the QCD phase diagram where the first order phase transition ends would be the QCD critical point.

The Quark Gluon Plasma is studied in heavy ion collision where nucleus of heavy atoms like Gold(Au) and Lead(Pb) are accelerated to relativistic speeds and then collided. There are no direct signatures of QGP like electromagnetic plasma so they are studied by studying the final state hadrons since the lifetime of QGP is very short ( $\approx$  few fm/c)[19]. Some of the signatures of QGP are Jet Quenching in central Au+Au collision[20] [21], enhanced production of strange hadrons in relativistic heavy ion collisions[22] and observation of large elliptic flow  $v_2$  [20].

## 1.4 $V^0$ particles and decays

$V^0$  particles are unstable subatomic particles which decay into a pair of daughter particles. Since they are neutral and the tracks of the decay products is seen as a letter 'V' in bubble chamber Figure 1.4 they are called as  $V^0$  particles. For this analysis we have considered the  $V^0$  particles:  $\Lambda^0$  and  $K_s^0$  and their decay channels are given in Table 1.1.

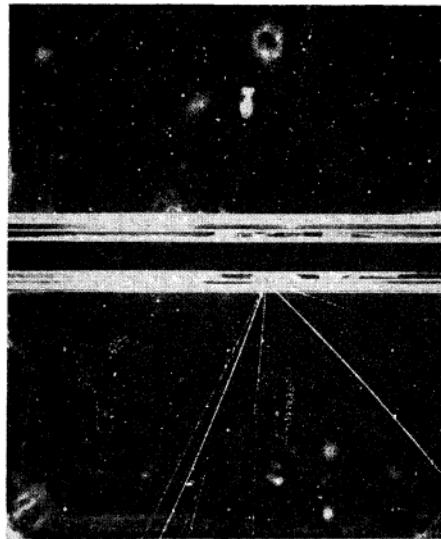


Figure 1.4: Example of the production of a  $V^0$  particle by a single charged particle. The initiating particle enters the lead plate between the chambers almost vertically from above and slightly to the right of the center of the chamber. In the lead plate it undergoes a nuclear collision, from which six charged particles, the  $V^0$  particle, and an unknown number of neutral particles emerge. The  $V^0$  particle decays near the bottom of the lower chamber into a heavily-ionizing particle and a negative particle of about minimum ionization [23].

| Particle    | Decay Channel   | Branching Ratio (%) | Decay Length (cm) | Mass ( $MeV/c^2$ )   |
|-------------|-----------------|---------------------|-------------------|----------------------|
| $K_s^0$     | $\pi^+ + \pi^-$ | $69.20 \pm 0.05$    | $2.684 \pm 0.001$ | $497.611 \pm 0.013$  |
| $\Lambda^0$ | $\pi^+ + p$     | $63.90 \pm 0.50$    | $7.890 \pm 0.060$ | $1115.683 \pm 0.006$ |

Table 1.1: Properties of  $K_s^0$  and  $\Lambda$  [11]

## 1.5 Pentaquark

Experimental searches for pentaquark hadrons comprised of light flavors have a long and vivid history[24]. Pentaquarks are exotic hadrons having 4 quarks and 1 antiquark [25][26][27]. The search for pentaquarks began in 1976 almost 10 years after they were proposed[28]. But there was no definite conclusion. In 2003, a second wave of observations occurred motivated by some theoretical predictions regarding existence of exotic hadrons[29][30][31]. The model serving as the basis for these searches was known as the chiral quark soliton model[30]. It predicts ten exotic states from the anti-decuplet that could not be constructed from only three quarks. It is an relativistic field theoretical model of the nucleon and it was inspired by an effective field theory proposed by Tony Skyrme[29] where the nucleons were "solitons". The best known candidate was the  $\Theta(1540)^+$  state, first observed by the LEPS collaboration in Japan[32]. Several experiments searched for this state[32][33], but there was no conclusive evidence since more data and subsequent more sensitive experiments did not confirm these claims. So the 2006 Particle Data Group listing[34] included a statement: "The conclusion that pentaquarks in general, and that  $\Theta^+$ , in particular, do not exist, appears compelling."

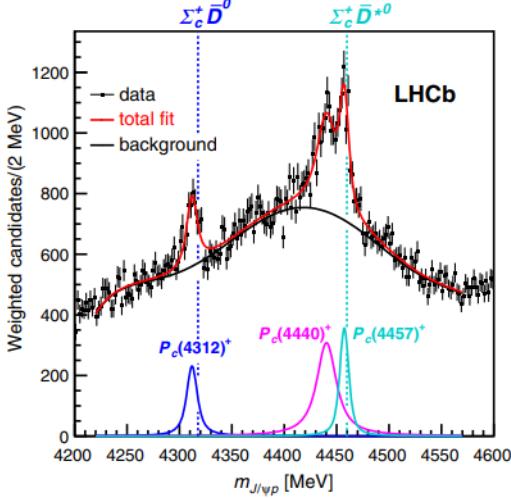


Figure 1.5:  $m_{J/\psi p}$  distribution[36].

But recent advancements in detector technologies and energy scale of the particle accelerators have led to the search for other pentaquark states. In 2015 the LHCb collaboration reported a significant  $J/\psi p$  structure in  $\Lambda_0^b \rightarrow J/\psi p K^-$  decays which cannot be accounted for by conventional particles [35]. These structures were reported as Pentaquark candidates,  $P_c(4380)^+$  and  $P_c(4450)^+$  with statistical significance of 9 and 12 standard deviations respectively Figure 1.5. In 2019 LHCb collaboration announced discovery of a new pentaquark particle,  $P_c(4312)^+$  which decays to proton and  $J/\psi$  (with quark content charm and anticharm quark). It had statistical significance of 7.3 sigma[36].

## 1.6 Thesis Motivation

In this thesis we present the results of search for pentaquark candidates in the  $\Lambda K_s^0$  decay mode using the Kalman-Filter(KF) Particle reconstruction code with STAR Experiment at RHIC. This analysis was motivated by previous searches done on data collected for Au+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV,in the year 2001 and  $\sqrt{s_{NN}} = 39$  GeV in the year 2010 using the STAR detector[37]. There was a narrow peak observed at  $1734 \pm 0.5 \pm 5$  MeV in the analysis of  $\sqrt{s_{NN}} = 200$  GeV Au+Au collision data Figure 1.6. If this peak corresponds to a real particle it would be a candidate for the  $N^0$  ( $uds d\bar{s}$ ) state (octet, antidecuplet or 27-plet) or  $\Xi^0$  ( $udss d\bar{s}$ ) state [30]. Since the luminosity of  $\sqrt{s_{NN}} = 39$  GeV 2010 data ( $L_{avg}(cm^{-2}s^{-1}) = 1.3 \times 10^{26}$ ) is similar to that of the data used in this analysis [38] ( $L_{avg}(cm^{-2}s^{-1}) = 1.5 \times 10^{26}$ ) we have used the dataset. The previous analysis on the Au+Au collisions at  $\sqrt{s_{NN}} = 39$  GeV, 2010 dataset used the standard STAR code for  $V^0$  particle reconstruction and rotational background analysis but did not find a signal peak. We have used the KFParticle reconstruction code for particle reconstruction and TMVA package of root for background rejection. The invariant mass of  $\Lambda K_s^0$  <sup>4</sup> is plotted to check for signal peak.

---

<sup>4</sup>Here  $\Lambda$  corresponds to  $\Lambda^0$  baryon with quark content ( $uds$ ) and  $K_s^0$  corresponds to K-short meson with quark content ( $d\bar{s}$ ) or  $s\bar{d}$

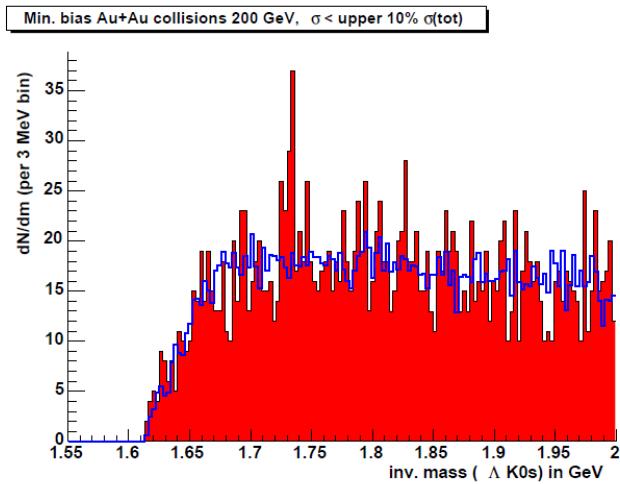


Figure 1.6: Invariant mass distribution  $\Lambda K_s^0$  (in GeV) in min. bias Au+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV measured with the STAR experiment(red) together with the estimated background using the mixed events technique (blue) [38].

# Chapter 2

## Experimental Setup

This chapter briefly describes the particle accelerator and the detector experimental facility and the detector subsystems used to collect data for this analysis.

### 2.1 Relativistic Heavy Ion Collider

The Relativistic Heavy Ion Collider (RHIC) is a particle accelerator built in Brookhaven National Laboratory (BNL), long island, Upton, New York, USA, for studying the properties of QGP formed during heavy ion collisions [39]. RHIC started in 2000 and has been running for 22 years now. RHIC is one of the only two operating heavy ion colliders (other is the LHC) in the world. There have been four major experiments which have operated in RHIC. These are BRAHMS, PHENIX, PHOBOS and STAR [40] [41] [42]

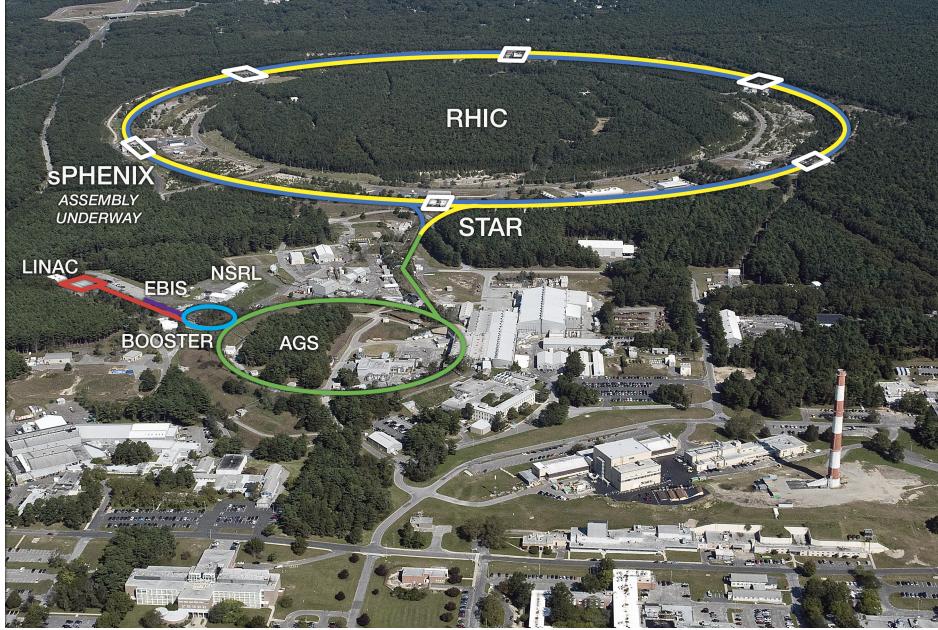


Figure 2.1: Aerial view of the RHIC accelerator complex at Brookhaven National Laboratory, Upton, New York [45].

[43]. Currently, only STAR is active because PHOBOS and BRAHMS were decommissioned in 2005 and 2006 respectively and PHENIX is being modernized into new experiment called sPHENIX [44].

The collider has two superconducting rings of circumference about 3.8 km. The ring through which ions travel in a clockwise direction is termed as the "blue ring" and in counter-clockwise direction is termed as the "yellow ring". Both of the superconducting rings intersect at six interaction points where the collisions take place and different experiments are setup as shown in Figure 2.1. The ions before being injected into RHIC are accelerated in a complex of smaller accelerators- Tandem Van de Graaff, Booster and Alternating Gradient Synchrotron (AGS). Unlike proton beams it is a very complicated procedure for producing fully stripped high intensity and high

brightness heavy ion beams, such as the Au beams, that are suitable for a collider. This is due to the fact that electrons are more tightly bound to a heavy nucleus, and effective stripping of the final K-shell electrons requires the acceleration of the partially stripped ions to tens of GeV per nucleon. In case of Au ion beam the negatively charged Au ion beams from the pulsed sputter ion source at the Tandem Van de Graaff(100 $\mu$ A in a 700 $\mu$ s pulse) are passed through a carbon stripping foil at the Tandem's high voltage terminal to partially strip off their electrons and then they are accelerated to 1 MeV per nucleon in the second stage of the Tandem. After further stripping at the exit of the Tandem and charge selection by bending magnets, beams of gold ions with a charge state of +32 are delivered to the Booster Synchrotron and accelerated to 95 MeV per nucleon [46]. These ions are stripped again at the exit of the Booster and reaches a Helium like configuration having charge state of +77. The ion beam is then injected to AGS for acceleration to the RHIC injection energy of 10.8 GeV per nucleon. The Au ion is injected into AGS in 24 bunches, where they are de-bunched and re-bunched to four bunches which are then ejected out. These four bunches are ejected at the top energy, one bunch at a time, and transferred to RHIC through the AGS-to-RHIC Beam Transfer Line. Gold ions are fully stripped to a charge state of +79 at the exit from the AGS. Then they are accelerated in the rings of RHIC to the desired collision energy [46].

RHIC is also the only spin polarised proton collider ever built and it is capable of colliding spin-polarized proton upto  $\sqrt{s_{NN}} = 500$  GeV. The RHIC can achieve energies upto 100 GeV per nucleon for  $^{197}_{79}\text{Au}$  ions. For proton beams it can achieve 250 GeV per nucleon and 96.5 GeV per nucleon for heavier

ions like  $^{238}_{92}\text{U}$ .

## 2.2 Solenoidal Tracker at RHIC

The Solenoidal Tracker at RHIC (STAR) experiment was constructed to understand at the fundamental level the hadronic interactions at high energy densities. Since these interactions produce a large number of particles STAR was designed to record high multiplicity events and cover full azimuth within pseudorapidity region  $|\eta| < 1$ . The STAR detector consists of a number of detectors which measure the various data needed for analysis, like the Time Projection Chamber (TPC) which is placed inside a solenoidal magnet to enable tracking, momentum analysis and particle identification via the ionisation energy loss per distance( $\frac{dE}{dx}$ ). It also contains a Time of flight(TOF) detector for particle identification at higher momenta. The Barrel Electro-Magnetic Calorimeter(BEMC) and the Barrel Shower Maximum Detector (BSMD) is present outside the magnet and is used for triggering of high  $p_T$  particles and jets [47]. The STAR magnet can create magnetic field with magnitude of 0.5 Tesla in full field mode and 0.25 Tesla in half field mode with the field parallel to the z direction. The overview of STAR detector can be seen in Figure 2.2. All the results shown in this thesis are based on data collected using the STAR detector mainly the TPC and Trigger detectors.

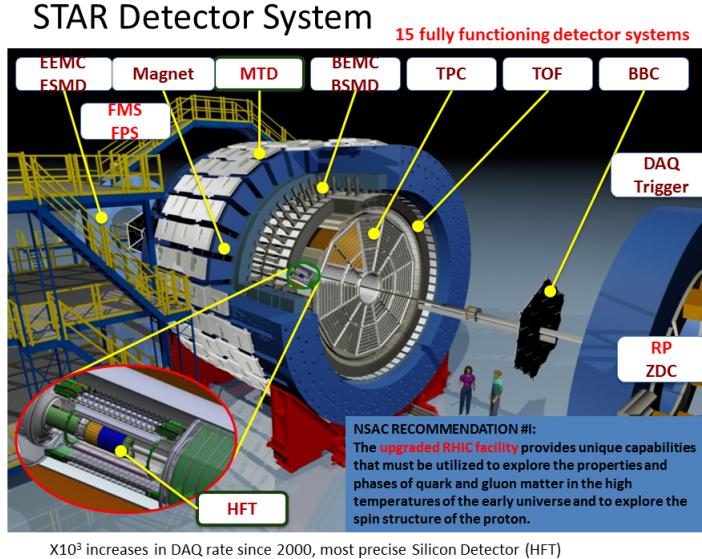


Figure 2.2: 3-D representation of the STAR Detector System[48].

### 2.2.1 Time Projection Chamber

The Time Projection Chamber (TPC) is the primary tracking device for STAR. Its acceptance covers  $\pm 1.8$  units of pseudo-rapidity ( $\eta$ ) through the full azimuthal angle and over the full range of multiplicities. Particles are identified over a momentum range from 100 MeV/c to greater than 1 GeV/c, and momenta are measured over a range of 100 MeV/c to 30 GeV/c [49]. The TPC is 4.2 m long and 4 m in diameter. It is an empty volume of gas in an well defined, uniform electric field of  $\approx 135$  V/cm. The TPC is filled with P10 gas (10% methane, 90% argon) regulated at 2 mbar above atmospheric pressure [50]. The paths of primary ionising particles passing through the gas chamber are reconstructed through the secondary electrons which drift to the readout end caps at the end of the chamber. The uniform electric field

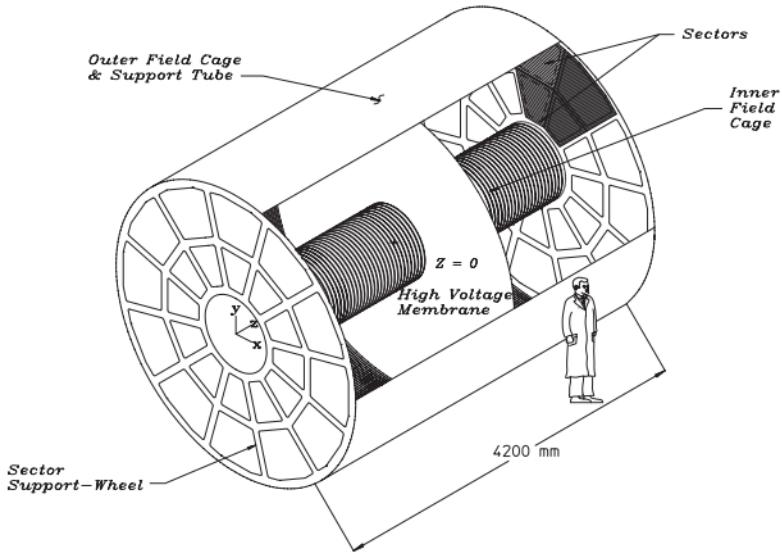


Figure 2.3: The Time Projection Chamber detector of STAR[49].

is defined by a thin conductive Central Membrane (CM) at the center of the TPC.

### 2.2.2 Time Of Flight

The Time of Flight (TOF) detector is responsible for particle identification at high momenta ( $> 1 \text{ GeV}/c$ ). The STAR TOF is based on Multi-gap Resistive Plate Chamber (MRPC) technology [51]. The TOF detector covers a pseudo-rapidity region of  $|\eta| < 0.9$  and full azimuth angle ( $\Delta\phi = 2\pi$ ). The TOF detector surrounds the TPC and comprises of 120 highly segmented trays. Each tray consists of 32 Multi-gap Resistive Plate Chamber (MRPC) modules that are placed along beam (z) direction. The TOF detector is comprised of two parts : the Vertex Position Detector (VPD) which is the

start detector and the other is the TOF tray which is the stop detector. The charged particles produced during heavy ion collision can be identified from their time of flight measured in the TOF detector along with their momentum ( $p$ ) from the TPC detector. The time of flight ( $\Delta t$ ) is the difference in time between the start time provided by the VPD's and the stop time provided by the TOF trays. The total time resolution of TOF is about  $100\ ps$ . The path length( $s$ ) calculated from TPC can be used to calculate the velocity( $\beta$ ) of the particle. The momentum( $p$ ) from TPC can be used to calculate the mass of particle using the equation Equation 2.3.

$$\beta = \frac{s}{c\Delta t} \quad (2.1)$$

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \quad (2.2)$$

$$m = \frac{p}{\gamma\beta c} \quad (2.3)$$

Here  $c$  is the velocity of light.

# Chapter 3

## Kalman Filter Particle Reconstruction

In this chapter we discuss the Kalman Filter (KF) algorithm and the KF-Particle reconstruction code implemented for particle reconstruction of short lived particles in detected in the STAR detector. The KFParticle code was developed in FIAS (Frankfurt Institute for Advanced Studies) initially for the CBM and the ALICE experiments at the LHC [52]. It is based on the Kalman Filter algorithm and was motivated to speed up computing time for the increase in the size of data from the detectors.

## 3.1 Kalman Filter Algorithm

The Kalman Filter Algorithm is a method for estimating states of a dynamical system [53]. The theory for the algorithm was developed by Rudolf Emil Kálmán[54]. The application of Kalman Filter to track fitting is straightforward if one interprets the track as a discrete dynamic system[55]. In decayed particle reconstruction the algorithm is mainly used for primary and secondary vertex reconstruction using an extended version of the algorithm known as Extended Kalman Filter [52] [56].

### 3.1.1 Steps

First let us define the notations which will be used:

$x^T$  - denotes transpose of vector  $x$ .

$\langle \rangle$  - denotes the expectation value.

Covariance matrix of a vector  $x$  ( $\text{cov}(x)$ ) =  $\langle (x - \langle x \rangle) \cdot (x - \langle x \rangle)^T \rangle$

State vector  $r^t$  - denotes the vector of real numbers that represents the unknown quantities to be estimated (for example, parameters of a track). If the state vector changes from one measurement to another it is denoted by  $r_k^t$

Measurement  $m_k$  - where  $k = 1, 2, \dots, n$  denotes known measured quan-

entity which linearly depends on state vector:

$$m_k = H_k r^t + \eta_k \quad (3.1)$$

where  $\eta_k$  denotes an error of the  $k$ -th measurement and matrix  $H_k$  denotes the model of measurement. It is assumed that the error is unbiased and its covariance matrix  $V_k$  is known i.e.

$$\begin{aligned} \langle \eta_k \rangle &= 0 \\ \langle \eta_k \cdot \eta_k^T \rangle &= V_k \end{aligned} \quad (3.2)$$

If the measurements  $m_k$  nonlinearly depend on  $r^t$ , it is necessary to linearize the model of measurement and this procedure is known as the extended Kalman Filter method.

$\tilde{r}_k$  - denotes the optimal estimation of the vector  $r_k^t$  according to the first  $k - 1$  measurements.

$A_k$  - denotes the extrapolation operator which describes deterministic changes of the state vector  $r^t$  between the two measurements.

$Q_k$  - denotes the process noise which describes the random deviations of the state vector.

The algorithm can be summarised in the flowchart as shown in Figure 3.1.

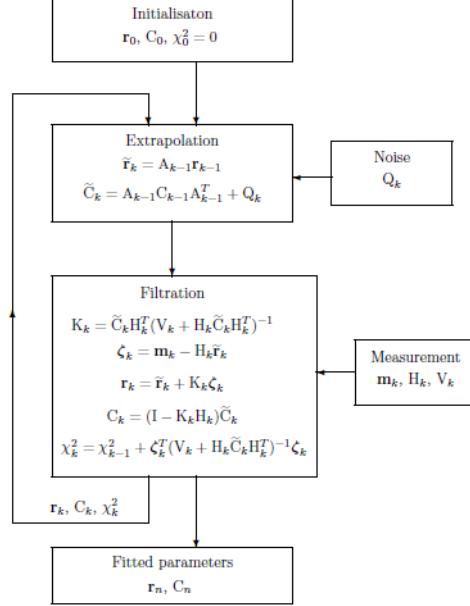


Figure 3.1: Scheme of Kalman Filter Algorithm [52].

$\tilde{C}_k$  denotes the covariance matrix of the best estimator obtained in the previous step and extrapolated to  $k$ -th measurement.  $\zeta_k$  is called residual and the matrix  $K_k$  is the gain matrix. The  $\chi^2_k$  value is the total  $\chi^2$  deviation of the estimation from the measurements. The algorithm steps 2 and 3 are repeated  $n$  times, for each measurement  $m_k$ ;  $k = 1, \dots, n$ . After the filtration of the last measurement  $m_n$ , the obtained estimator  $r_n$  is the desired best estimator with the covariance matrix  $C_n$ .

### 3.1.2 KFParticle STAR Package

The main advantage of using KFParticle reconstruction package is taking into account the experimental errors in the procedure of identification of

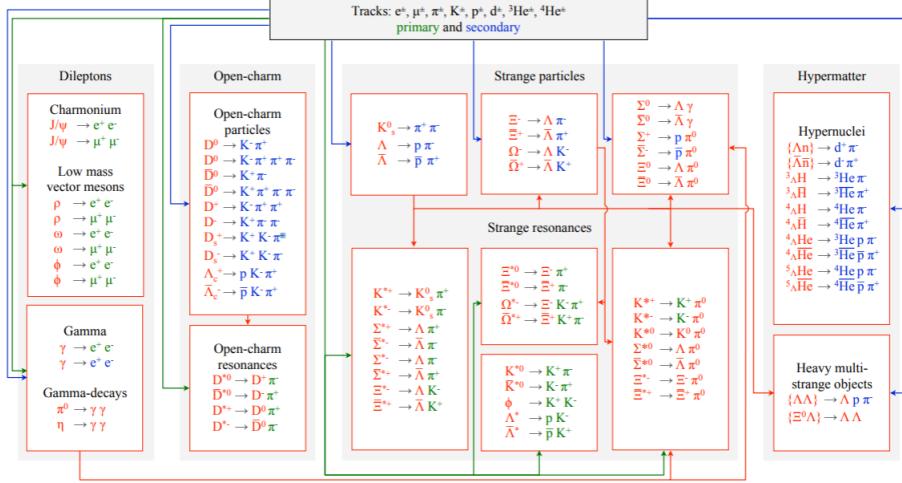


Figure 3.2: List of decay channels currently available in KFParticle[56].

particles as well as its independence of detector geometry as a result of which it can be implemented in any collider experiments [52] [56]. The algorithm is implemented in STAR in same way the usual STAR reconstruction software was implemented in the form of StKFParticleAnalysisMaker with Init(), Make() and Finish() functions. The main classes which the user needs to use for analysis is the StKFParticleAnalysisMaker and StKFParticleInterface classes. StKFParticleInterface Class controls the processing of the events and user can add their needed cuts in the ProcessEvent() function. The KFParticle package currently offers 70 decay channels that can be analysed and user can modify the code for their needed decay channels. The below figure shows the decay channels Figure 3.2.

Other than the track quality and event selection cuts KFParticle provides the option to cut on  $\chi^2_{prim}$ ,  $\chi^2_{fit}$ ,  $\chi^2_{topo}$  and  $l/\Delta l$ . The  $\chi^2_{prim}$  cut is a criterion

to separate tracks into primary and secondary. This criterion represents the probability of the trajectory intersecting primary vertex within uncertainties assuming the particle parameters follow Gaussian distribution and  $\chi^2_{prim}$  follows  $\chi^2$  distribution. For example if  $\chi^2_{prim} > 18.6$  there is a probability of 0.01% that the track is primary and this is the default value used in the KFParticle code. Similarly  $\chi^2_{fit}$  characterizes probability of daughter trajectories intersecting within their uncertainties,  $\chi^2_{topo}$  gives the probability that two tracks originate from the same decay vertex (It minimizes distance of closest approach between them). and  $l/\Delta l$  is distance from the primary vertex to the decay point normalised on its error.

# Chapter 4

## Toolkit for Multivariate Analysis

Artificial Intelligence (AI) and Machine Learning (ML) can be considered as one of the most remarkable achievements of science and technology. They have become an integral part of every research involving large datasets. As a result, high energy physics is no exception either. Machine Learning can be thought of as multivariate analysis (MVA) which is the process of utilising multiple variables to reach an output. ROOT has developed its own package for multivariate analysis called the Toolkit for Multivariate Analysis (TMVA). Although this package is specifically aimed at High Energy physics (HEP) it is not limited to this field only. It contains various Machine Learning algorithms or methods which can be used for better distinction between signal and background in HEP or for other data analysis purposes other than HEP.

The methods available in the TMVA package can be listed as follows[57]

1. Rectangular cut optimisation
2. Projective likelihood estimator (PDE approach)
3. Likelihood estimator using self-adapting phase-space binning (PDE-Foam)
4. k-Nearest Neighbour (k-NN) Classifier
5. Fisher discriminants (linear discriminant analysis)
6. Linear discriminant analysis (LD)
7. Function discriminant analysis (FDA)
8. Artificial Neural Networks (nonlinear discriminant analysis)
9. Deep Learning
10. Support Vector Machine (SVM)
11. Boosted Decision and Regression Trees
12. Predictive learning via rule ensembles (RuleFit)

The procedure to use TMVA in one's analysis can be seen in the flowchart  
Figure 4.1.

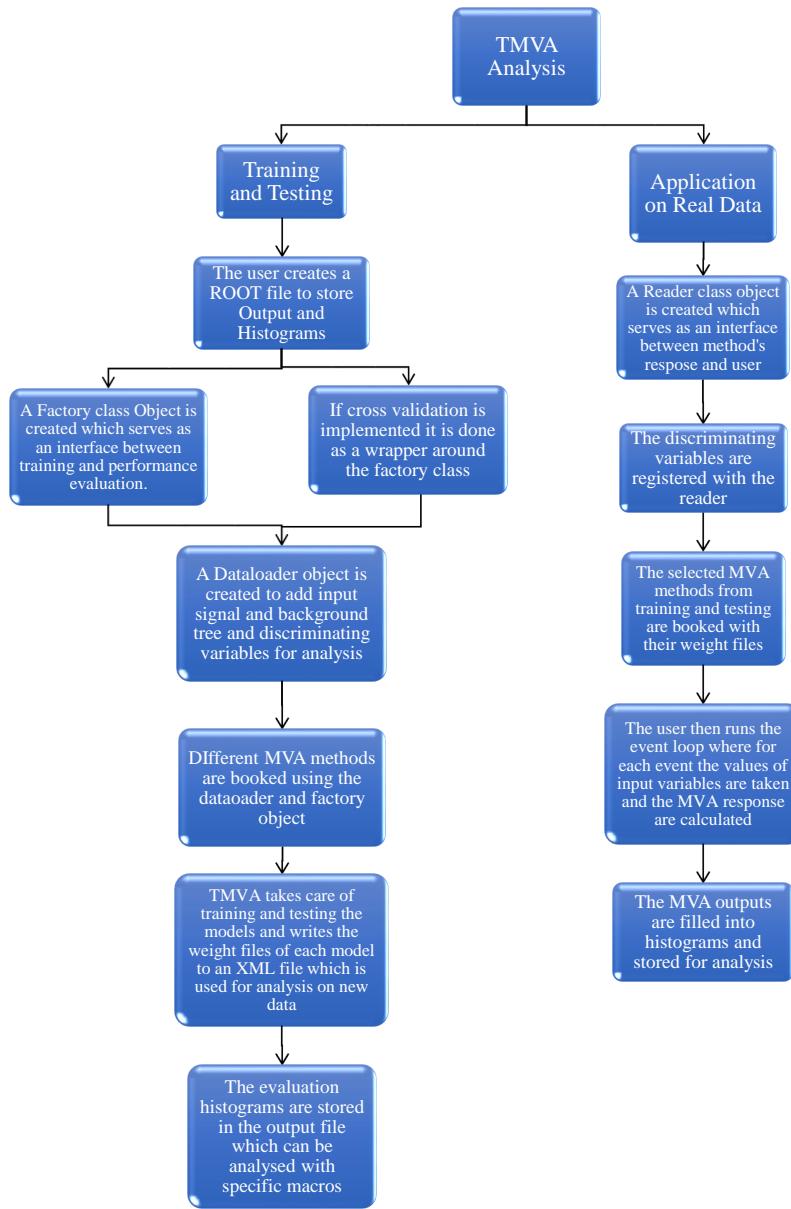


Figure 4.1: Flowchart showing steps involved in using TMVA for analysis

## 4.1 Boosted Decision Tree

Boosted Decision Tree (BDT) are one of the most popular ML algorithms used in classification problems in different fields. It is an example of a supervised learning algorithm which means that during the training phase the output for a particular input is known and mapped and the algorithm is trained on it. It was also used in the search for Higgs Boson at Compact Muon Solenoid (CMS) detector in CERN[2] and by the LHCb group for the search for pentaquarks[35].

A decision tree is a binary tree structured classifier which means that repeated yes/no or left/right decision are taken on a variable until the stop criterion is met. Each decision taken on the input is classified into what is called a leaf node. Each node can be classified as signal(+1) like or background(-1) like. The leaf nodes are classified as signal or background depending on the purity  $p = S/S + B$  where  $S$  is the number of Signal events and  $B$  is the number of background events. At each layer a single training variable is used to find the cut value which provides best separation between signal and background. That variable is divided up into NCuts. The BDT scans each interval to determine whether it looks signal like or background like and uses that to determine that nodes purity. A variety of separation criteria can be configured to assess the performance of a variable and a specific cut requirement. The criteria's are : Gini Index, Cross entropy, Misclassification error, Statistical significance and Average squared error. We have used Gini Index in our analysis [57]. BDT is prone to overtraining and the same variable may be used at several nodes, while others might not be used at all. This can

lead to bias when applied to real data which happened in this analysis. So usually a collection of decision trees is used which is called a forest.

In order to improve the performance of the classifier a procedure called boosting is used which is a way of enhancing the classification performance of a weak MVA methods by sequentially applying an MVA algorithm to reweighted (boosted) versions of the training data and then taking a weighted majority vote of the sequence of MVA algorithms thus produced [57]. There are different types of boosting methods and the one used in this analysis is called Adaptive Boost(AdaBoost). In this procedure events that were misclassified during the training of a decision tree are given a higher event weight in the training of the following tree. The boost weight( $\alpha$ ) is derived from the misclassification rate,  $err$ , of the previous tree

$$\alpha = \frac{1 - err}{err} \quad (4.1)$$

where  $err$  is the ratio of total weight of mistakes to total weights of all points. The result of an individual classifier is defined as  $h(x)$  with  $x$  being the tuple of input variables.  $h(x)$  takes value +1 for signal and -1 for background. The boosted event classification ( $y_{Boost}(x)$ ) is then given by

$$y_{Boost}(x) = \frac{1}{N_{collection}} \cdot \sum_i^{N_{collection}} \ln(\alpha_i) \cdot h_i(x) \quad (4.2)$$

where  $N_{collection}$  is the collection of weak classifiers. Small (large) values for  $y_{Boost}(x)$  indicate a background-like (signal-like) event.

## 4.2 Multi Layer Perceptron

Multi Layer Perceptron (MLP) is a type of Artificial Neural Network (ANN) in which the neurons are organised in layers and direct connections are allowed from a given layer to the following layer. All neural net implementations in TMVA are of this type. The first layer of a MLP is the input layer, the last one the output layer, and all others are hidden layers. For a classification problem with  $n_{var}$  input variables the input layer consists of  $n_{var}$  neurons that hold the input values,  $x_1, \dots, x_{nvar}$ , and one neuron in the output layer that holds the output variable, the neural net estimator  $y_{ANN}$ .

TMVA provides two methods for training of the neural network. The most commonly used training algorithm for neural networks is the Back-propagation (BP) algorithm is used to find the gradient of the cost function with respect to the weights of the network. This is done by using chain rule in which the derivative is calculated from the last layer to the first layer, which can be thought of as moving backwards which leads to its name. Back propagation is used by all neural networks in TMVA. After getting the cost function stochastic gradient descent optimisation technique is used to minimise the cost function which results in a trained network. The other method provided by TMVA for training is called the Broyden-Fletcher-Goldfarb-Shannon(BFGS) method [58]. differs from back propagation by the use of second derivatives of the error function to adapt the synapse weight. The advantage of the BFGS method compared to BP is the smaller number of iterations. But for large networks computing time increases.

A good neural network is always a trade off between using a network architecture that is flexible enough to allow for the modelling of complex functions and avoiding overtraining. The Bayesian extension of MLP offers a means to allow for more complex network architectures while at the same time regularizing the model complexity adaptively to avoid unwanted overfitting effects [57]. But this increases the computing time. Enabling the Bayesian extension of the MLP adds an additional term to the network error function. In this analysis we have trained the network with BFGS method with Bayesian regulators which is called the Multilayer Perceptron Bayesian Neural Network(MLPBNN).

# Chapter 5

## Experimental Analysis

This chapter discusses the results of the reconstruction of the pentaquark state in  $\Lambda K_s^0$  decay channel using the KFParticle Reconstruction Code and TMVA BDT and MLPBNN algorithms. We have used the Au+Au  $39GeV$  STAR dataset for our analysis. Let us first look at some information about the dataset.

### 5.1 Dataset Information

The Dataset used for analysis in this thesis were collected by STAR experiment at RHIC in the year 2010 for Au+Au collision at  $\sqrt{s_{NN}} = 39GeV$ . The analysis uses  $\mu$ Dst ("MicroDst or MuDst") files which contain the information about each Au ion collision. The information about each Au+Au ion collision is known as event. The MuDst files contains the momentum

and trigger information. The total number of runs were 13950. The dataset contains approx. 255.7 million events out of which approx. 175.2 million events were used in this analysis. The dataset information can be seen in the table Table 5.1.

| Trigger setup name | Stream name | Collision | Year of data taken | Production Tag | Trigger label | Offline trigger ID | Number of Events | Run Number        |
|--------------------|-------------|-----------|--------------------|----------------|---------------|--------------------|------------------|-------------------|
| AuAu39_production  | st_physics  | auau39    | 2010               | P10ik          | MinBias(mb)   | 280001             | ~255.7 Million   | 11098073-11112023 |

Table 5.1: Data-set and trigger information for Au+Au collisions at  $\sqrt{s_{NN}} = 39$  GeV [59]

### 5.1.1 Event Quality Assurance

The Event quality assurance ensures the selection of good events for analysis. The events were required to have a primary vertex position along the longitudinal direction ( $V_z$ ) within  $\pm 40$  cm from the center of the TPC detector. A cut of less than 2cm from the center of the beam pipe on the radial direction of the primary vertex( $V_r = \sqrt{V_x^2 + V_y^2}$ ) is applied to reduce effects of beam and beam pipe interactions.

| System/Energy | $V_z$ (cm)   | $V_r$ (cm) |
|---------------|--------------|------------|
| Au+Au, 39 GeV | $ V_z  < 40$ | $V_r < 2$  |

Table 5.2: Event Selection Cuts for Au+Au collisions at  $\sqrt{s_{NN}} = 39$  GeV

### 5.1.2 Centrality Selection

In collision experiments the impact parameter ( $b$ ) which is the perpendicular distance between the centres of the two colliding nuclei is an important

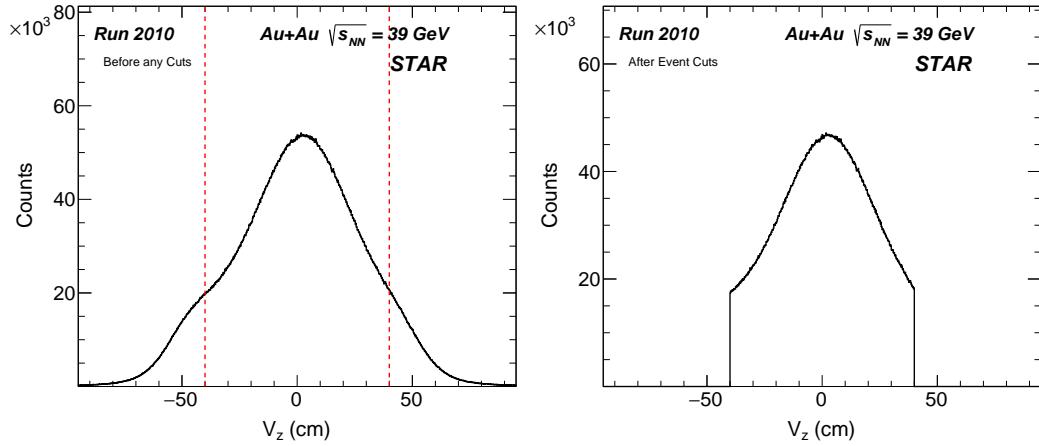


Figure 5.1: (Left) primary vertex position along the longitudinal direction distribution( $V_z$ ) before event cuts. (Right) primary vertex position along the longitudinal direction distribution( $V_z$ ) after event cuts

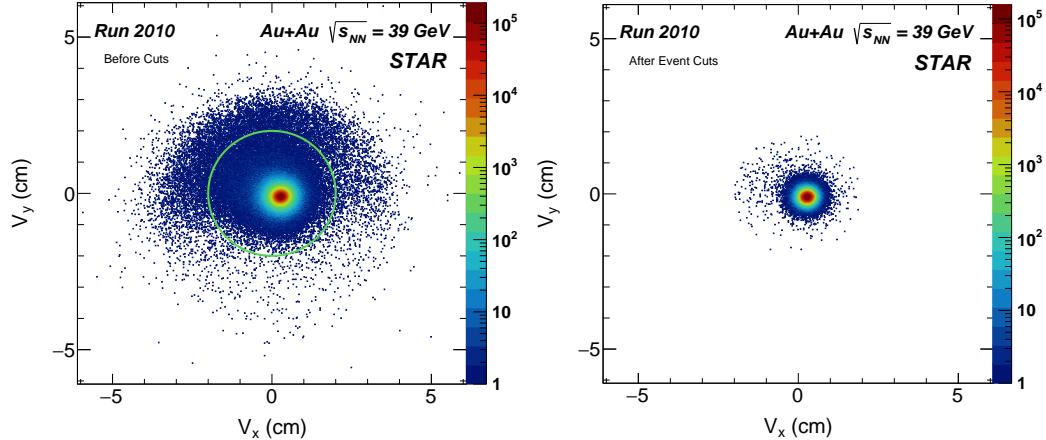


Figure 5.2: (Left) primary vertex position along radial direction distribution( $V_r$ ) before event cuts. (Right) primary vertex position along radial direction distribution( $V_r$ ) after event cuts

parameter to distinguish between different types of collision. The collisions in which the impact parameter is low is said to be central collision while events with increasing impact parameters are said to be semi central,semi

peripheral and peripheral respectively. Information about the centrality can give information about the interaction taking place in the collision and help us to understand its properties.

The impact parameter cannot be measured directly in experiment. As a result of this the centrality has to be inferred and this is based on the fact (which is supported by observational evidence) that particle multiplicity of an event increases monotonically with decreasing impact parameter. As a result of this the centrality can be determined by a quantity which is experimentally observable called the reference multiplicity( $N_{ch}$ ). The reference multiplicity is defined as the number of primary charged tracks from the TPC detector with number of hit fit points greater than and equal to 15, pseudo-rapidity <sup>1</sup> range of  $|\eta| < 0.5$  , distance of closest approach (DCA) to the primary vertex less than 3 cm and  $V_z < 40$  cm. The centrality is divided to bins which correspond to the fraction of the measured total cross-section. If there was a perfect accelerator and a perfect detector, the multiplicity range for the centrality determination could be measured accurately by integrating the multiplicity distribution but this is not the case and significant efforts are made to make correction to the particle multiplicity distribution and these have been done previously for the datasets in STAR and stored in the "StReffMultCorr" package and one can use it to get the centrality classification of the events [60].

---

<sup>1</sup> $\eta = -\ln[\tan(\theta/2)]$  and  $\theta = \tan^{-1}(\frac{p_T}{p_z})$

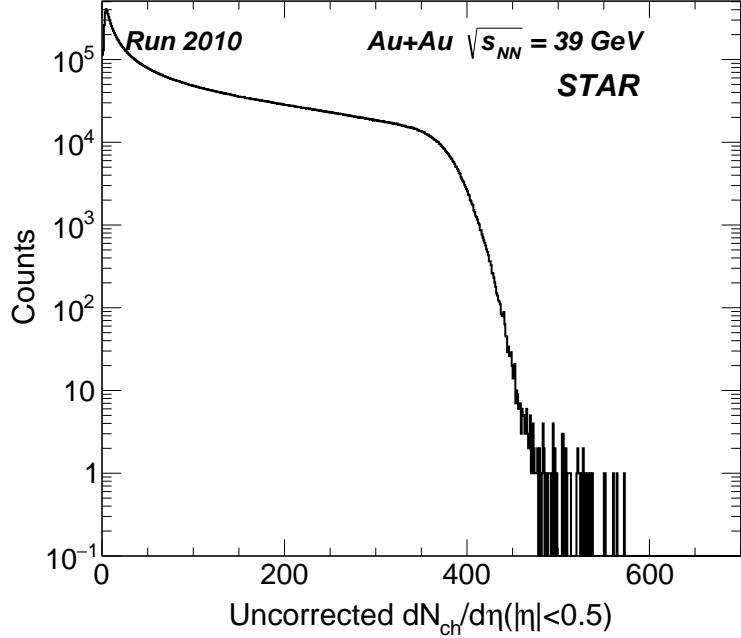


Figure 5.3: Uncorrected charged particle multiplicity for  $|\eta| < 0.5$

### 5.1.3 Track Quality Assurance

To ensure that the charged particles produced from the collision are reconstructed from the collision by the tracking algorithm several track selection cuts were added to select the tracks. These are added to the process event function of the StKFParticleInterface class. The short tracks are removed by selecting tracks which have atleast 15 hit fit points in the TPC. The same tracks which were misidentified as two separate tracks by the tracking algorithm were removed by selecting tracks with number of fit points to maximum number of all possible fit points ratio  $> 0.52$ . Tracks with number of hit points greater than 10 hits are selected to improve the particle identification with measure of ionization energy loss ( $dE/dx$ ) from TPC detector.

The pseudorapidity ( $\eta$ ) range is selected as  $|\eta| < 1$  and the transverse momentum ( $p_T$ ) is selected as  $0.15 < p_T < 10\text{GeV}/c$  and additional cuts are added by the KFParticle on the  $dE/dx$  resolution of the TPC detector which is  $0.04 \leq dedxErrorFit \leq 0.12$ .

| Track Parameter                 | Cuts Applied                       |
|---------------------------------|------------------------------------|
| Number of fit points            | $nHitsFit > 15$                    |
| $nHitsFit/nHitsPoss$            | $\text{Ratio} > 0.52$              |
| Transverse Momentum             | $0.15 < p_T < 10$                  |
| Pseudorapidity                  | $ \eta  < 1$                       |
| Number of fit points for $dEdx$ | $nHitsFit dEdx > 10$               |
| $dE/dx$ resolution              | $0.04 \leq dedxErrorFit \leq 0.12$ |

Table 5.3: Track Parameter cuts for Au+Au collisions at  $\sqrt{s_{NN}} = 39\text{ GeV}$

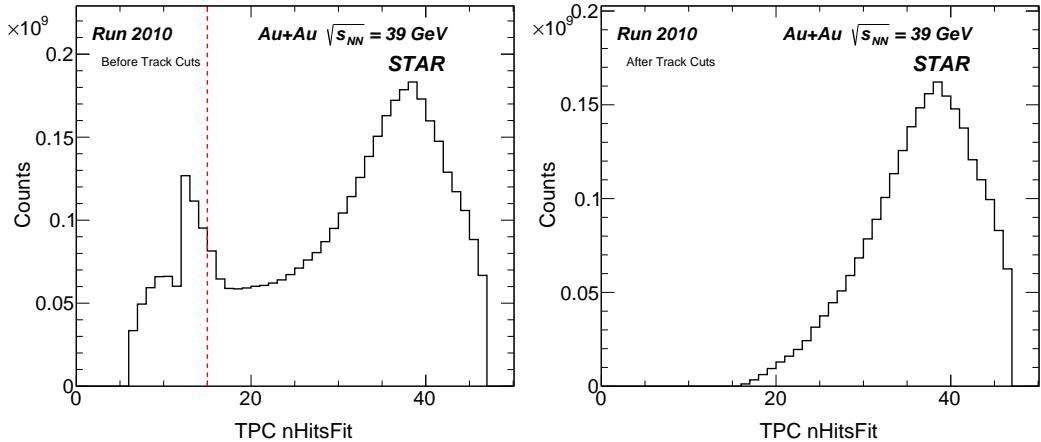


Figure 5.4: (Left) Distribution of TPC Hits before track cuts.(Right) Distribution of TPC Hits after track cuts.

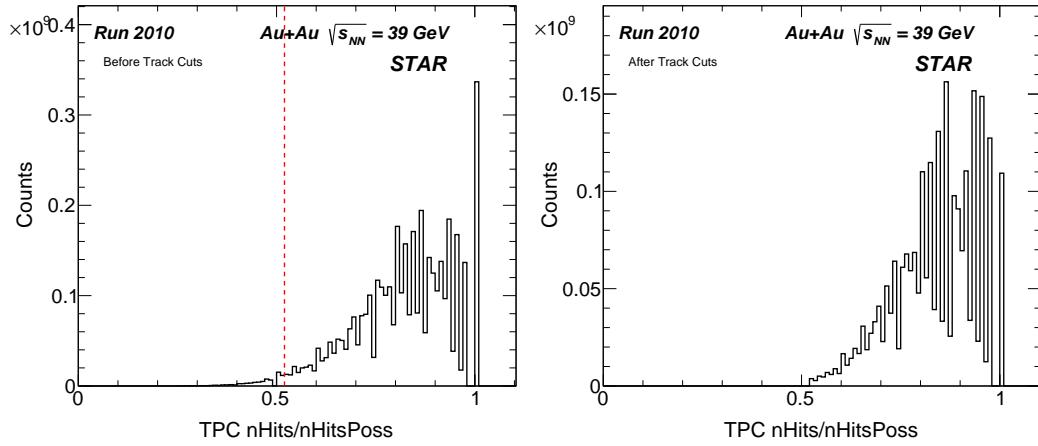


Figure 5.5: (Left) Distribution of ratio of TPC Hits and possible Hits before track cuts.(Right) Distribution of ratio of TPC Hits and possible Hits after track cuts.

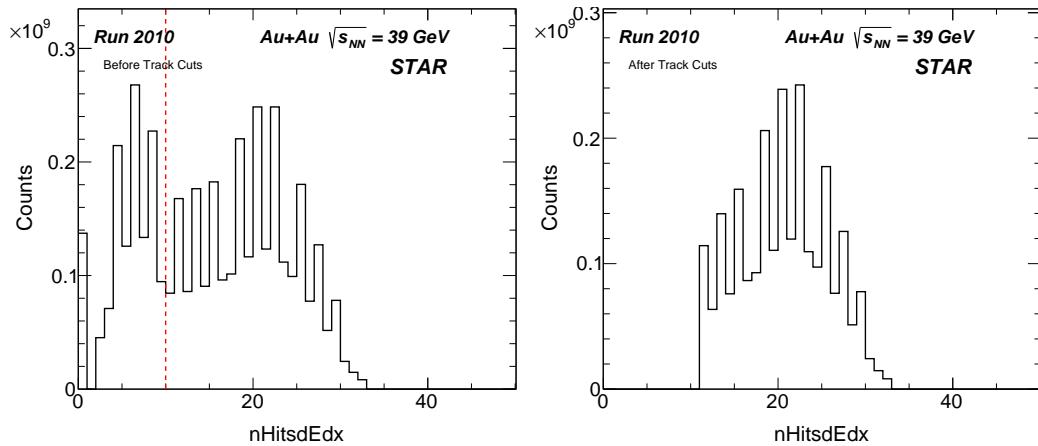


Figure 5.6: (Left) Distribution of dEdx Hits before track cuts.(Right) )Distribution of dEdx Hits after track cuts

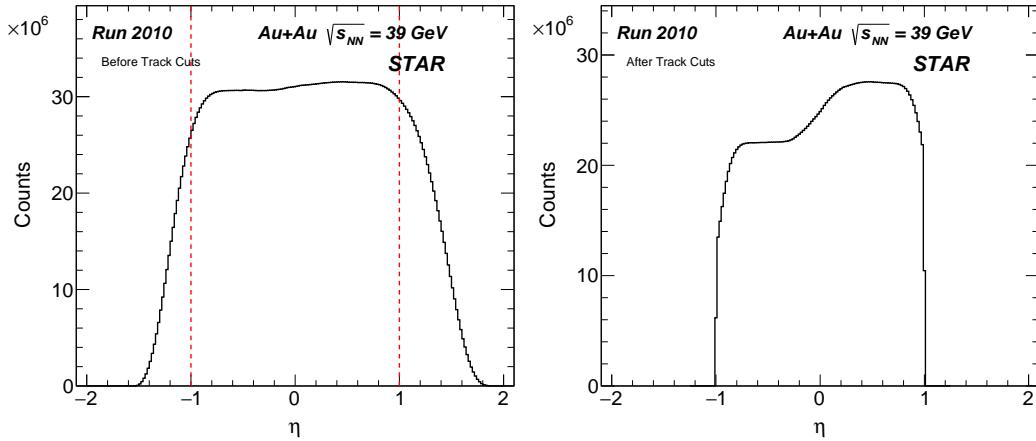


Figure 5.7: (Left) Distribution of Pseudorapidity before track cuts.(Right) Distribution of Pseudorapidity after track cuts

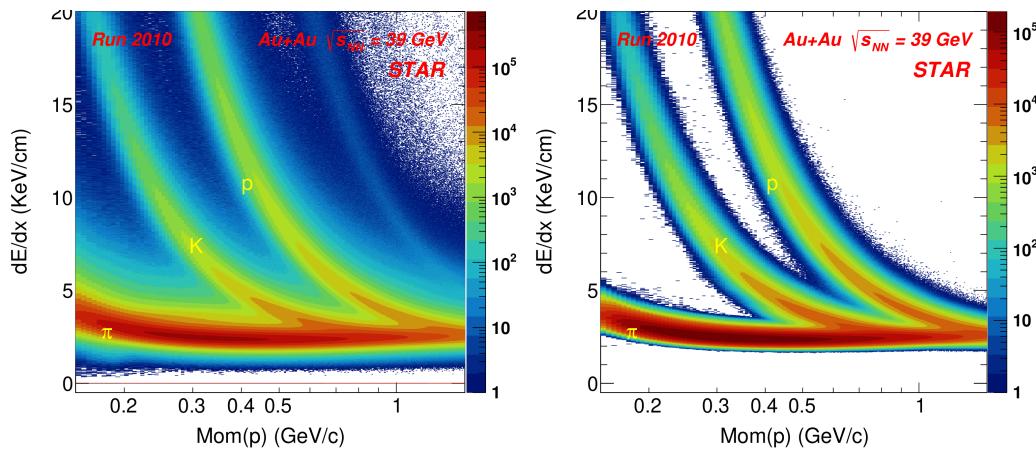


Figure 5.8: (Left)Energy loss of charged particles in TPC vs Momentum before track cuts.(Right)Energy loss of charged particles in TPC vs Momentum after cuts

## 5.2 KFParticle Analysis

In this section we discuss the default KFParticle cuts used to reconstruct the PQ particle which is referred to as the parent particle and various properties of the parent particle which can be calculated using the KFParticle package. First the  $V^0$  particles  $\Lambda^0$  and  $K_s^0$  are reconstructed using the grand daughter tracks of pions( $\pi^+, \pi^-$ ) and protons( $p$ ). The PQ mother particle is then reconstructed from the  $V^0$  particles considering  $\Lambda^0$  and  $K_s^0$  as the daughter particles. The schematic representation for the PQ decay is shown in Figure 5.9. The KFParticle cuts are discussed in the table (Table 5.4)

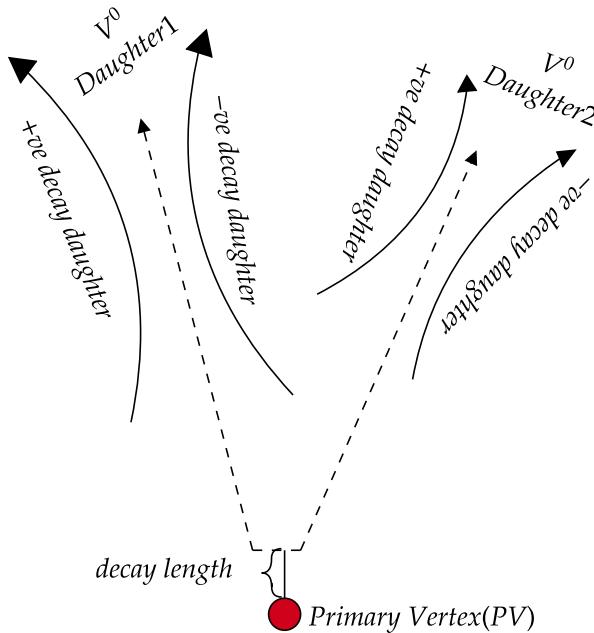


Figure 5.9: Schematic diagram for PQ reconstruction and decay

| Cut Parameter   | Cuts Description   | $\Lambda^0$            | $K_s^0$                | PQ                     |
|-----------------|--|------------------------|------------------------|------------------------|
| $\chi_{prim}^2$ | Criteria to separate tracks into primary and secondary           | $\chi_{prim}^2 > 18.6$ | $\chi_{prim}^2 > 18.6$ | $\chi_{prim}^2 > 18.6$ |
| $\chi_{fit}^2$  | $\chi^2$ of daughter trajectories intersecting                   | $\chi_{fit}^2 < 3$     | $\chi_{fit}^2 < 3$     | $\chi_{fit}^2 < 3$     |
| $\chi_{topo}^2$ | $\chi^2$ of that two tracks originate from the same decay vertex | $\chi_{topo}^2 < 3$    | $\chi_{topo}^2 < 3$    | $\chi_{topo}^2 < 3$    |
| $l/dl$          | decay length normalised on its error                             | $l/dl > 10$            | $l/dl > 10$            | $l/dl < 3$             |

Table 5.4: KFParticle Cuts for Au+Au collisions at  $\sqrt{s_{NN}} = 39$  GeV

### 5.2.1 Armenteros Podolanski Plot

The Armenteros Podolanski plot [61] is a method of knowing the type of the reconstructed  $V^0$  particle. It is a two dimensional plot, of Armenteros transverse momentum  $p_t^{Arm}$  of the oppositely charged decay products with respect to the  $V_0$  versus the longitudinal momentum asymmetry( $\alpha$ ).

$$\alpha = \frac{p_L^+ - p_L^-}{p_L^+ + p_L^-} \quad (5.1)$$

$$p_t^{Arm} = p^{+,-} \sin(\phi^{+,-}) \quad (5.2)$$

where  $p_L^\pm = p^\pm \cos(\phi^\pm)$  and  $\phi^\pm = \cos^{-1}\left(\frac{p_{V^0} \cdot p^\pm}{|p_{V^0}| |p^\pm|}\right)$  as can be seen in the diagram Figure 5.10

In case of  $K_s^0$  we can see that the center of the ellipse is at  $\alpha = 0$  since  $K_s^0$  decays to equal mass pions while in case of  $\Lambda$  decay there is difference in mass of the decay products. The Armenteros plot for the reconstructed

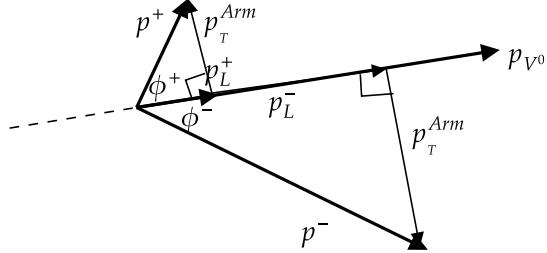


Figure 5.10: Calculation of  $p_T^{Arm}$

$V^0$  and PQ particles can be seen in Figure 5.11, Figure 5.12, Figure 5.13. The graphs for real data are plotted for a single response cut  $> 0.03$  for MLPBNN trained with data as background. These plots are obtained by using GetArmenterosPodolanski function of the KFParticle package.

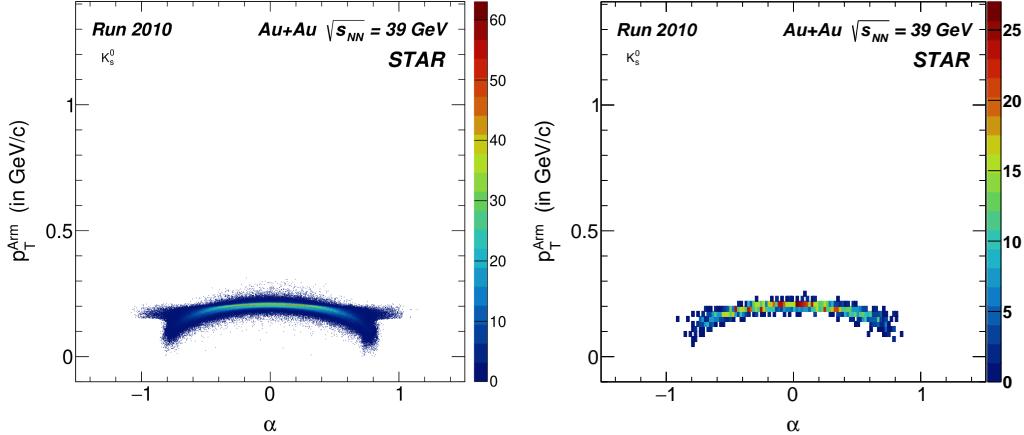


Figure 5.11: (Left) Armenteros-Podolanski plot for  $K_s^0$  for real data after MLPBNN response cut  $> 0.03$ (Data as Bkg) . (Right) Armenteros-Podolanski plot for  $K_s^0$  for 1PQ/event simulation

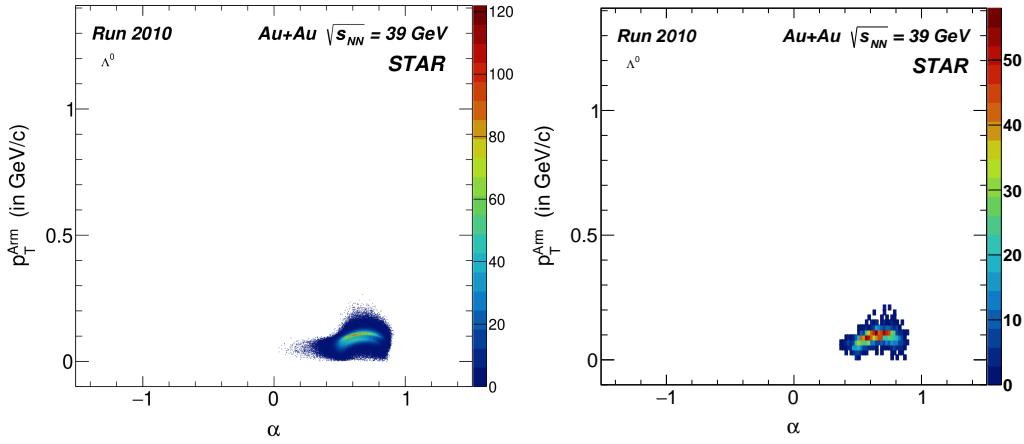


Figure 5.12: (Left) Armenteros-Podolanski plot for  $\Lambda^0$  for real data after MLPBNN response cut  $> 0.03$ (Data as Bkg). (Right) Armenteros-Podolanski plot for  $\Lambda^0$  for 1PQ/event simulation

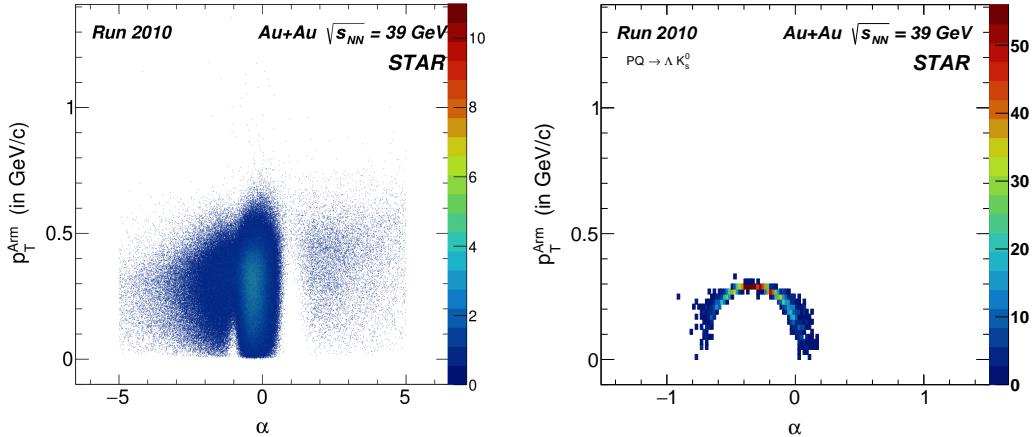


Figure 5.13: (Left) Armenteros-Podolanski plot for reconstructed PQ-candidates from real tracks after MLPBNN response cut  $> 0.03$ (Data as Bkg). (Right) Armenteros-Podolanski plot for reconstructed PQ from 1PQ/event simulation tracks

### 5.2.2 TMVA Analysis

In this section we present the results of the application of BDT and MLPBNN after training it on 1PQ per event Monte Carlo Simulation data simulated using the GEANT4 Simulation toolkit[62]. The parameters used for the generation of simulated 1PQ/event are mass =  $1.73 \text{ GeV}/c^2$ , temperature ( $T$ ) =  $0.4255 \text{ GeV}$ . The mean transverse momentum  $\langle p_T \rangle \approx 1.3 \text{ GeV}/c$  for the simulated PQ state. The mass of  $\omega$  particle is  $1.69 \text{ GeV}/c^2$  and the mean transverse momentum of  $\omega$  particle is  $\langle p_T \rangle = 1.29 \pm 0.14 \text{ GeV}/c$ . Since the mass of the  $\omega$  particle is comparable to our expected mass range the temperature  $T = 0.4255 \text{ GeV}$  is used. The 1PQ/event is used as the signal tree for training in TMVA. For Background tree we used 20  $K_s^0$  and 20  $\Lambda^0$  per event which are not from the same parent particle. The temperature used for generating the data is  $0.325 \text{ GeV}$  and the mean transverse momentum  $\langle p_T \rangle = 0.8 \text{ GeV}/c$ . The simulation data are generated in the form of  $\mu\text{DsT.root}$  files which contain the information about the generated tracks and momentum information. These are then used to reconstruct the PQ state and its mass is calculated by invariant mass reconstruction using the  $\Lambda^0$  and  $K_s^0$  momentum and mass information as shown in Equation 5.3.

$$m_{inv} = \sqrt{(\sqrt{p_1^2 + m_1^2} + \sqrt{p_2^2 + m_2^2}) - (\vec{p}_1 + \vec{p}_2)^2} \quad (5.3)$$

where  $\vec{p}_1, \vec{p}_2$  are the momentum vectors and  $m_1, m_2$  are the rest masses of the daughter particles. The information about the generated data is given in the table below Table 5.5

| Dataset                                  | Total Events Generated | Total PQ reconstructed for training and testing |
|--|------------------------|---|
| 1PQ/Event Signal                         | 2.43 Million           | 1761  |
| 20 $K_s^0$ and 20 $\Lambda^0$ Background | 101420                 | 43136   |
| Data For Background                      | -                      | 40000   |

Table 5.5: Dataset used for TMVA Training and Testing

### MLPBNN Results for $20\Lambda^0$ and $20K_s^0$ as Background

The MLPBNN is trained using 24 variables and 5 fold crossvalidation. The cuts used on the dataset is same as given in Table 5.4. The  $tanh$  function is used as the neuron activation function. The training is done for 60 training cycles since it is computationally expensive method. The number of hidden layers used are 5 more than the number of input variables used which is 29. The test for overtraining is done on after each 5th epochs (where epoch means training the network once with all training data for one cycle). The variables used for training and testing are listed below:

- Chi2fitPQ
- QtPQ
- DcaABKs
- Chi2topoPQ
- AlphaPQ
- DecayLenKs
- ldlPQ
- Chi2fitKs
- QtKs
- DcaPQ
- Chi2topoKs
- AlphaKs
- DcaABPQ
- ldlKs
- Chi2fitL
- DecayLenPQ
- DcaKs
- Chi2topoL

- ldlL
- DcaABL
- QtL
- DcaL
- DecayLenL
- AlphaL

After training the MLPBNN it is used on the real dataset to get the invariant mass histograms based on the response given by the classifier. The distributions for the variables used for training can be seen in the (Figure 5.14,Figure 5.15).

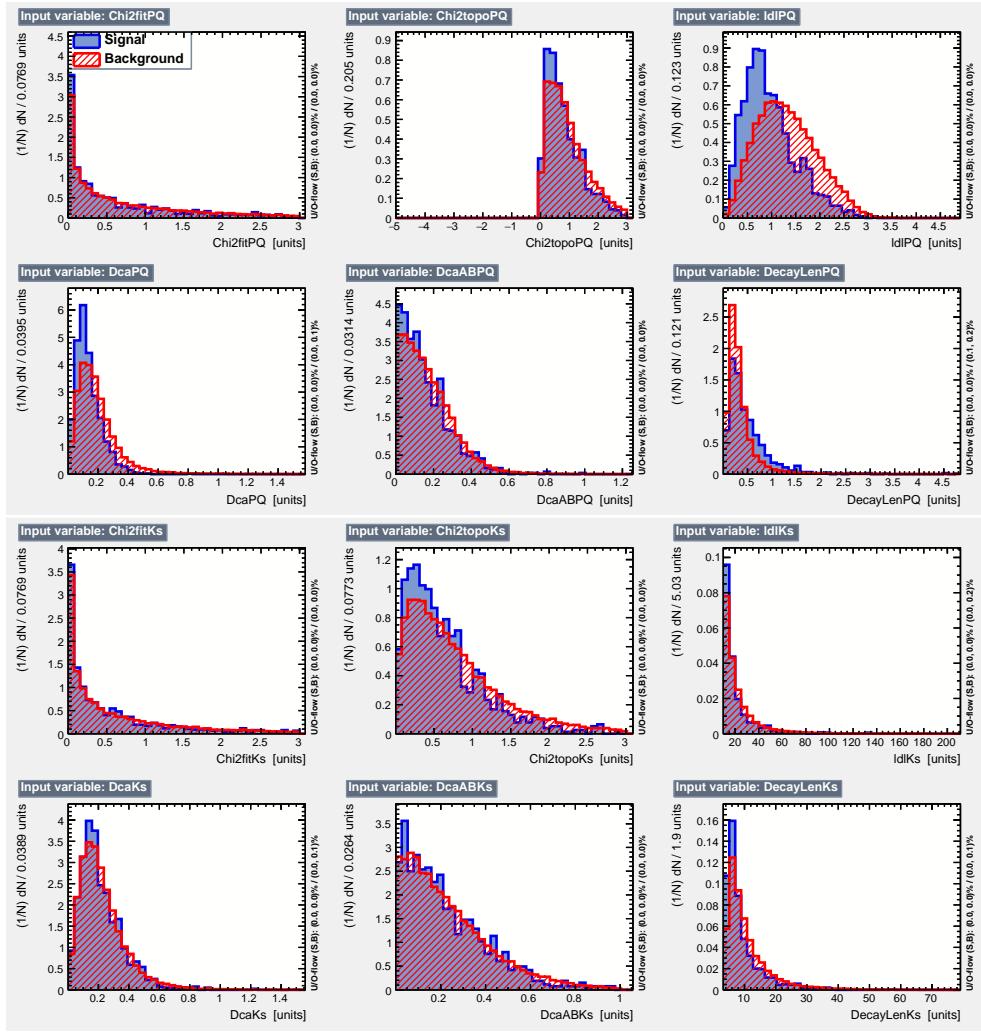


Figure 5.14: Distribution of the input variables used. Here Blue distribution represents the signal and red distribution represents the background

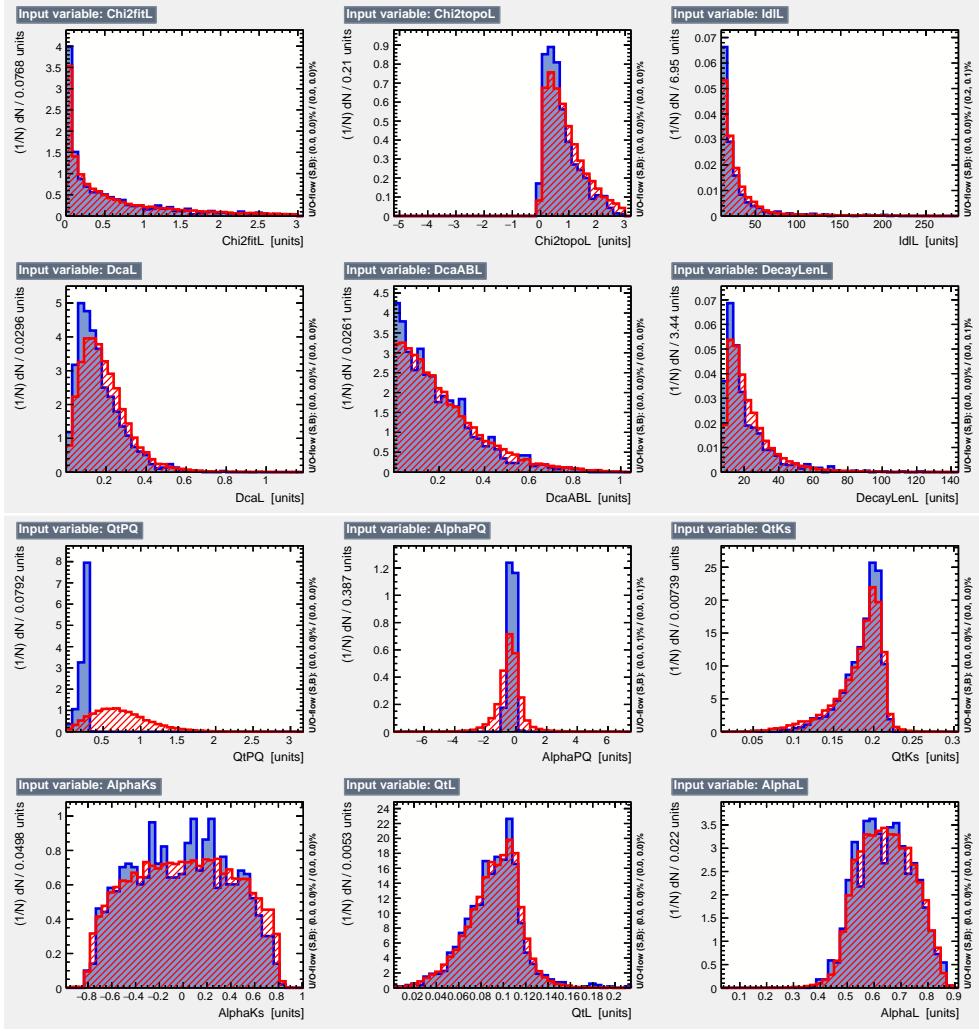


Figure 5.15: Distribution of the input variables used. Here Blue distribution represents the signal and red distribution represents the background

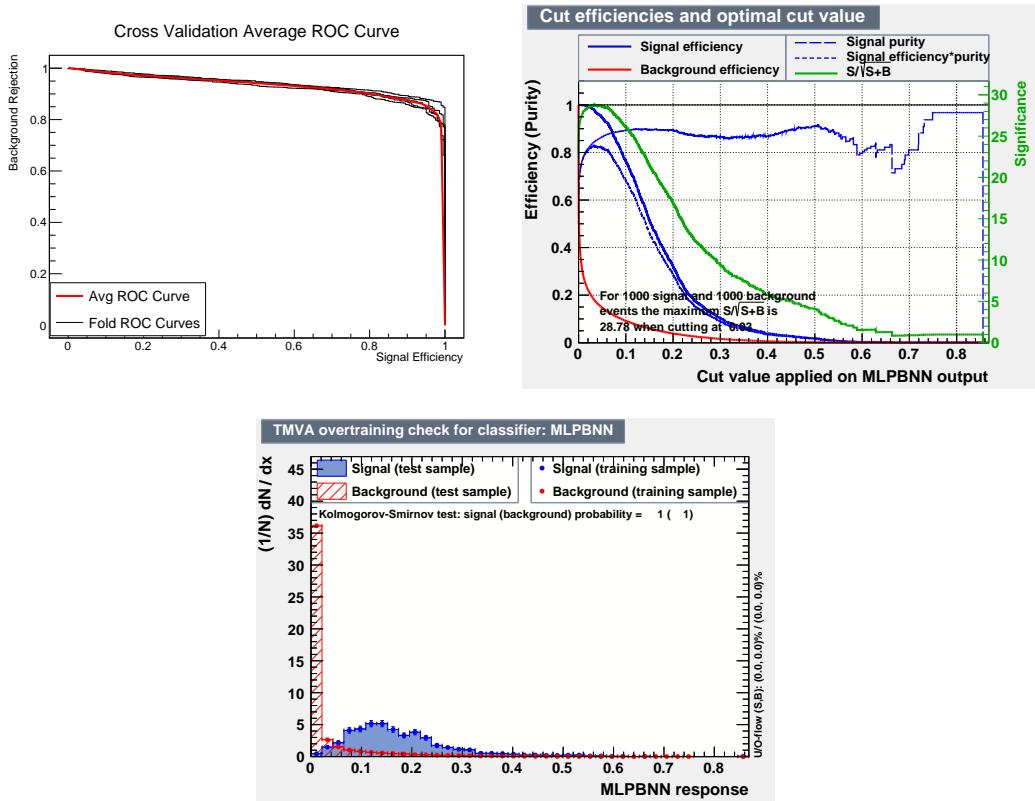


Figure 5.16: (TopLeft)ROC curve for MLPBNN method. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by MLPBNN. (Bottom) Overtraining check of the classifier.

The information about the training and testing of the classifier can be obtained by looking at the Receiver Operating Characteristic (ROC) curve given in top left Figure 5.16. This shows how well a classifier is trained as it shows the relation between background rejection and signal efficiency. An ideal classifier would be able to have background rejection as one and signal efficiency also 1. But in practice this is not the case and with increase in signal efficiency there is decrease in background rejection so a good classifier always try to maximise the area under the ROC curve. The purity is defined as the ratio of number of signal events to total number of events. As seen in the top right of Figure 5.16 the optimal cut value is calculated where the significance i.e,  $\frac{S}{S+B}$  is maximum. The bottom graph of Figure 5.16 shows the classifier output distribution and the overtraining check for the classifier. We see there is good agreement between the distributions for training and testing samples. The Kolmogorov-Smirnov test is a statistical test to check compatibility between distributions i.e if a sample comes from a specific distribution[63]. KS test is theoretically applied to unbinned data but TMVA approximates it for the case of binned data as a result of which we get the scores of 1 in this case. The total number of events on which the MLPBNN was applied was  $\approx 4.17$  Million and the total entries in the PQ invariant mass histogram for each response cut set is given in Table 5.6. We have used different MLPBNN response values from 0 to 0.1 in 9 sets to select the events for the invariant mass reconstruction of the PQ and we get the following graphs. We have plotted the invariant mass histogram in different centrality bins of 0-80%, 10-80%, 10-50% and 50-80%.

| <b>MLPBNN Response Cut</b> | <b>Total PQ Entries</b> |
|----------------------------|-------------------------|
| > 0                        | 3946119                 |
| > 0.01                     | 1404723                 |
| > 0.02                     | 1124731                 |
| > 0.03                     | 960027                  |
| > 0.04                     | 841675                  |
| > 0.05                     | 749198                  |
| > 0.07                     | 609528                  |
| > 0.09                     | 504524                  |
| > 0.1                      | 460881                  |

Table 5.6: Total events which pass the corresponding MLPBNN response cuts.

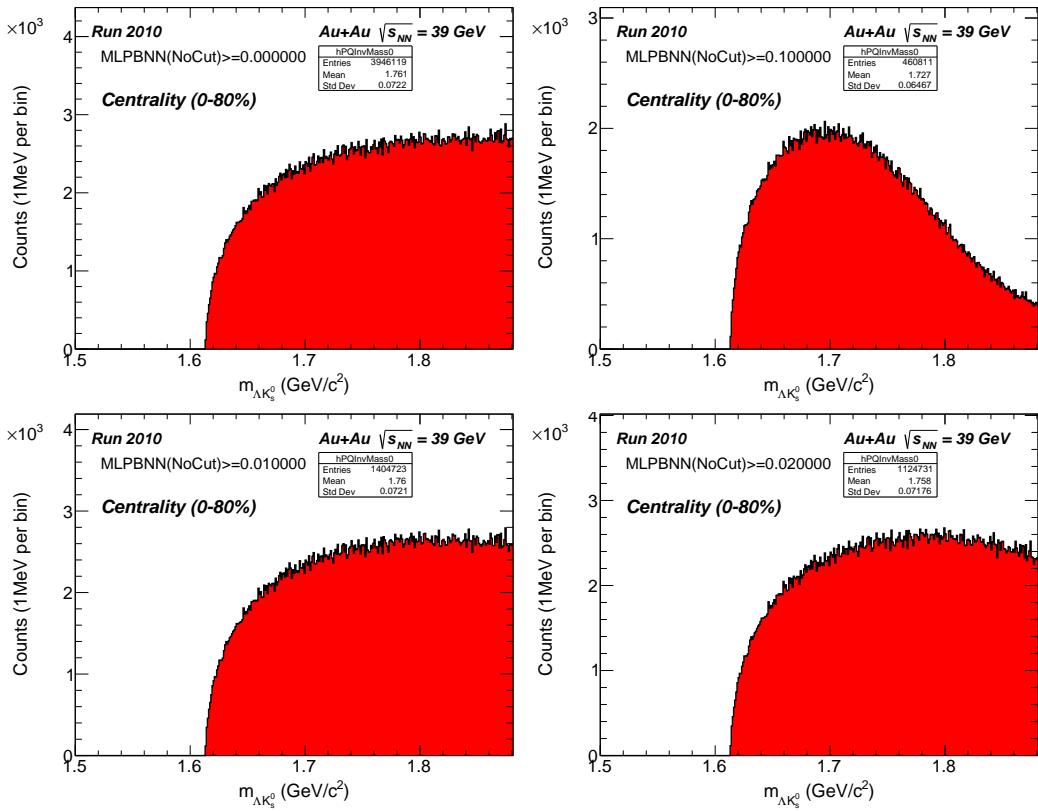


Figure 5.17:  $\Lambda^0 K_s^0$  invariant mass for centrality 0-80% after MLPBNN response cuts of (from top left to right)  $> (0, 0.1, 0.01 \text{ and } 0.02)$  with 1MeV per bin

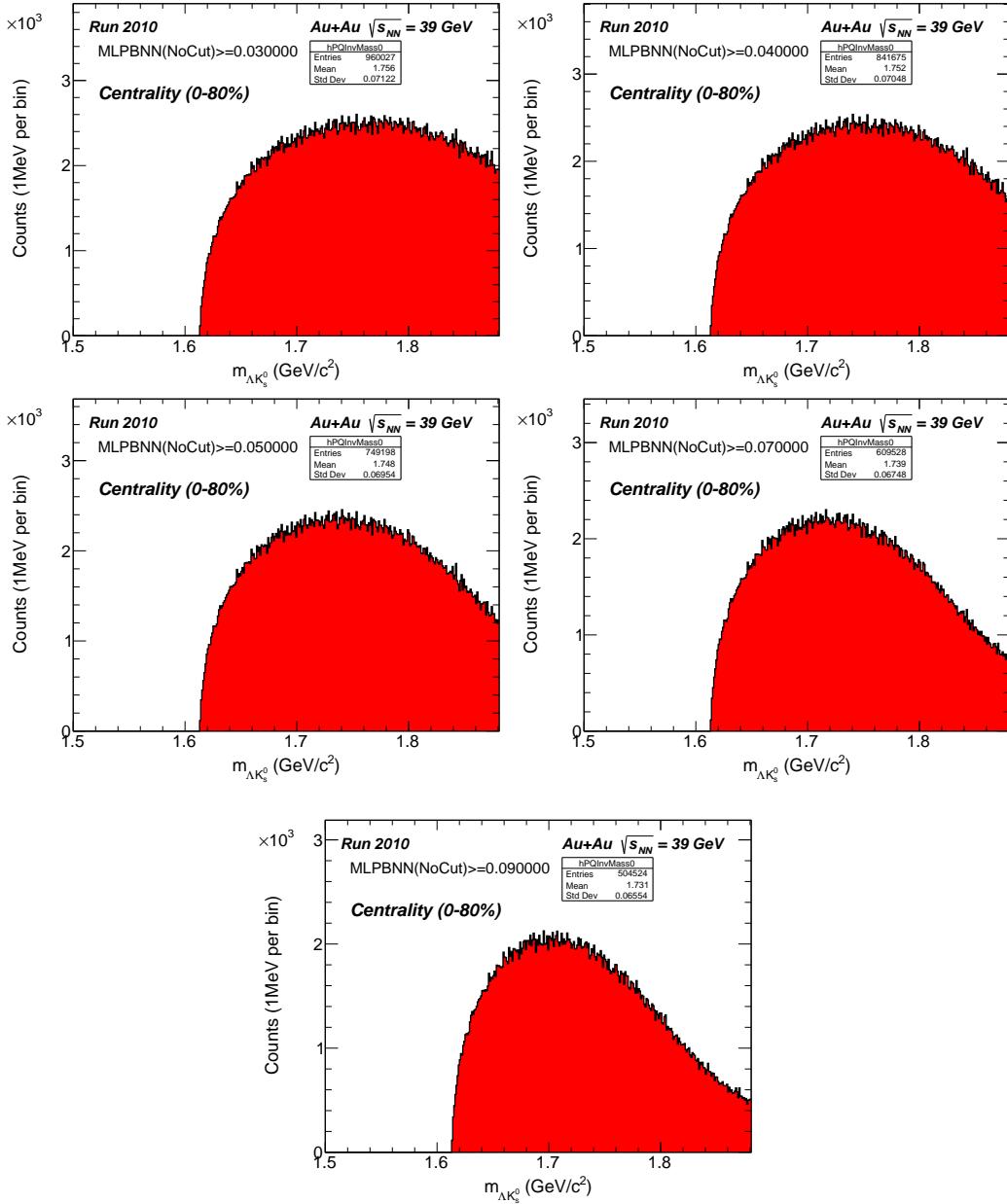


Figure 5.18:  $\Lambda^0 K_s^0$  invariant mass for centrality 0-80% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07$  and  $0.09)$  with 1MeV per bin

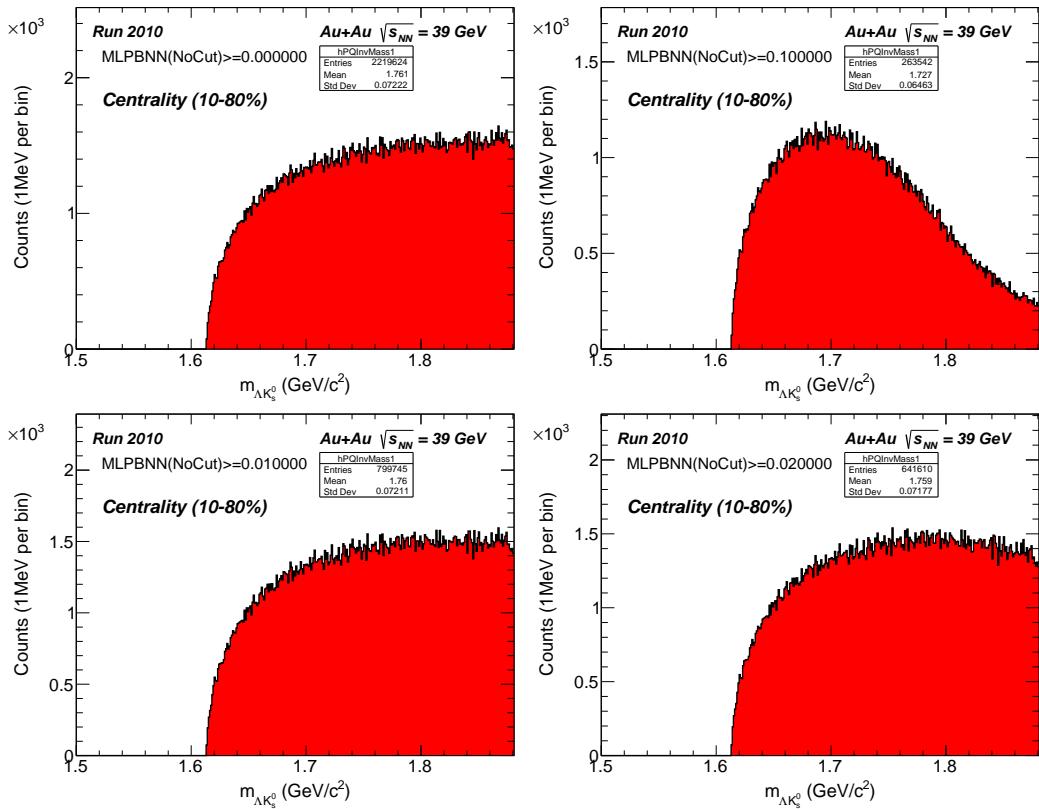


Figure 5.19:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-80% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0, 0.1, 0.01 \text{ and } 0.02)$  with 1MeV per bin

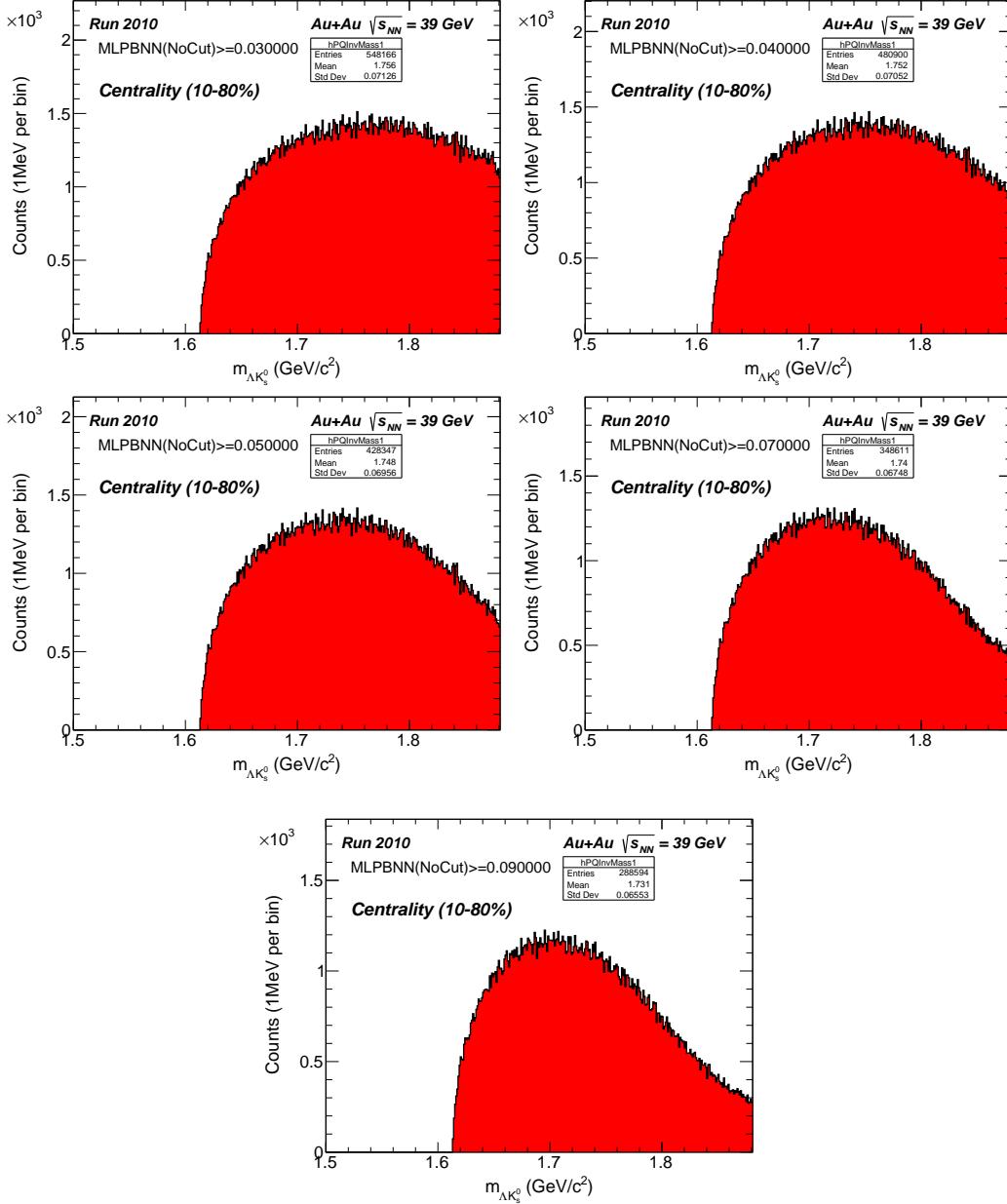


Figure 5.20:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-80% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

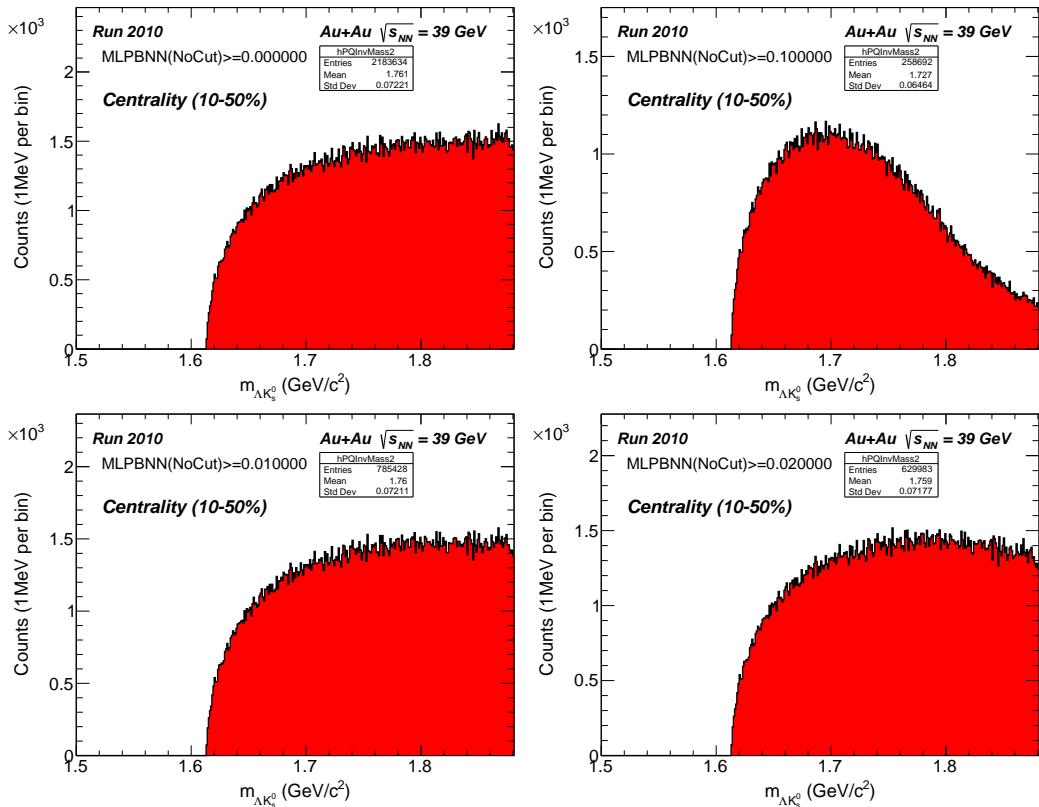


Figure 5.21:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-50% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0, 0.1, 0.01 \text{ and } 0.02)$  with 1MeV per bin

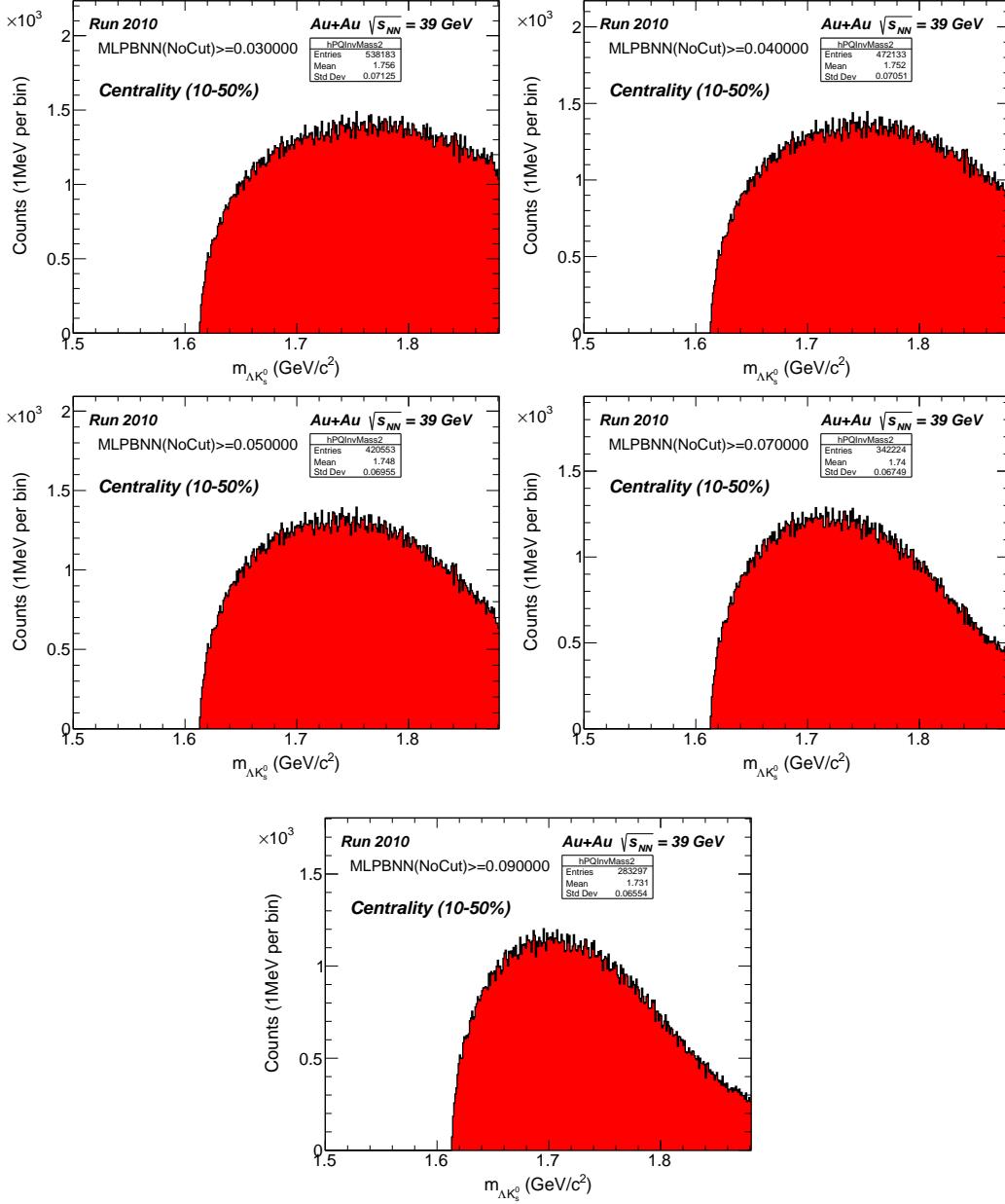


Figure 5.22:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-50% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

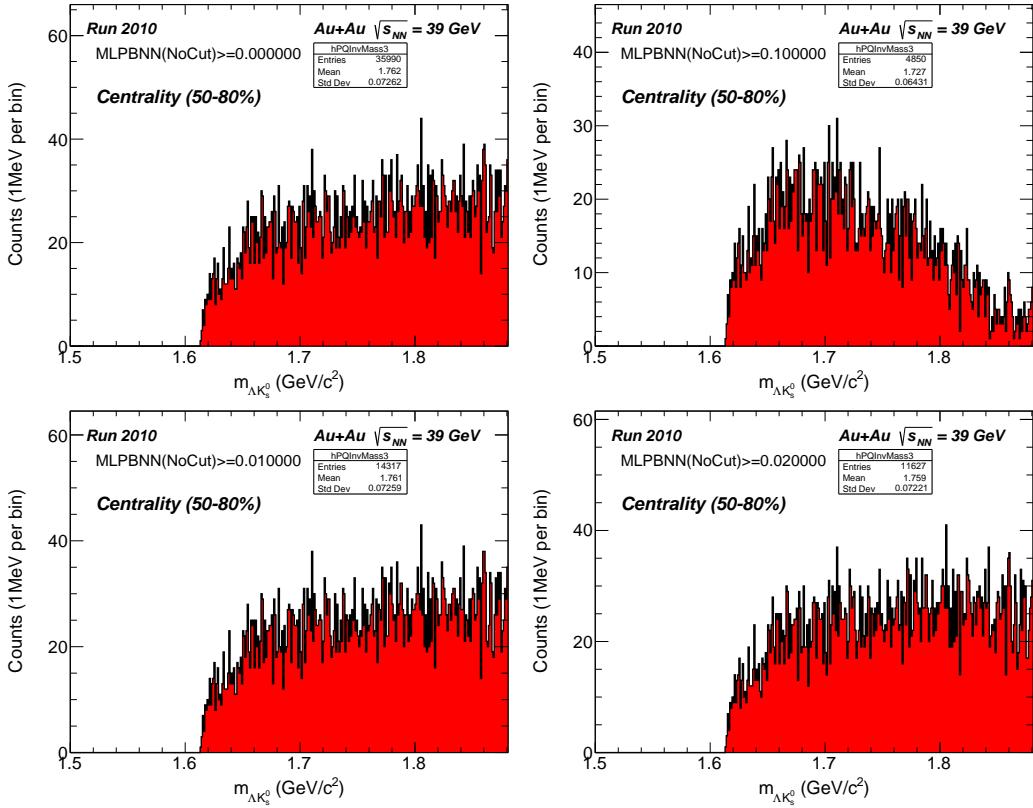


Figure 5.23:  $\Lambda^0 K_s^0$  invariant mass for centrality 50-80% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $\geq (0, 0.1, 0.01 \text{ and } 0.02)$  with 1MeV per bin

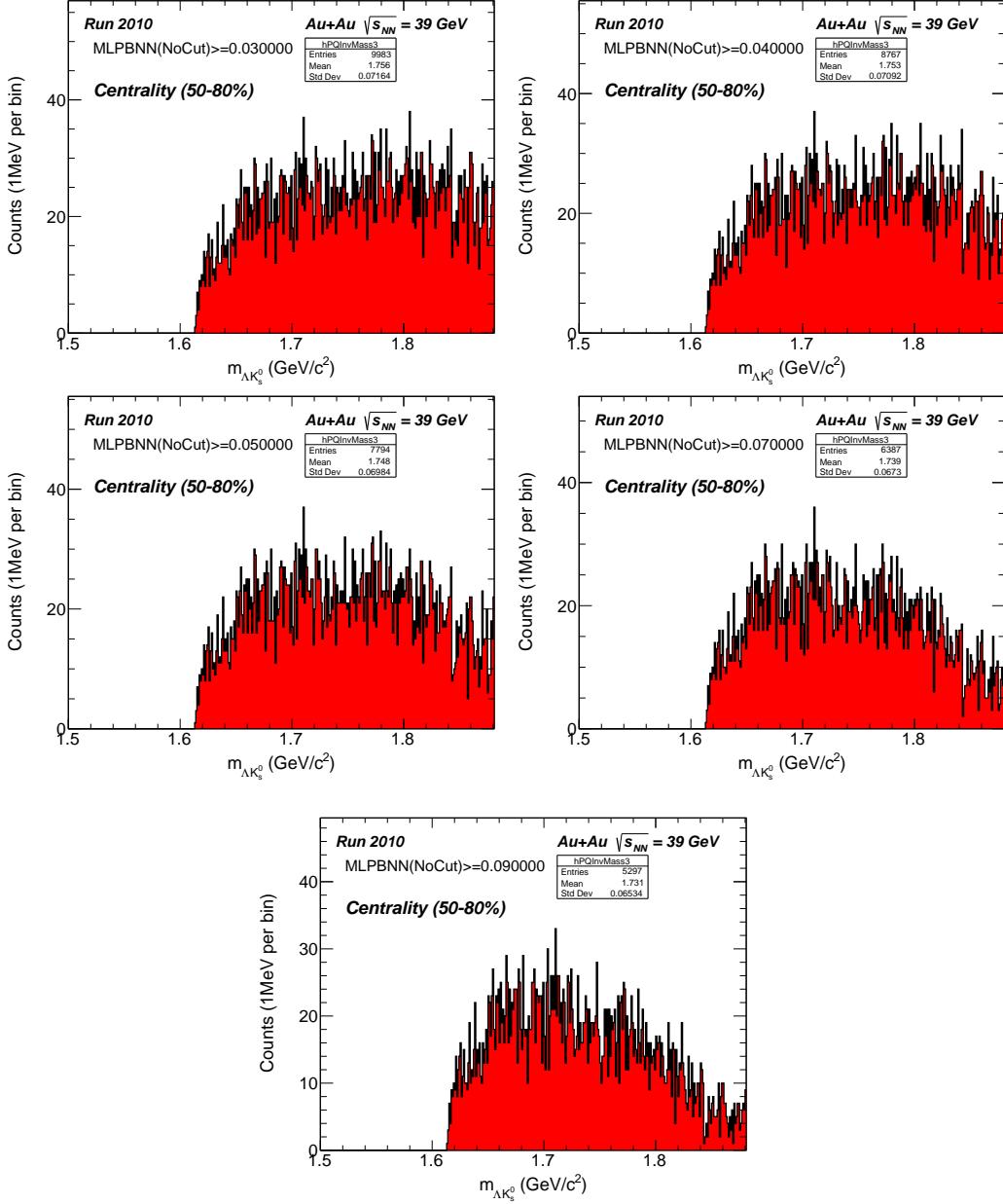


Figure 5.24:  $\Lambda^0 K_s^0$  invariant mass for centrality 50-80% after MLPBNN response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

## MLPBNN Results for Data as Background

The MLPBNN method is trained using the same variables listed in section 5.2.2 and the cut set given in Table 5.4. The neural network options are same as discussed in the previous section. We used real data as background tree for training the classifier and the training information can be found from the ROC and response cut graphs Figure 5.25. The total number of events on which the MLPBNN was applied was  $\approx 3.94$  Million since we exclude the data which we used as background to train to avoid biasing the classifier and the total entries in the PQ invariant mass histogram for each response cut set is given in Table 5.7. In this case also we plotted the  $\Lambda^0 K_s^0$  invariant mass histogram in different centrality bins (0-80%, 10-80%, 10-50% and 50-80%) and same set of cuts as before for the MLPBNN response cut.

| MLPBNN Response Cut | Total PQ Entries |
|---------------------|------------------|
| > 0                 | 3731641          |
| > 0.01              | 1138027          |
| > 0.02              | 906290           |
| > 0.03              | 769385           |
| > 0.04              | 670825           |
| > 0.05              | 593525           |
| > 0.07              | 475015           |
| > 0.09              | 384323           |
| > 0.1               | 346380           |

Table 5.7: Total events which pass the corresponding MLPBNN response cuts.

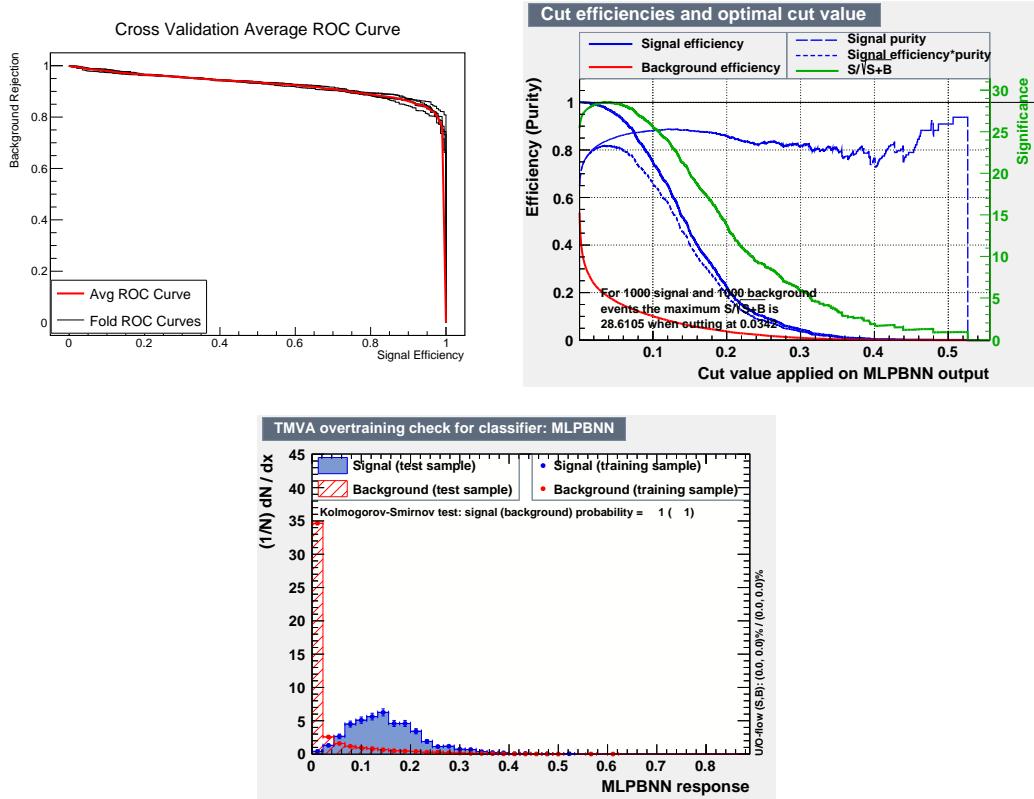


Figure 5.25: (TopLeft)ROC curve for MLPBNN method with real data as background. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by MLPBNN with real data as background. (Bottom) Overtraining check of the classifier with real data as background.

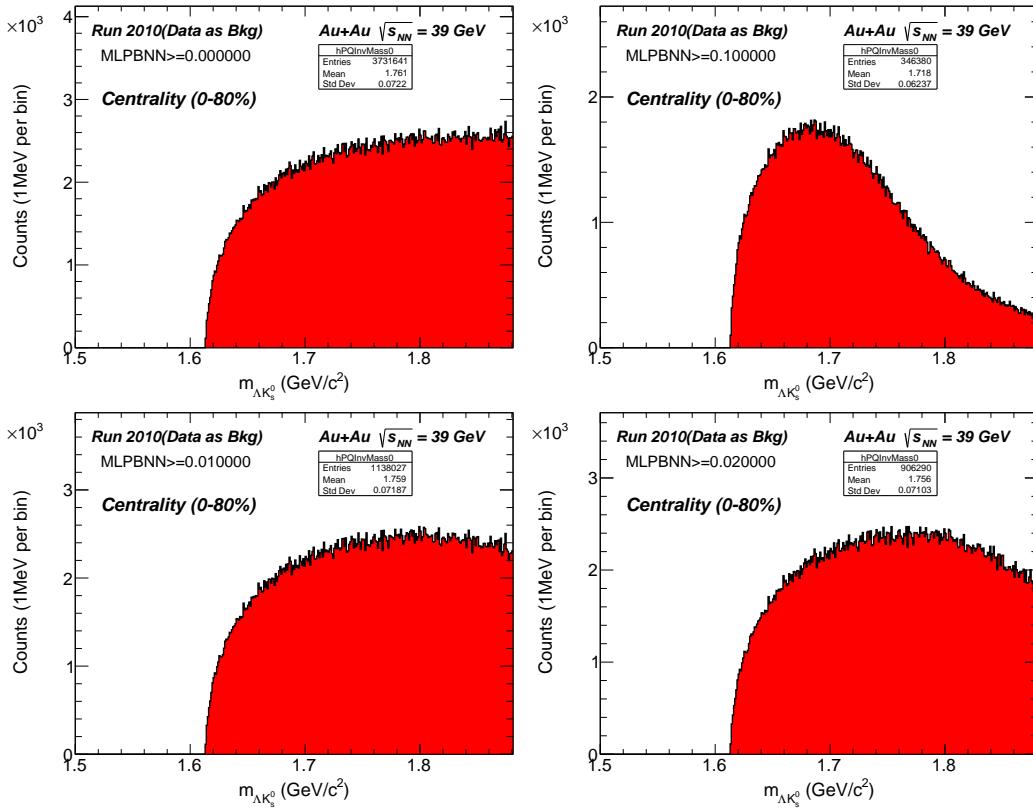


Figure 5.26:  $\Lambda^0 K_s^0$  invariant mass for centrality 0-80% after MLPBNN response cuts with real data as background (from top left to right)  $> (0, 0.1, 0.01$  and  $0.02)$  with 1MeV per bin

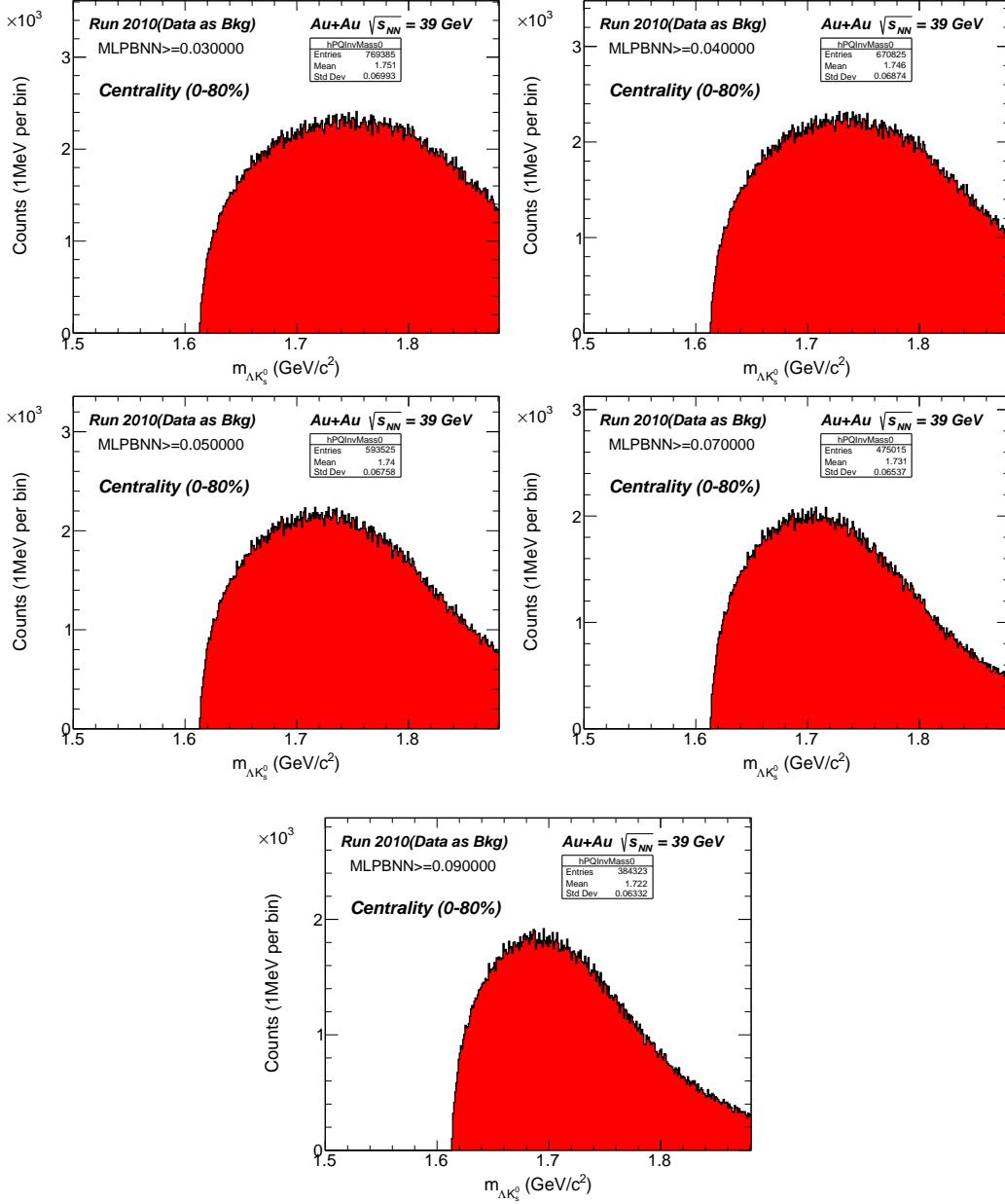


Figure 5.27:  $\Lambda^0 K_s^0$  invariant mass for centrality 0-80% after MLPBNN response cuts with real data as background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

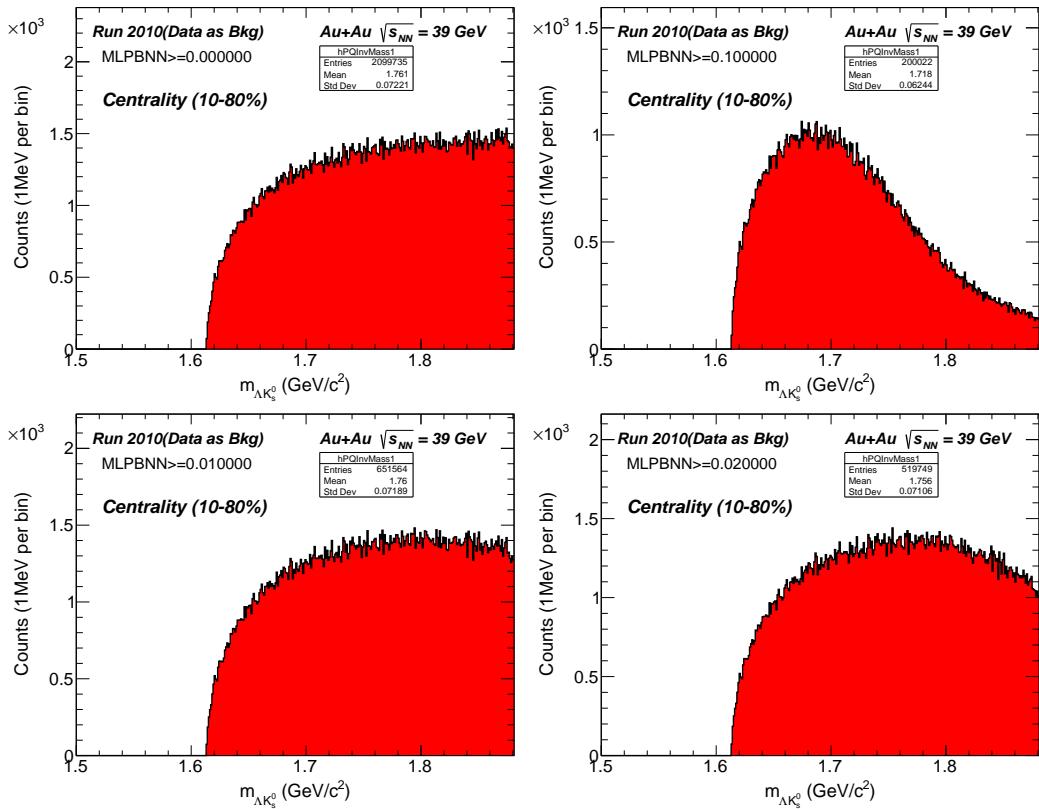


Figure 5.28:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-80% after MLPBNN response cuts with real data as background (from top left to right)  $> (0, 0.1, 0.01$  and  $0.02)$  with 1MeV per bin

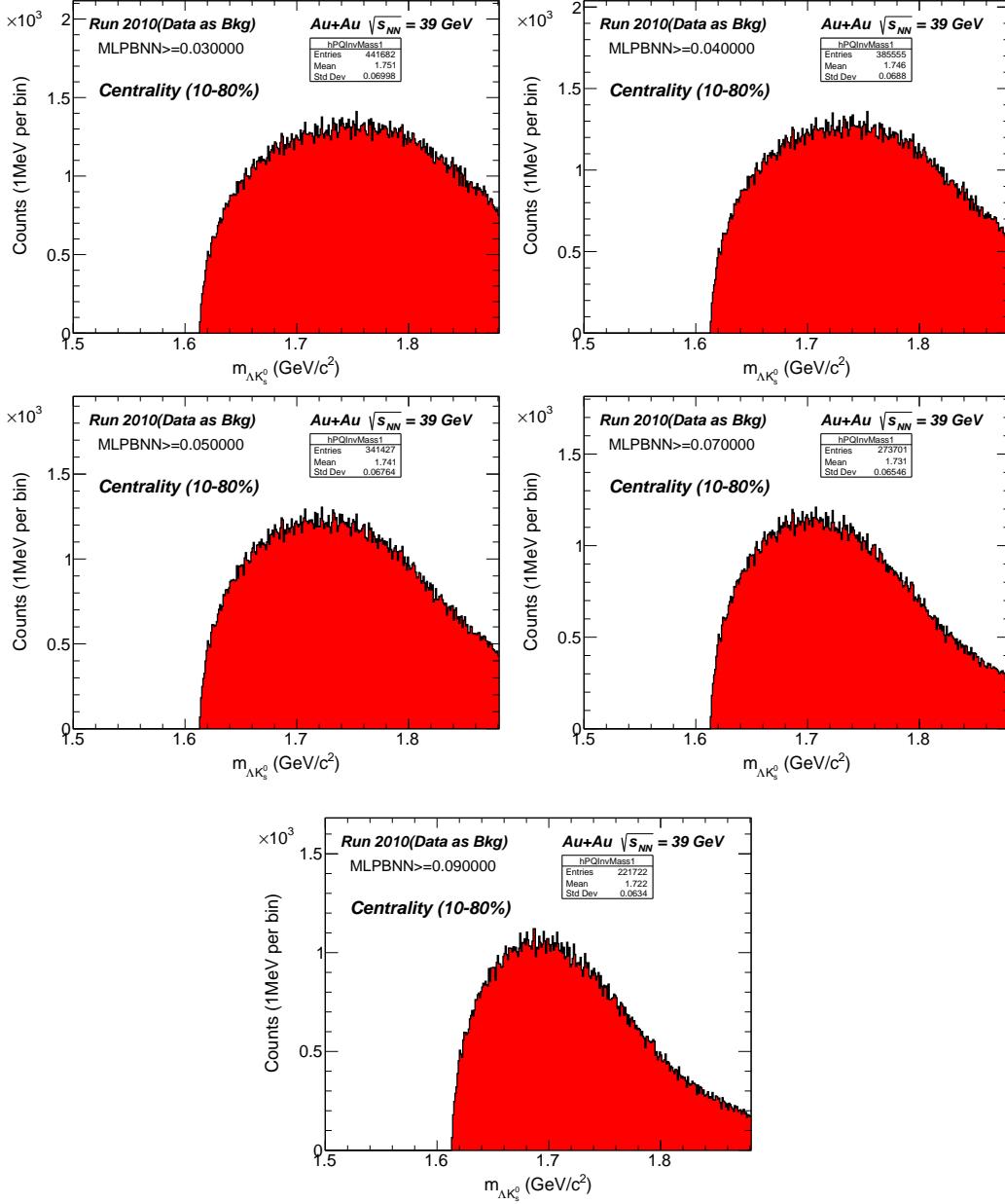


Figure 5.29:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-80% after MLPBNN response cuts with real data as background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

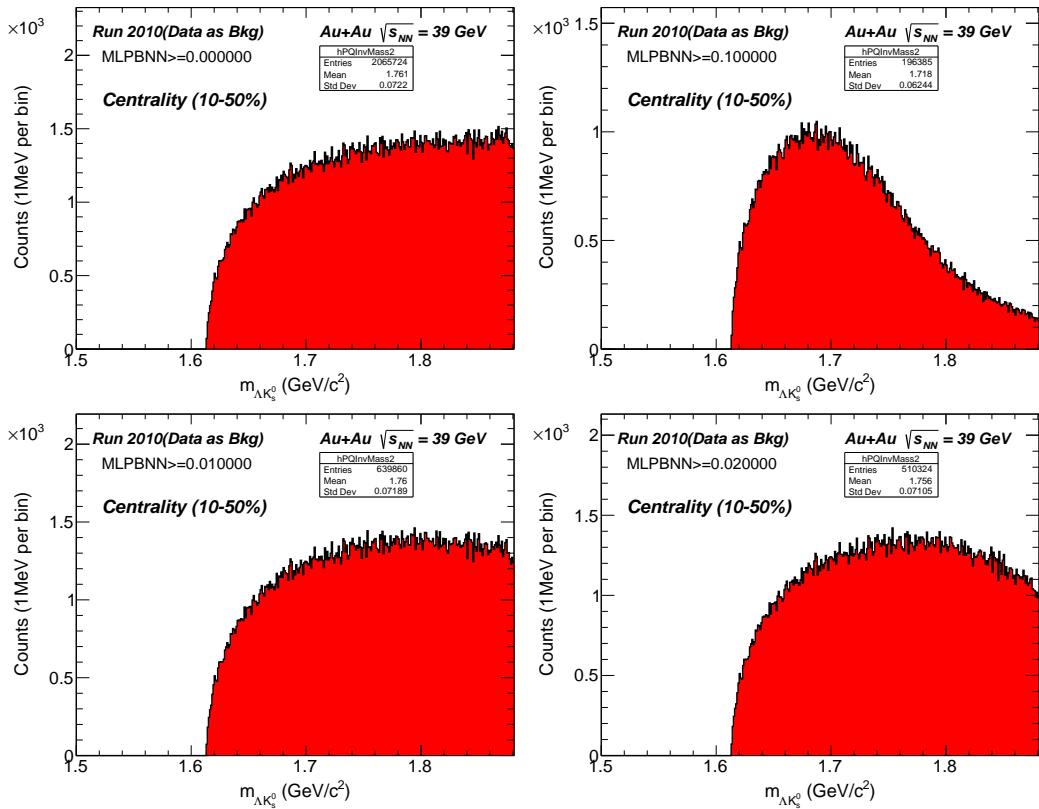


Figure 5.30:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-50% after MLPBNN response cuts with real data as background (from top left to right)  $> (0, 0.1, 0.01$  and  $0.02)$  with 1MeV per bin

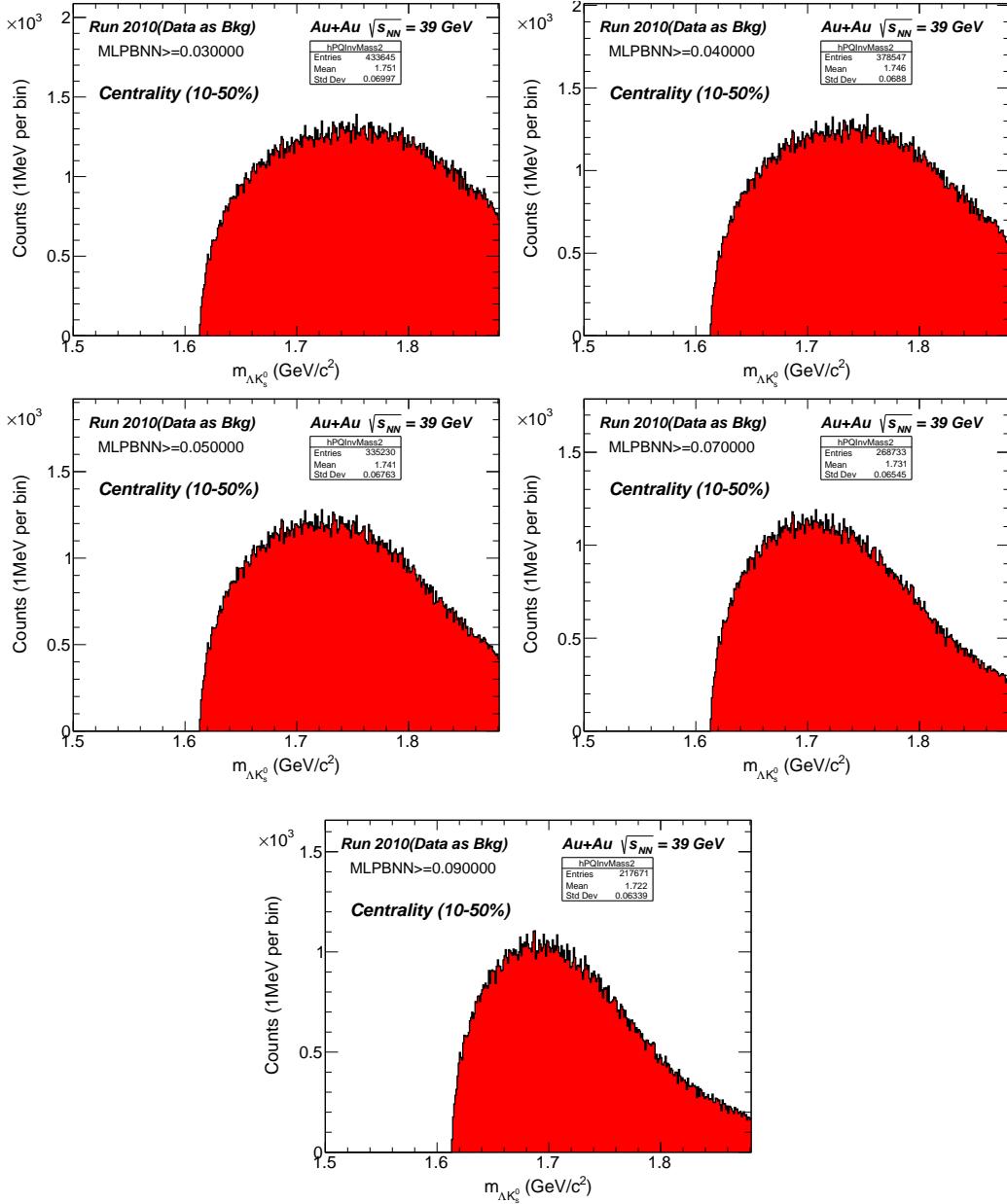


Figure 5.31:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-50% after MLPBNN response cuts with real data as background (from top left to right)  $> (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

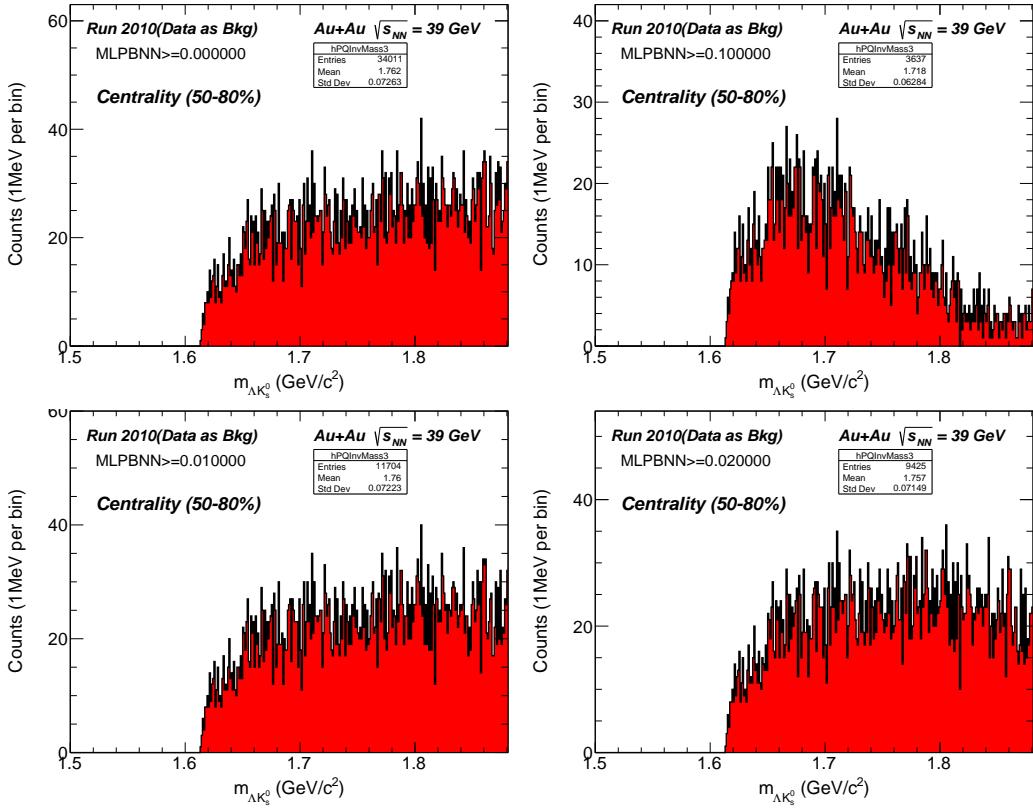


Figure 5.32:  $\Lambda^0 K_s^0$  invariant mass for centrality 50-80% after MLPBNN response cuts with real data as background (from top left to right)  $> (0, 0.1, 0.01$  and  $0.02)$  with 1MeV per bin

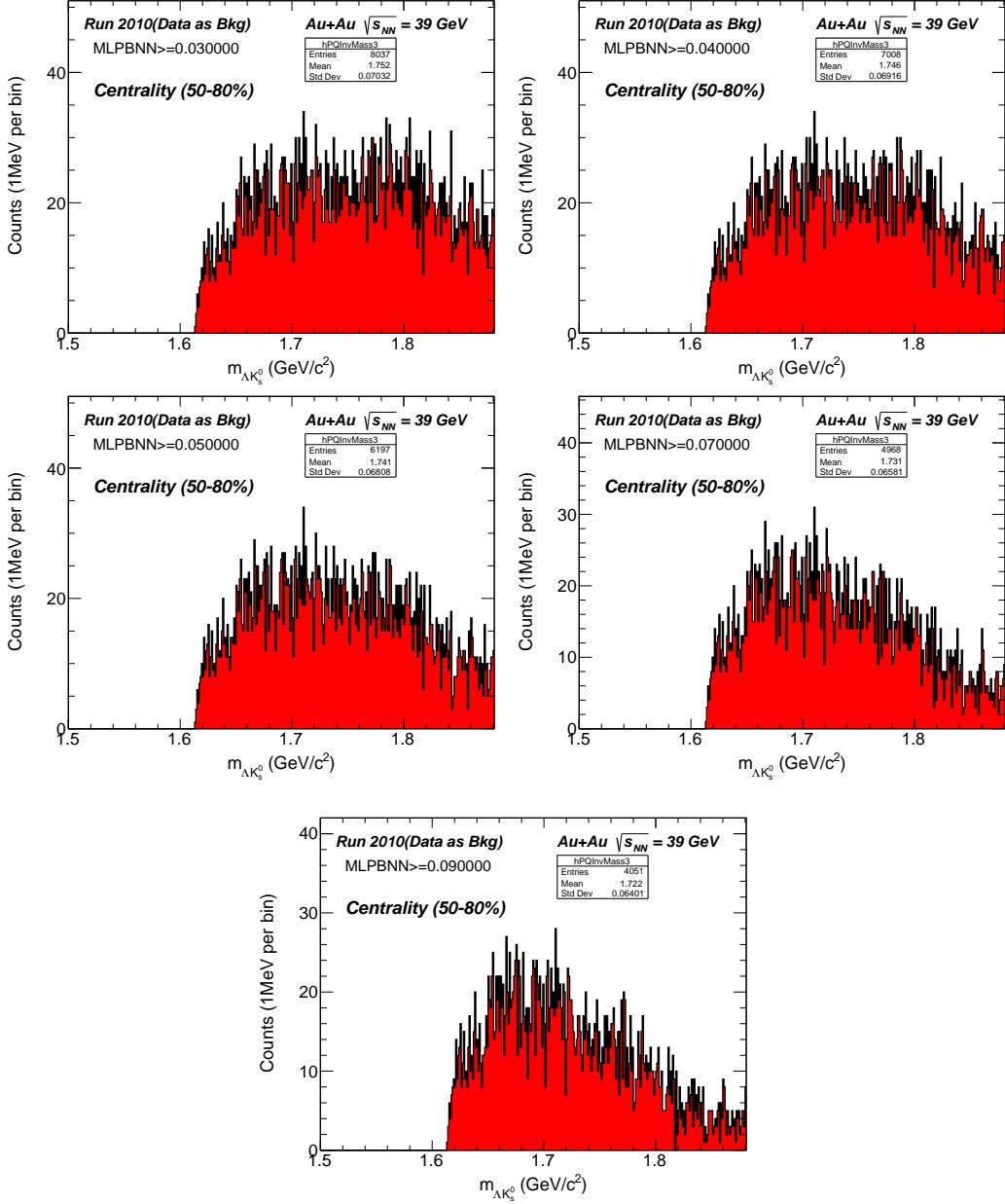


Figure 5.33:  $\Lambda^0 K_s^0$  invariant mass for centrality 50-80% after MLPBNN response cuts with real data as background (from top left to right)  $\geq (0.03, 0.04, 0.05, 0.07 \text{ and } 0.09)$  with 1MeV per bin

## **BDT results for $20\Lambda^0$ and $20K_s^0$ as Background**

The BDT method is trained using 24 variables and 5 fold crossvalidation. An additional cutset is applied while training and application part of TMVA which is given in Table 5.8. The mass cut of  $\pm 3\sigma$  corresponds to  $498.2 \pm 10.5$  MeV and  $1116 \pm 4.2$  MeV for  $K_s^0$  and  $\Lambda^0$  respectively. The pAngle is defined as the pointing angle which is the cosine of the angle between the particle momentum and position w.r.t the primary vertex. The cuts on  $\alpha$  which is the longitudinal momentum asymmetry in Armenteros plot and  $p_T^{Arm}$  were decided by comparing the plots from data and simulation. After applying the cutset the total number of signal events reduces to 153 from 1761 and background events to 391 from 43160. The total trees used were 850 with maximum depth of each tree as 3. The Gini Index separation criteria was used along with AdaBoost. We have used BDT response values of -0.4 and -0.35 after inferring from the BDT response cut graph Figure 5.34 to select the events for the invariant mass reconstruction of the PQ and we get the following graphs. The total number of events on which the BDT was applied was  $\approx 4.17$  Million. We have plotted the invariant mass histogram in different centrality bins of 0-80%, 10-80%, 10-50% and 50-80%.

| Cuts            | $\Lambda^0$            | $K_s^0$               | PQ                     |
|-----------------|------------------------|-----------------------|------------------------|
| $\chi_{fit}^2$  | <2                     | <2                    | <2                     |
| $\chi_{topo}^2$ | <2                     | <2                    | <2                     |
| $l/\Delta l$    | >12                    | >12                   | <2                     |
| pAngle          | >0.995                 | >0.995                | >0.99                  |
| Mass            | $\pm 3\sigma$          | $\pm 3\sigma$         | -                      |
| $p_T^{Arm}$     | $0 < p_T^{Arm} < 0.15$ | $0.05 < Q_t < 0.25$   | $0 < p_T^{Arm} < 0.35$ |
| $\alpha$        | $0.4 < \alpha < 0.9$   | $-0.8 < \alpha < 0.8$ | $-0.8 < \alpha < 0.2$  |

Table 5.8: Cut Set used for analysis using BDT

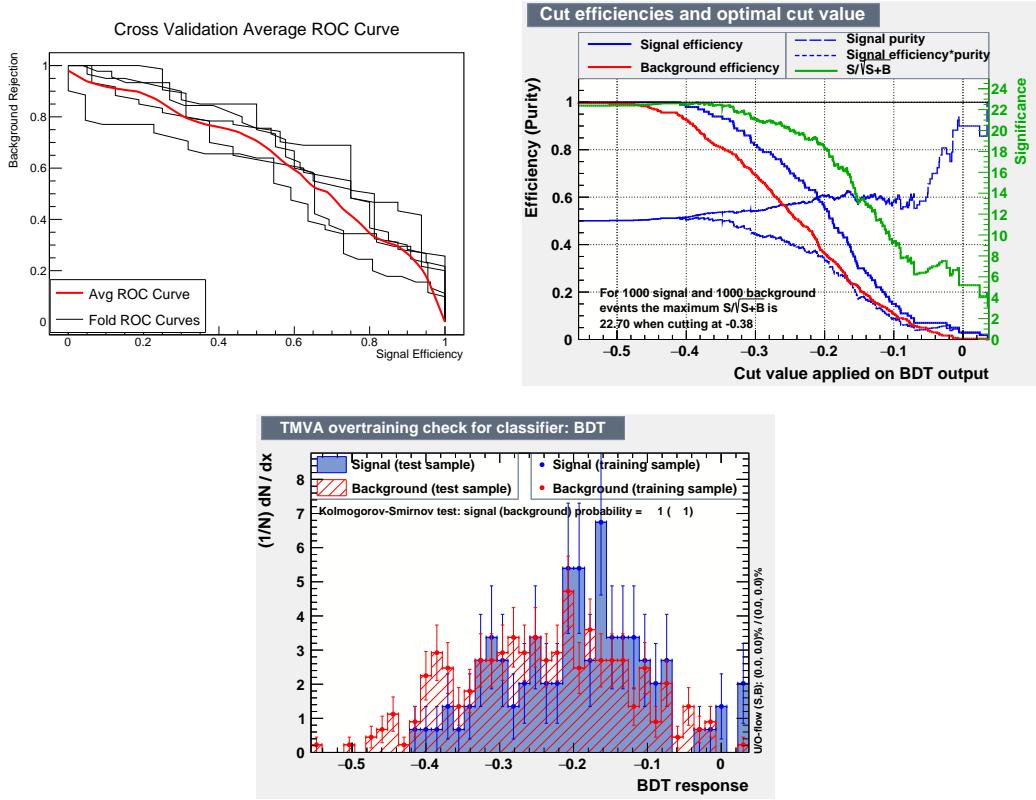


Figure 5.34: (TopLeft)ROC curve for BDT method  $20\Lambda^0$  and  $20K_s^0$  as background. The black curves represent the individual ROC curves for each cross-validation fold and red curve is the average of all the folds. (TopRight)The optimal cut value calculated by BDT. (Bottom) Overtraining check of the classifier.

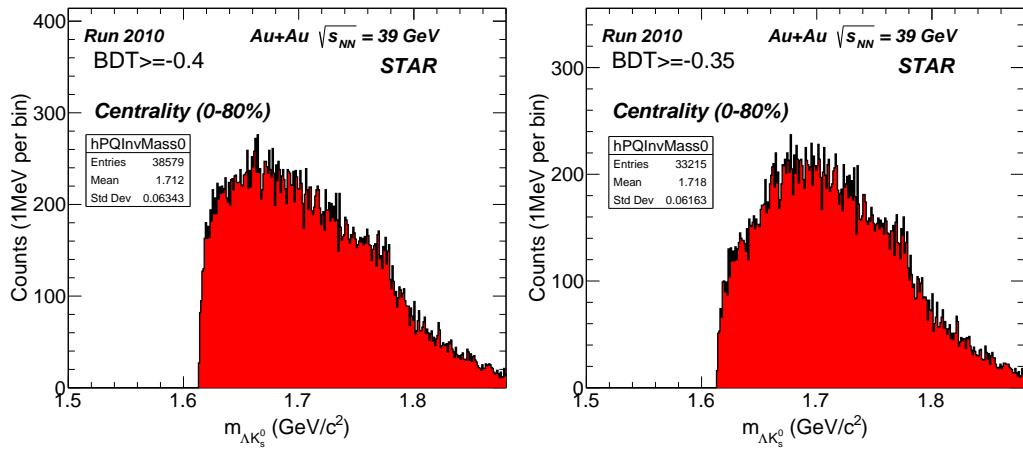


Figure 5.35:  $\Lambda^0 K_s^0$  invariant mass for centrality 0-80% after BDT response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from left to right)  $> (-0.4$  and  $-0.35)$  with 1MeV per bin

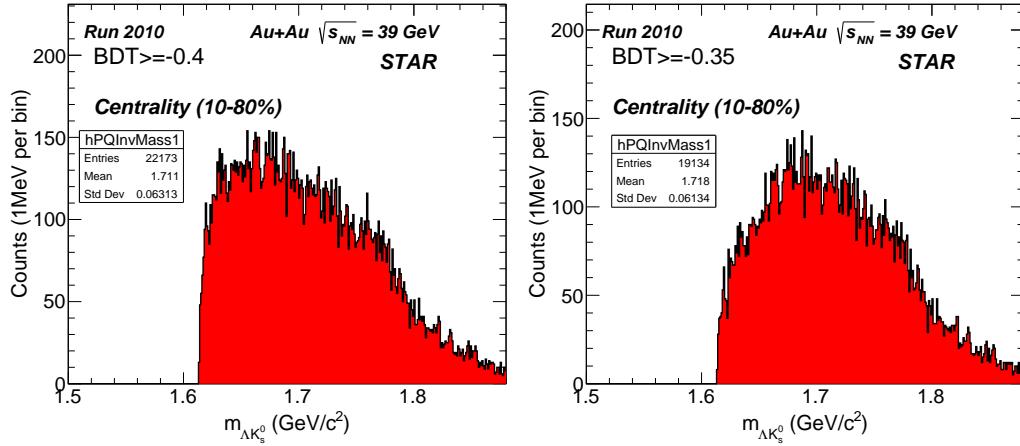


Figure 5.36:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-80% after BDT response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from left to right)  $> (-0.4$  and  $-0.35)$  with 1MeV per bin

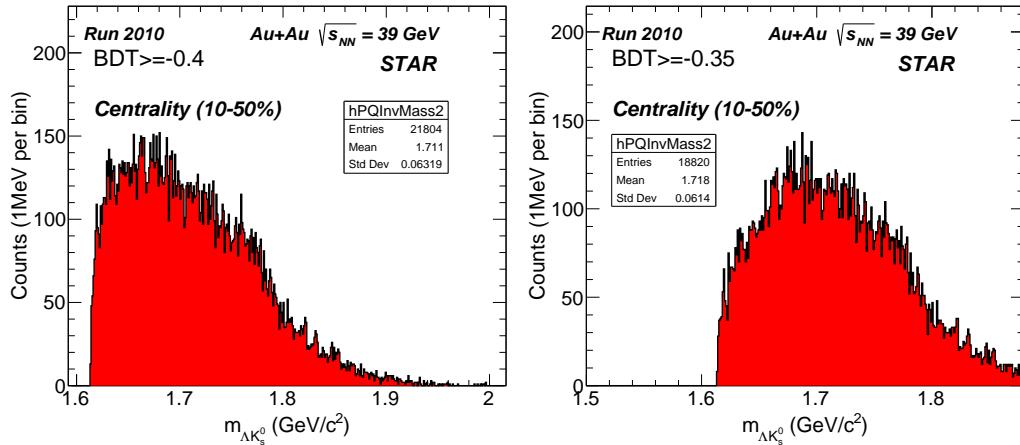


Figure 5.37:  $\Lambda^0 K_s^0$  invariant mass for centrality 10-50% after BDT response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from left to right)  $> (-0.4$  and  $-0.35)$  with 1MeV per bin

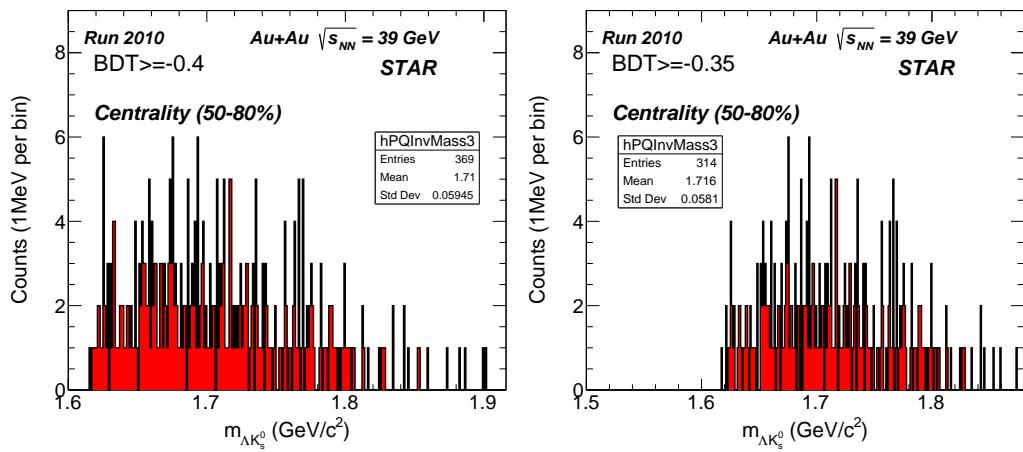


Figure 5.38:  $\Lambda^0 K_s^0$  invariant mass for centrality 50-80% after BDT response cuts with  $20\Lambda^0$  and  $20K_s^0$  as Background (from left to right)  $> (-0.4$  and  $-0.35)$  with 1MeV per bin

# Chapter 6

## Conclusion

In this thesis we present the preliminary results of the analysis of Au+Au 39 GeV data using the STAR detector in RHIC using the KFParticle particle reconstruction package and TMVA package of ROOT. We don't observe a signal for the PQ state in this analysis.

Further statistical analysis is required to give information about the upper limits of production of a hypothetical pentaquark that can be inferred by the non observation of the searched candidate. The number of events used for training BDT is very small as result of which the classifier becomes very biased as a result of which the training of the classifier is not optimal as can be seen from the ROC curve for BDT, so for better results we need to produce more simulation data for training. For analysis using TMVA we need to produce the signal and background trees having the same input variables in order to train our classifier.

For future extensions of this analysis one can train the classifiers with more

training and testing data which can produce better results. The analysis can be performed on  $\sqrt{s_{NN}} = 200$  GeV Au+Au STAR dataset with HFT detector which can help improve the statistics. Before training the classifier several additional cuts on the variables can be added which help in distinguishing signal from background. Different background reconstruction techniques like rotational background and mixed events technique can be applied after analysis through TMVA to further improve the signal peak over the background. Other classification methods can also be trained and applied which can perform better than the classifiers used in this analysis. In this analysis we have studied the  $\Lambda^0 K_s^0$  decay channel, other decay channels can also be studied like the  $\Lambda^0 K^+$  and the  $\Lambda^0 K^-$  channels.

# Bibliography

- [1] G. Aad, T. Abajyan, and e. al, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, vol. 716, no. 1, pp. 1–29, 2012.
- [2] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, and et al. Tumasyan, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, pp. 30–61, sep 2012.
- [3] R. N. Cahn, “The eighteen arbitrary parameters of the standard model in your everyday life,” *Rev. Mod. Phys.*, vol. 68, pp. 951–959, jul 1996.
- [4] “The Standard Model | CERN <https://home.cern/science/physics/standard-model>.”
- [5] E. R. Siegel, “Beyond The Galaxy: How Humanity Looked Beyond Our Milky Way And Discovered The Entire Universe,” 2015.
- [6] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?,” *Phys. Rev. Lett.*, vol. 30, pp. 1346–1349, jun 1973.

- [7] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories,” *Phys. Rev. Lett.*, vol. 30, pp. 1343–1346, jun 1973.
- [8] D. Griffiths, *Introduction to elementary particles*. 2008.
- [9] J. C. Collins and M. J. Perry, “Superdense Matter: Neutrons or Asymptotically Free Quarks?,” *Phys. Rev. Lett.*, vol. 34, pp. 1353–1356, may 1975.
- [10] N. Cabibbo and G. Parisi, “Exponential hadronic spectrum and quark liberation,” *Physics Letters B*, vol. 59, no. 1, pp. 67–69, 1975.
- [11] P. A. Zyla and Others, “Review of Particle Physics,” *PTEP*, vol. 2020, no. 8, p. 083C01, 2020.
- [12] F. Karsch, “Lattice QCD at High Temperature and Density,” in *Lectures on Quark Matter* (L. Plessas Willibald and Mathelitsch, ed.), pp. 209–249, Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.
- [13] T. K. Nayak, “Probing the QCD phase structure using event-by-event fluctuations,” *Journal of Physics: Conference Series*, vol. 1602, p. 012003, jul 2020.
- [14] E. V. Shuryak, “Quantum chromodynamics and the theory of superdense matter,” *Physics Reports*, vol. 61, no. 2, pp. 71–158, 1980.
- [15] M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, “Color superconductivity in dense quark matter,” *Rev. Mod. Phys.*, vol. 80, pp. 1455–1515, nov 2008.

- [16] F. Karsch, “Lattice results on QCD thermodynamics,” *Nuclear Physics A*, vol. 698, no. 1, pp. 199–208, 2002.
- [17] F. R. Brown, F. P. Butler, H. Chen, N. H. Christ, Z. Dong, W. Schaffer, L. I. Unger, and A. Vaccarino, “On the existence of a phase transition for QCD with three light quarks,” *Phys. Rev. Lett.*, vol. 65, pp. 2491–2494, nov 1990.
- [18] Y. Hatta and T. Ikeda, “Universality, the QCD critical and tricritical point, and the quark number susceptibility,” *Phys. Rev. D*, vol. 67, p. 14028, jan 2003.
- [19] D. H. Rischke and M. Gyulassy, “The maximum lifetime of the quark-gluon plasma,” *Nuclear Physics A*, vol. 597, no. 4, pp. 701–726, 1996.
- [20] J. Adams, M. M. Aggarwal, and e. al, “Experimental and theoretical challenges in the search for the quark-gluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions,” *Nuclear Physics A*, vol. 757, no. 1, pp. 102–183, 2005.
- [21] J. Adams, C. Adler, and e. al, “Evidence from  $d+\{A\}\{u\}$  Measurements for Final-State Suppression of High- $\{p\}_T$  Hadrons in  $\{A\}\{u\}+\{A\}\{u\}$  Collisions at RHIC,” *Phys. Rev. Lett.*, vol. 91, p. 72304, aug 2003.
- [22] J. Rafelski and B. Müller, “Strangeness Production in the Quark-Gluon Plasma,” *Phys. Rev. Lett.*, vol. 48, pp. 1066–1069, apr 1982.
- [23] R. B. Leighton and S. D. Wanlass, “The Occurrence of Heavy Mesons in Penetrating Showers,” *Physical Review*, vol. 86, p. 426, may 1952.

- [24] “PDG Pentaquark Review`https://pdg.lbl.gov/2020/reviews/rpp2020-rev-pentaquarks.pdf`.”
- [25] M. Gell-Mann, “A schematic model of baryons and mesons,” *Physics Letters*, vol. 8, no. 3, pp. 214–215, 1964.
- [26] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 2,” in *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978* (D. B. Lichtenberg and S. P. Rosen, eds.), pp. 22–101, apr 1964.
- [27] H. J. Lipkin, “New possibilities for exotic hadrons — anticharmed strange baryons,” *Physics Letters B*, vol. 195, no. 3, pp. 484–488, 1987.
- [28] T. G. Trippe, A. Barbaro-Galtieri, and e. al, “Review of particle properties,” *Reviews of Modern Physics*, vol. 48, pp. S1–S245, apr 1976.
- [29] M. Praszałowicz, “Pentaquark in the Skyrme model,” *Physics Letters B*, vol. 575, no. 3, pp. 234–241, 2003.
- [30] D. Diakonov, V. Petrov, and M. Polyakov, “Exotic anti-decuplet of baryons: prediction from chiral solitons,” *Zeitschrift für Physik A Hadrons and Nuclei*, vol. 359, no. 3, pp. 305–314, 1997.
- [31] H. Weigel, “Radial excitations of low-lying baryons and the Z+ penta-quark,” *The European Physical Journal A - Hadrons and Nuclei*, vol. 2, no. 4, pp. 391–402, 1998.

- [32] T. Nakano, D. S. Ahn, and e. al, “Evidence for a narrow  $s = +1$  baryon resonance in photoproduction from the neutron,” *Phys. Rev. Lett.*, vol. 91, p. 012002, Jul 2003.
- [33] V. V. Barmin, V. S. Borisov, and e. al, “Observation of a baryon resonance with positive strangeness in  $K^+$  collisions with Xe nuclei,” *Physics of Atomic Nuclei 2003 66:9*, vol. 66, pp. 1715–1718, apr 2003.
- [34] W. M. Yao, C. Amsler, and e. al, “Review of Particle Physics,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 33, p. 1, jul 2006.
- [35] R. Aaij, B. Adeva, and e. al, “Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda b0 \rightarrow j/\psi K-p$  Decays,” *Physical Review Letters*, vol. 115, p. 072001, aug 2015.
- [36] R. Aaij, C. Abellán Beteta, and e. al, “Observation of a Narrow Pentaquark State,  $P_c(4312)^{+}$ , and of the Two-Peak Structure of the  $P_c(4450)^{+}$ ,” *Physical review letters*, vol. 122, p. 222001, jun 2019.
- [37] H. A. M, “Searches for pentaquark states with the star experiment at rhic,” Master’s thesis, IISER Berhampur, 2021.
- [38] S. Kabana and C. STAR, “Exotic Particle Searches with STAR at RHIC,” *Acta Physica Hungarica A) Heavy Ion Physics*, vol. 24, pp. 321–328, oct 2005.
- [39] M. Harrison, T. Ludlam, and S. Ozaki, “RHIC project overview,” *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators*

*erators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 235–244, mar 2003.

- [40] M. Adamczyk and L. et al. Antvorskov, “The BRAHMS experiment at RHIC,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 437–468, mar 2003.
- [41] K. Adcox and S. S. et al. Adler, “PHENIX detector overview,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 469–479, mar 2003.
- [42] B. Back and M. et al. Baker, “The PHOBOS detector at RHIC,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 603–623, mar 2003.
- [43] K. H. et al. Ackermann, “STAR detector overview,” *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 624–632, mar 2003.
- [44] C. E. Pérez Lara, “The sPHENIX Experiment,” *EPJ Web of Conferences*, vol. 171, feb 2018.
- [45] “RHIC Complex | RHIC’s 2.4 mile ring has six intersection poi...[https://www.bnl.gov/today/body\\_pics/2021/12/rhic\\_complex\\_d3451011-hr.jpg](https://www.bnl.gov/today/body_pics/2021/12/rhic_complex_d3451011-hr.jpg).”

- [46] S. Ozaki and T. Roser, “Relativistic Heavy Ion Collider, its construction and upgrade,” *Progress of Theoretical and Experimental Physics*, vol. 2015, pp. 3–102, mar 2015.
- [47] M. Beddo and E. et al. Bielick, “The STAR Barrel Electromagnetic Calorimeter,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 725–739, mar 2003.
- [48] “STAR Future & Upgrade | The STAR experiment <https://drupal.star.bnl.gov/STAR/future>.”
- [49] M. et al. Anderson, “The STAR time projection chamber: a unique tool for studying high multiplicity events at RHIC,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 659–678, mar 2003.
- [50] L. Kotchenda, S. Kozlov, P. Kravtsov, A. Markov, M. Strikhanov, B. Stringfellow, V. Trofimov, R. Wells, and H. Wieman, “STAR TPC gas system,” *Nuclear Instruments and Methods in Physics Research. Section A, Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, pp. 703–712, mar 2003.
- [51] B. Bonner, H. Chen, G. Eppley, F. Geurts, J. Lamas-Valverde, C. Li, W. J. Llope, T. Nussbaum, E. Platner, and J. Roberts, “A single Time-of-Flight tray based on multigap resistive plate chambers for the STAR experiment at RHIC,” *Nuclear Instruments and Methods in Physics Re-*

*search Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 508, pp. 181–184, aug 2003.

- [52] S. Gorbunov, *On-line reconstruction algorithms for the CBM and AL-ICE experiments*. PhD thesis, 2013.
- [53] R. R. Fruhwirth and M. Regler, “Data analysis techniques for high-energy physics.,” p. 384, 2000.
- [54] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Journal of Basic Engineering*, vol. 82, pp. 35–45, mar 1960.
- [55] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, pp. 444–450, dec 1987.
- [56] M. Zyzak, *Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR*. PhD thesis, 2016.
- [57] K. Albertsson and S. et al. Gleyzer, “arXiv:physics/0703039 [Data Analysis, Statistics and Probability] TMVA 4 Toolkit for Multivariate Data Analysis with ROOT Users Guide,” tech. rep., 2005.
- [58] C. G. Broyden, “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations,” *IMA Journal of Applied Mathematics*, vol. 6, pp. 76–90, mar 1970.
- [59] “Production MuDst Summary on DD<https://www.star.bnl.gov/public/comp/prod/localdata/ProdDDstreams.html>.”

- [60] B. I. Abelev and M. M. et al. Aggarwal, “Systematic Measurements of Identified Particle Spectra in pp, d+Au and Au+Au Collisions from STAR,” 2009.
- [61] J. Podolanski and R. Armenteros, “III. Analysis of V-events,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 45, pp. 13–30, jan 1954.
- [62] J. Allison and K. et al. Amako, “Recent developments in Geant4,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186–225, nov 2016.
- [63] “1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm>.”