Identification of central events in nucleus-nucleus collisions by machine learning algorithms

Evgeny Andronov, Andrey Seryakov, Vladimir Kovalenko

Saint Petersburg State University

Introduction

Precise centrality determination is a key in a lot of studies dedicated to search for the critical point of strongly interacting matter.

Typical measures of centrality:

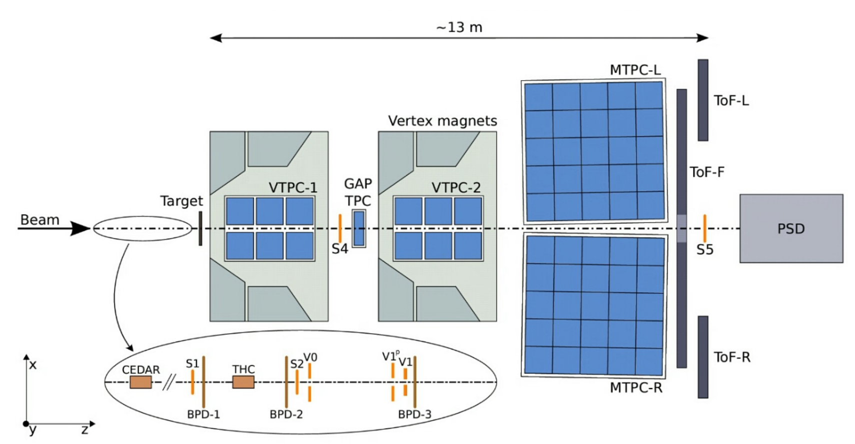
- impact parameter (unaccessible experimentally)
- number of nucleon participants

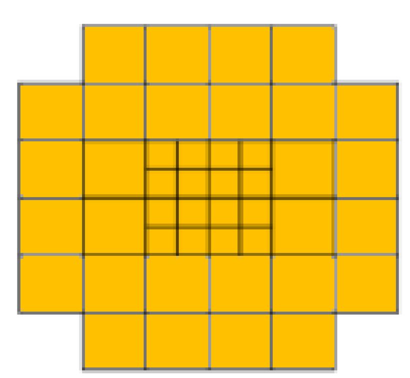
and their experimental proxies:

- forward energy
- multiplicity in a limited kinematic acceptance

In this work we investigate possibility to increase resolution capacity of forward energy detector of NA61/SHINE experiment [1] by means of machine learning techniques. Previously, this study has been performed for simulated Li+Be reaction [2] and for calorimeters with a beam hole [3]. In this contribution Ar+Sc collisions at 150A GeV/c simulated in EPOS1.99 model [4] are considered.

Projectile spectator detector of NA61/SHINE [1]

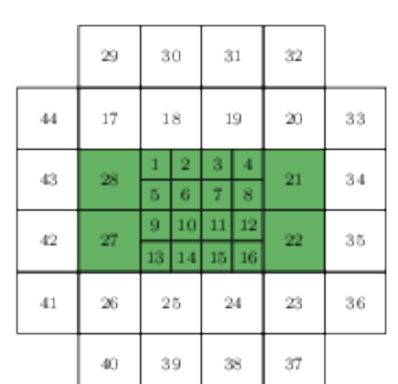


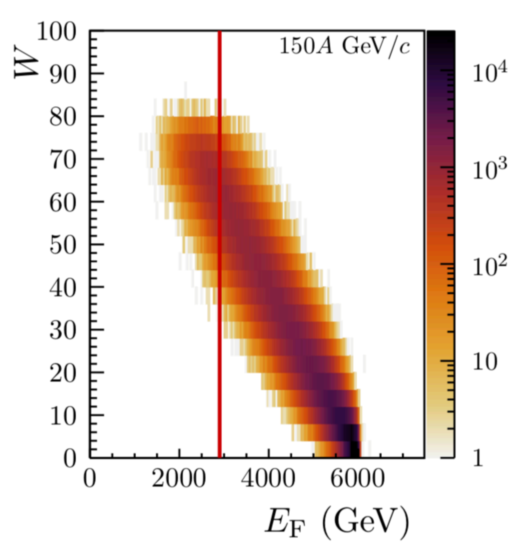


- 16 central 10*10 cm2; 28 peripheral 20*20 cm2
- 10 sections in deep
- potentially 440 features
- similar structure of calorimeters in BM@N, CBM, MPD (no beam hole)

Cut-based centrality selection:

Energy deposition in 20 modules [5]





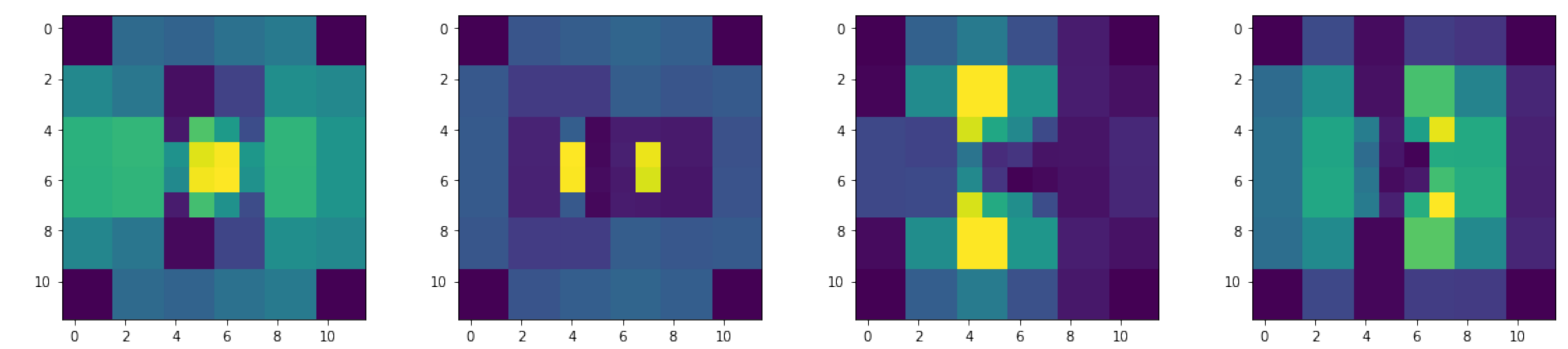
For 6% most central:

$$TPR = \frac{TP}{P} = 0.8918 \quad FPR = \frac{FP}{N} = 0.007$$

$$Prec = \frac{TP}{TP + FP} = 0.8921$$

Principle components analysis

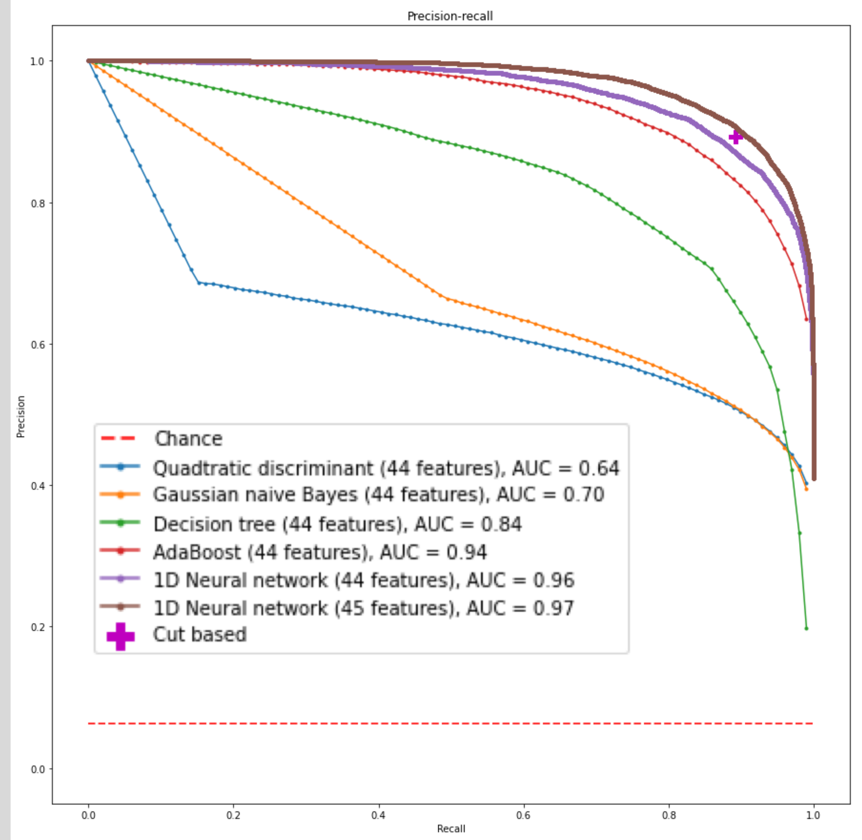
To understand significance of different modules we performed principal component analysis for a set of all 44 modules:

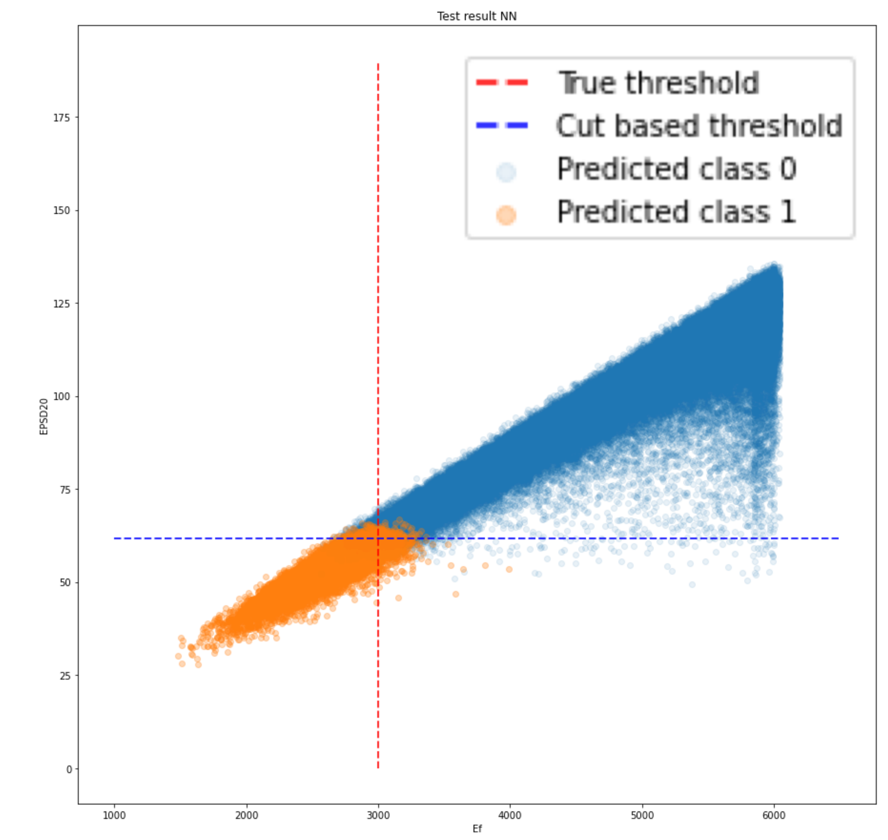


Four most significant principal components illustrate the major role of small central modules with a ring of closest large modules .

Classifiers: supervised learning

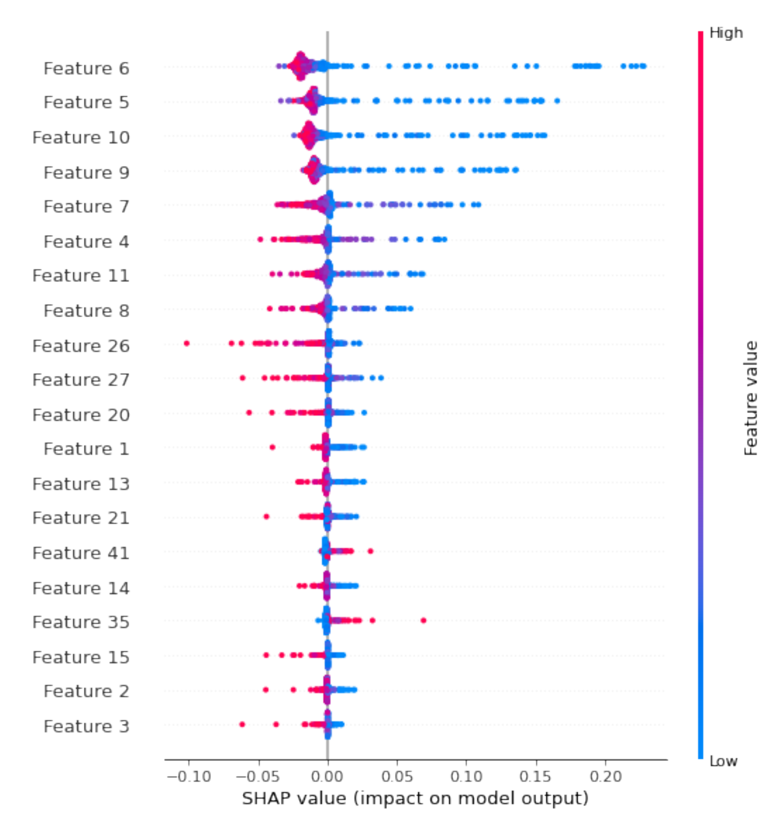
A number of classifying machine learning algorithms has been trained to distinguish between central and peripheral Ar+Sc collisions





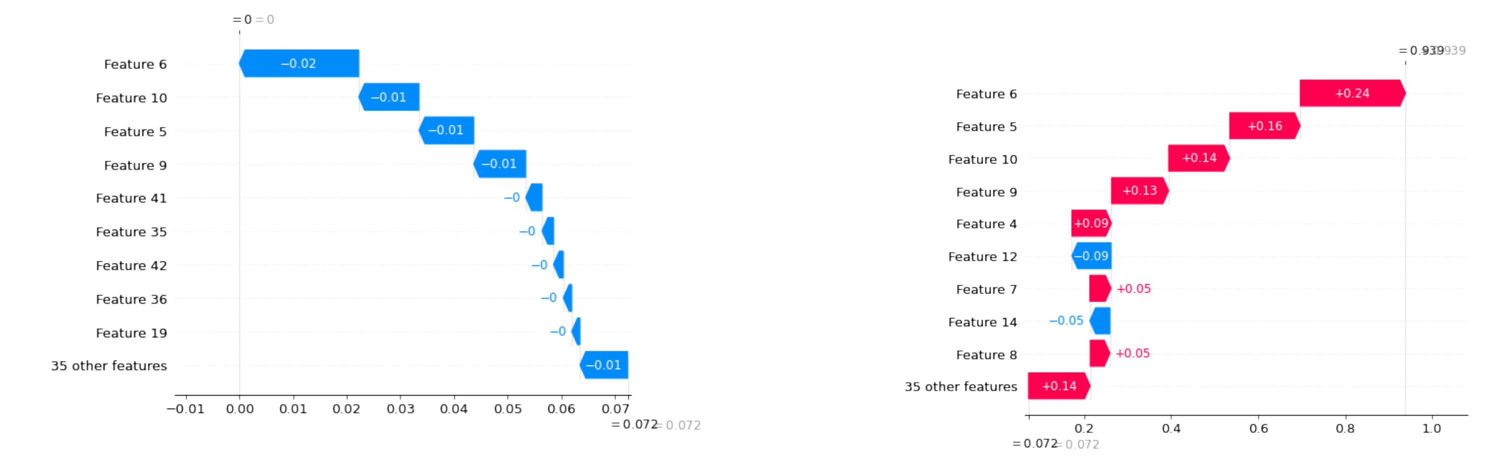
Classifiers based on 44 features do not increase precision for a given recall (TPR) in comparison to cut based Introduction of charged hadron multiplicity as the 45th feature allows to increase precision by ~1%.

Shapley values [6] of features estimated for Keras sequential convoluted neural network once again indicate significant role of central modules.



Feature importance

Shapley values for single events classified as peripheral (left) and central (right). For both events the dominating feature is module 6.



Symbolic classifier

Gplearn package [7] commonly used in genetic programming allows to perform symbolic classification. The idea of algorithm is fitting data with random functions of input features constructed using selected set of mathematical operators.

Symbolic classifier has been applied to this dataset with principal components as input features.

Optimal classifying function was found to be

$13 \cdot PC_0 - PC_1 - 24.027$ with TPR=0.93, FPR=0.01, Prec=0.83

Summary and Plans

Multiple machine learning techniques have been applied to improve centrality selection in Ar+Sc collisions using energy deposition in forward hadronic calorimeter as a feature.

All tested methods do not increase significantly precision of event selection in comparison to traditional cut based approach and indicate the importance of central modules in this procedure.

References

1. NA61/SHINE, JINST 9, P06005 (2014)
2. A. Seryakov, D. Uzhva, Phys.Part.Nucl. 51(3), 331 (2020)
3. N. Karpushkin et al., J.Phys.Conf.Ser. 1667(1), 012121 (2020)
4. K. Werner, F.-M. Liu, T. Pierog, Phys.Re.v. C 74, 044902 (2006)
5. NA61/SHINE, Eur. Phys. J. C81(5), 397 (2021)
6. S. Lundberg, S.-I. Lee, Adv. Neur. In. Proc. Syst. 30, 4765 (2017)
7. https://gplearn.readthedocs.io/en/stable/index.html

This work is supported by the Russian Science Foundation under grant 17-72-20045.

e.v.andronov@spbu.ru