Machine learning classification for D^0 meson signal extraction in d+Au collisions

Mini-workshop on Machine Learning for Particle Physics, ICNFP 2021
August 25, 2021

Jaroslav Bielčík, **Kateřina Hladká**, Lukáš Kramárik, Václav Kůs
Czech Technical University in Prague
Katerina.Hladka@fjfi.cvut.cz

FACULTY OF
NUCLEAR SCIENCES
AND PHYSICAL
ENGINEERING
CTU IN PRAGUE

The work was supported from the project LTT18002 of the Ministry of Education, Youth, and Sport of the Czech Republic

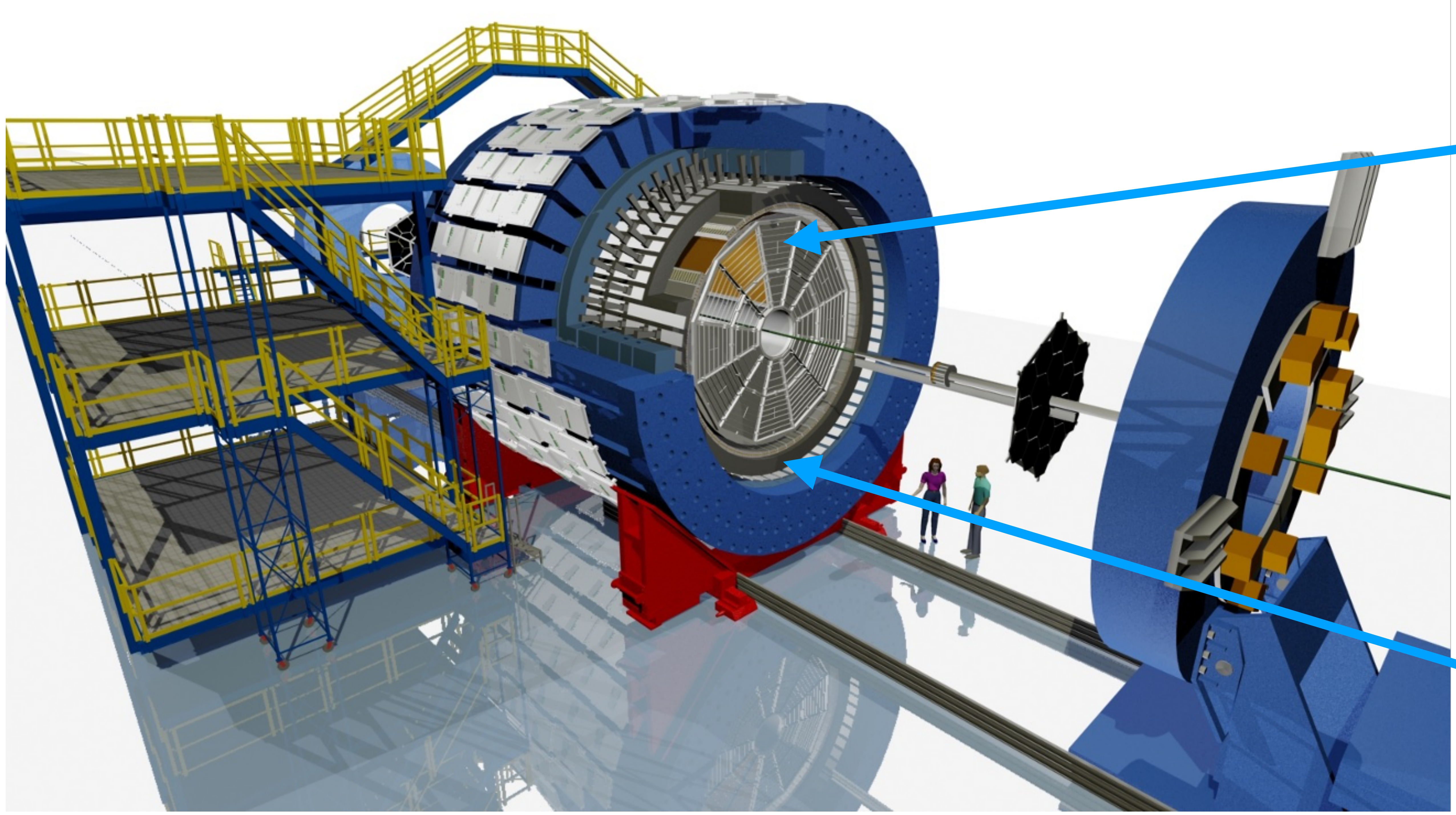
Outline

- The Solenoid Tracker At RHIC (STAR)
- Heavy-flavour quarks as a probe of QGP
- Dataset and topological properties of D^0 mesons
- Random forests
- Boosted decision trees
- Deep neural network
- Comparison of trained algorithms

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

The Solenoid Tracker At RHIC (STAR)

- Designed to study the strongly interacting matter
- Excels in tracking and identification of charged particles at mid-rapidity with full azimuthal coverage
- Most of the subsystems are immersed in 0.5 T solenoidal magnetic field



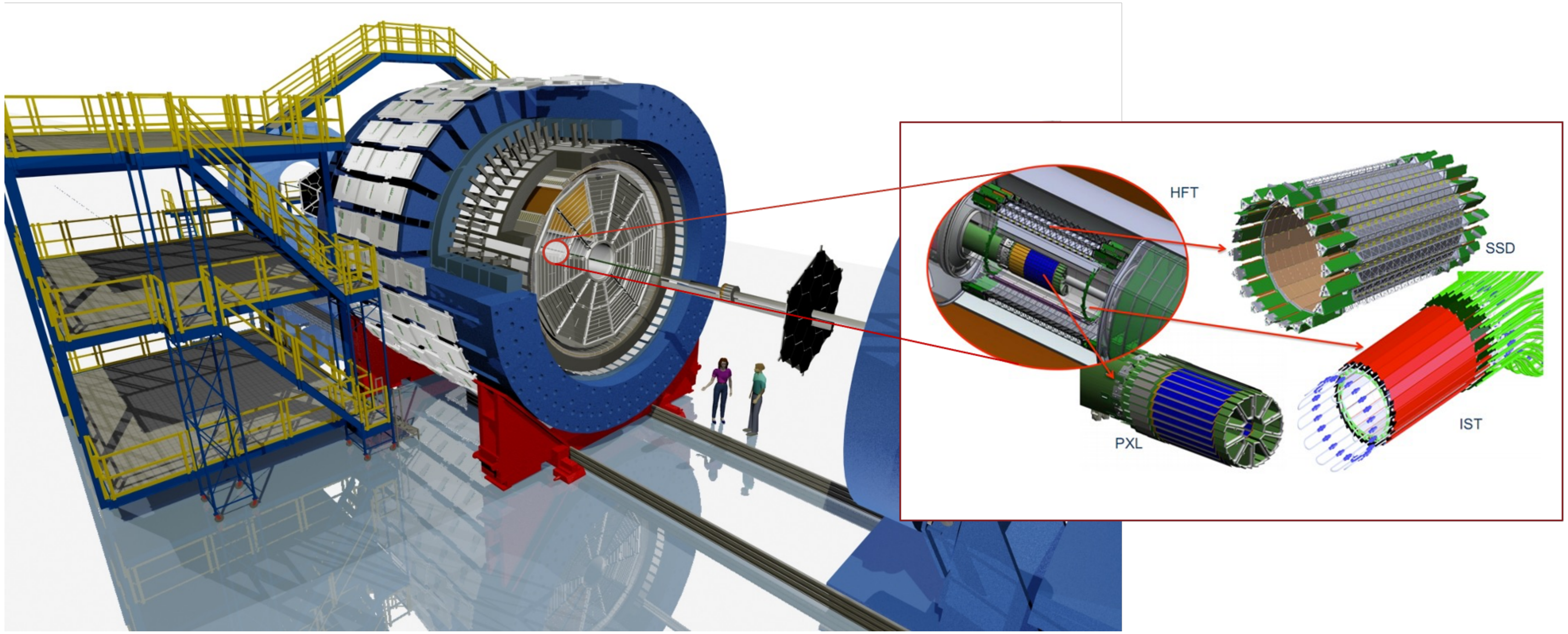
Time Projection Chamber (TPC)
- Main tracking device; momentum determination
- Particle identification via specific energy loss dE/dx

Time of Flight (TOF)
- Particle identification at low transverse momentum p_T via velocity β

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

The Solenoid Tracker At RHIC (STAR)

- Designed to study the strongly interacting matter
- Excels in tracking and identification of charged particles at mid-rapidity with full azimuthal coverage
- Most of the subsystems are immersed in 0.5 T solenoidal magnetic field



Heavy Flavor Tracker (HFT)

- Inner tracking system

- Excellent DCA_{xy} and DCA_z resolution: ≈50 μm for kaons at p_T = 750 MeV/c

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Heavy-flavour quarks as a probe of QGP

- **Heavy-flavour quarks possess large masses**
  - they are produced primarily at the initial stages of heavy-ion collisions
  - they experience the evolution of the collision

- **Charm quarks are confined in hadrons, that due their lifetimes could not be observed directly in the detector**

- **Open charm hadrons could be studied via hadronic decays:** $D^0(\overline{D^0}) \rightarrow K^-\pi^+(K^+\pi^-)$

- **Our goal is to reconstruct D^0 meson in order to study cold nuclear matter effects in asymmetric d+Au collisions at** $\sqrt{s_{NN}} = 200$ GeV

Motivation

Goal is to **compare** performance of the various machine learning methods (and packages) for finding D^0 signal

- **BDT** - **Boosted Decision Trees (TMVA, ROOT)**[1]
  Widely used, since implemented within ROOT package

- **RF - Random forest (scikit-learn, Python)**[2]
  Robust and interpretable, one of the most popular supervised machine learning methods, fast to train and simple to optimize

- **DNN - Deep neural network (Keras, Python)**[3]
  The most complex machine learning method tested in this study.

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

[1] *H. Voss, A. Hocker, J. Stelzer and F. Tegenfeldt, TMVA, the Toolkit for Multivariate Data Analysis with ROOT, PoS ACAT (2007) 040.*
[2] *F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.*
[3] *F. Chollet et al., "Keras." https://keras.io, 2015*

Dataset

- **HIJING full-event simulation of 2016 d+Au collisions at $\sqrt{s_{NN}}$ = 200 GeV**
  - official STAR's embedding to zerobias collisions
  - at least one D^0 meson decayed to K and π in the event
  - events with primary vertex $|V_z|< 6$ cm

- **$D^0 \rightarrow K\,\pi$ decay channel is studied**
  - pairs of kaons and pions are reconstructed in the same way as in the data
  - correct charge pairs are used for studies

- **Track selection**
  - HFT tracks
  - At least 15 space points in the TPC for track reconstruction
  - $p_T > 0.15$ GeV/c
  - Track pseudorapidity $|\eta|< 1$
  - TPC particle identification via dE/dx: pions: $|n\sigma_{pion}| < 3$, kaons: $|n\sigma_{kaon}| < 2$

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

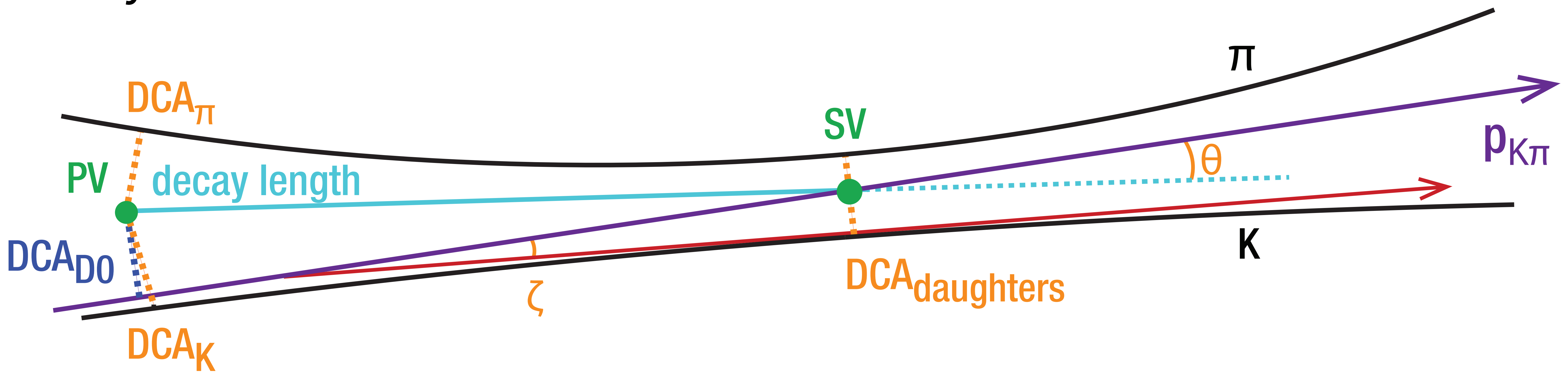
Topological reconstruction of D^0 mesons

- Topological properties of D^0 decays used for their reconstruction:

DCA = distance of closest approach

PV = primary vertex, place of the d+Au collision

SV = secondary vertex, place of the D^0 meson decay



- **decay length:**
  - distance between primary vertex (PV) and secondary vertex (SV) of D^0 meson candidate
  - ideally $\approx$ 200 µm

- **$DCA_{daughters}$** between kaon and pion tracks
  - ideally $\rightarrow$ 0 µm

- **cosine of angle θ** between D^0 momentum and decay length vector

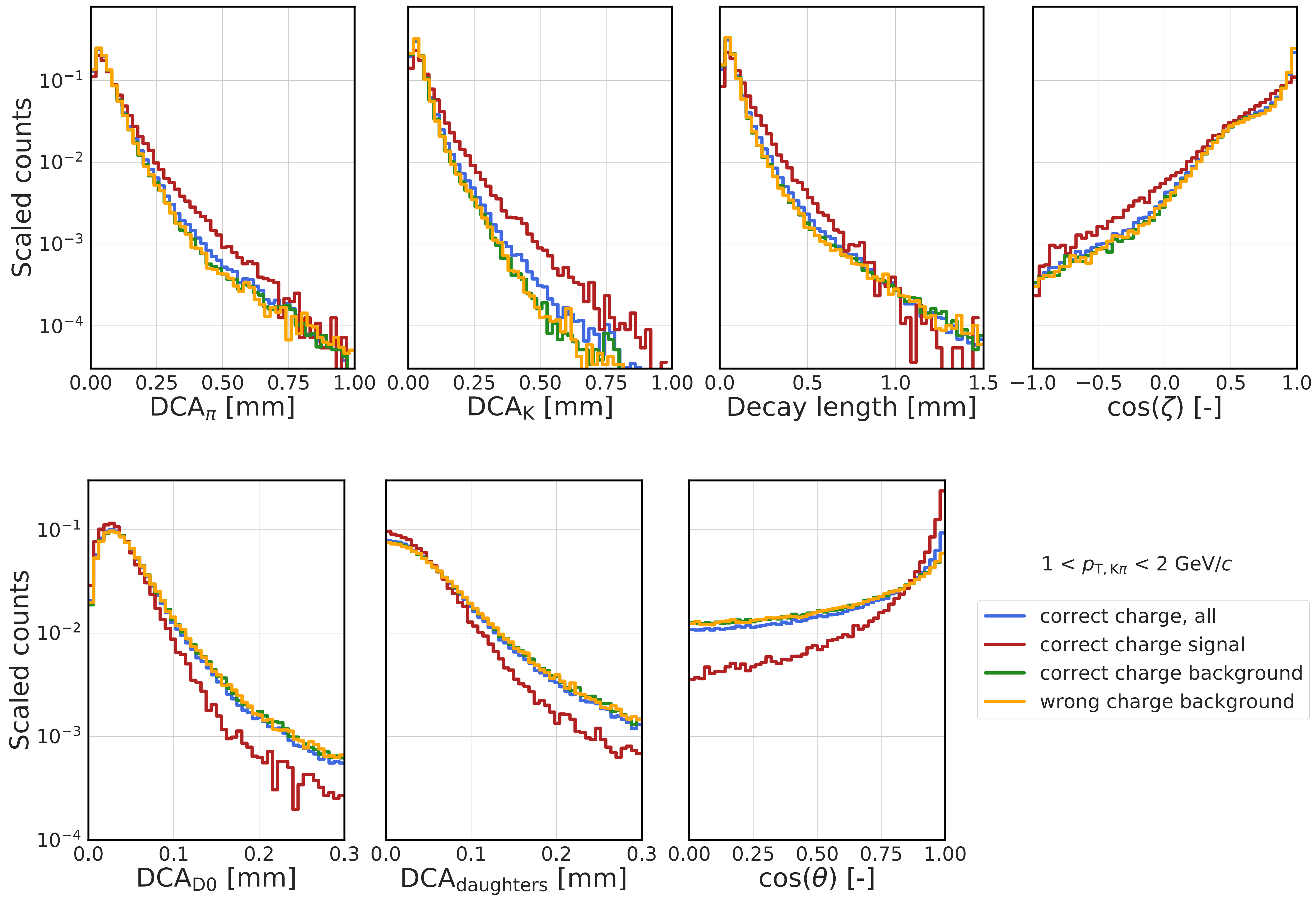
- **cosine of angle ζ** between D^0 momentum and kaon momentum

- **DCA_{D0}** of D^0 meson to PV
  - ideally $\rightarrow$ 0 µm

- **$DCA_{K,\pi}$** of kaon and pion to PV
  - ideally >> 0 µm

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Distribution of topological variables



- Correct and wrong charge combination pairs have **nearly similar distributions** which makes D^0 separation challenging

- D^0 signal shape is as we expect from topological properties of the decay

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Data pre-processing

- Each reconstructed correct-charge Kπ pair is associated to the tracks in the HIJING simulation
- Those coming from the real D^0 meson in the simulation are used as **signal** sample in the classification training, other correct-charge pairs are considered to be the **background**
- For the testing phase of the methods, combined signal and background correct-charge pairs are used
- All of the methods are independently trained and tested in the five D^0 meson **transverse momentum $p_{T,D0}$ intervals**
- $p_{T,D0}$ intervals for training: **0–1, 1–2, 2–3, 3–5 and 5–8 GeV/*c***
- For ML training, datasets **with** and **without** precuts were tested

All presented models were trained and evaluated over balanced dataset:
number of signal samples ≈ number of background samples

Variable	Min.	Max.
$DCA_{K,\pi}$ [mm]	0.002	2.0
$DCA_{daughters}$ [mm]	0.0	0.2
decay length [mm]	0.005	2.0
DCA_{D0} [mm]	0.0	0.5
$\cos\theta$	0.7	1.0

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Random forest

- **Random forest (RF):**
  - scikit-learn package used
  - Machine learning algorithm built with ensemble of independent trees trained using bagging (bootstrap aggregating)
  - Parameters may be optimized during the training:
    - maximum depth of trees
    - number of trees within ensemble
    - function used to measure quality of split of samples within tree's node (impurity measure)

- **Hyper-parameter space for RF optimization:**
  - maximum depth of trees d_{max}: {5, 7, 10, 12, 15, 17, 20}
  - number of trees N_{trees}: {200, 400, 600, 800, 1000, 1200}
  - impurity measure: {Gini, Entropy}

- **Dataset split:**
  - 60% of the samples used for training and hyper-parameter optimization (grid search) using 3-fold cross-validation
  - 40% of the samples used for classifier evaluation

Optimized parameters of random forest classifier

	With precuts					Without precuts				
$p_{\mathrm{T,D0}}$ [GeV/c]	0-1	1-2	2-3	3-5	5-8	0-1	1-2	2-3	3-5	5-8
d_{max}	10	13	12	10	7	15	15	12	12	7
N_{trees}	1200	600	1000	200	200	1200	600	1000	1200	200
Impurity measure	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy

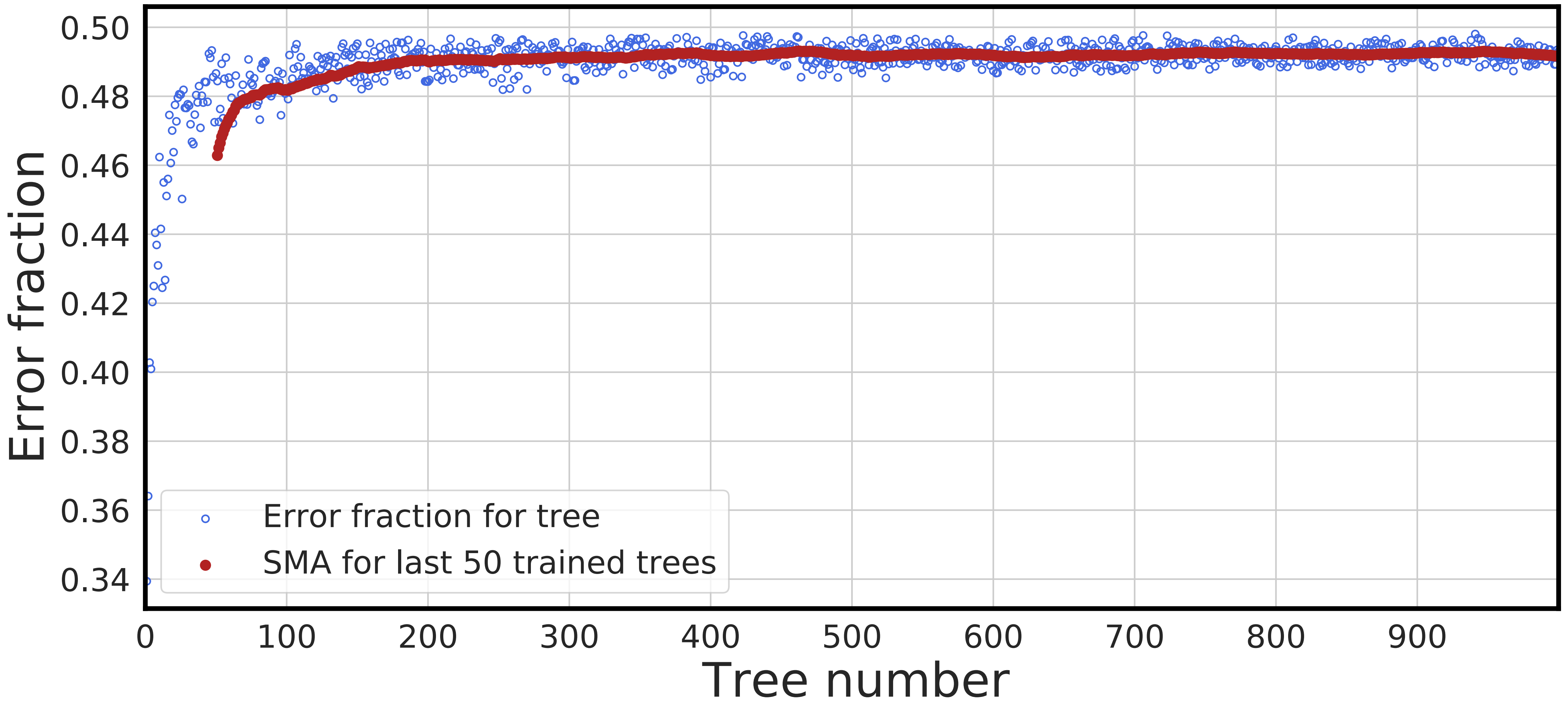
FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Boosted decision trees

- **Boosted decision trees (BDT):**
  - TMVA package used
  - Ensemble of shallow trees linked together using boosting algorithm (AdaBoost)
  - Training of individual trees is not independent, data misclassified by a tree have higher weight in the training of the next tree (class weights are balanced using weights coming from classification error fraction)

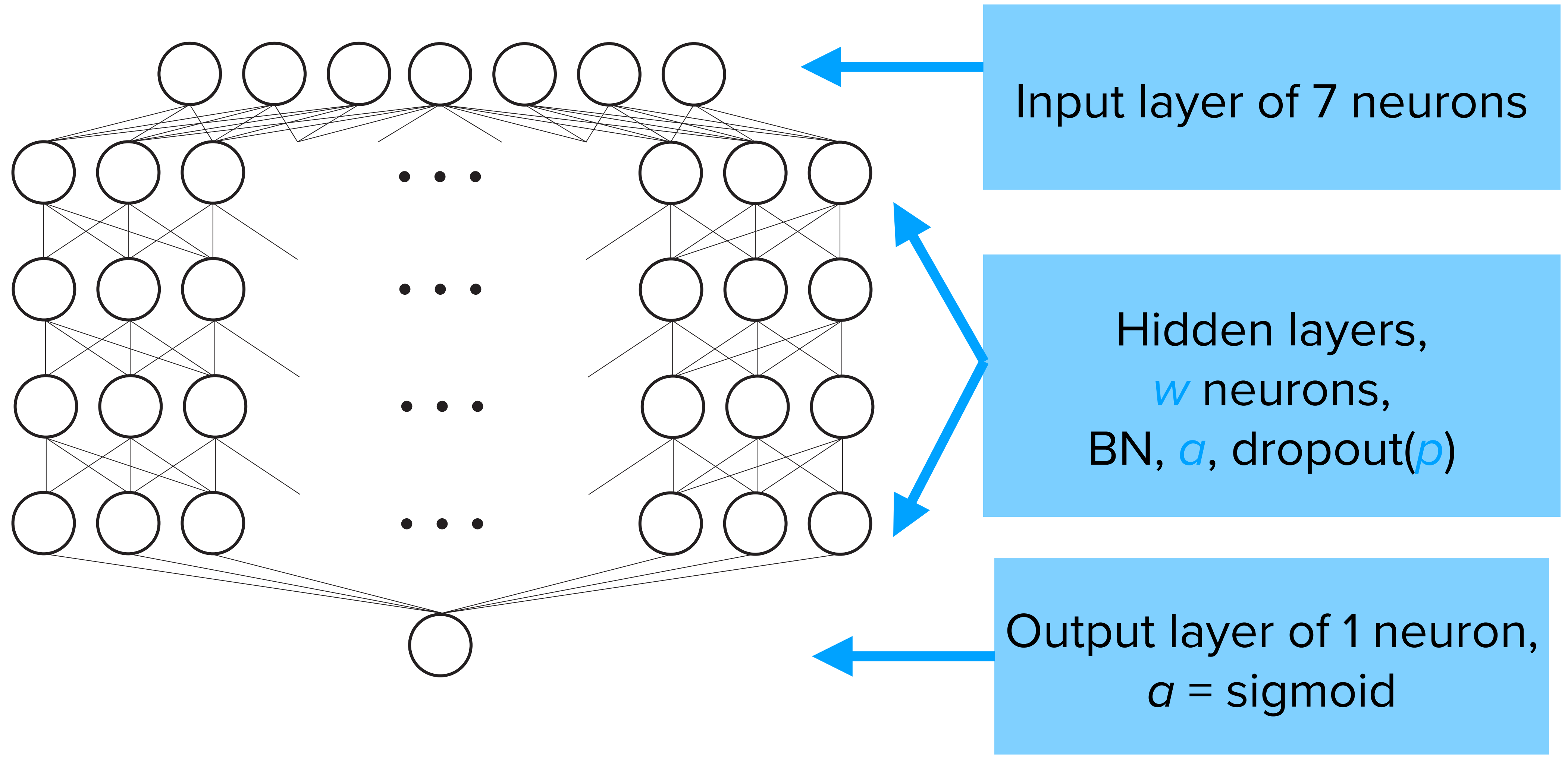
- **Error fraction:**
  - Calculated as how much signal events have positive BDT response (defined in [-1,1] range) and vice-versa for the background
  - Error fraction is getting higher with increasing N_{trees} and converges
  - The trees with error fraction close to 0.5 are very weak classifiers

- **BDT parameters:**
  - N_{trees} : selected as 500 (based on error fraction convergence)
  - Impurity measure: Gini
  - maximum tree depth d_{max}: 3
  - bagging is applied to minimize overtraining



FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Deep neural network



- Keras package used
- Supervised machine learning algorithm able to perform non-linear transformation of input space
- May address complex classification problems more effectively when optimized properly
- Requires large amount of CPU time, computational power and training data
- Feature-wise standardization was applied to topological variables

Hyper-parameter space for deep neural network setup optimization

Parameter	Activation a	Init. Learning rate $[\cdot 10^{-5}]$	Dropout rate p	Optimizer	Kernel init.	Batch size	Neurons in hidden layer w
	ReLU	0.1	0.2	Adam	Lecun uniform	64	64
Options	Tanh	0.5	0.3	Stoch. grad. Descent	He Normal	128	128
		1.0	0.4	AdaMax	He uniform		256

Deep neural network

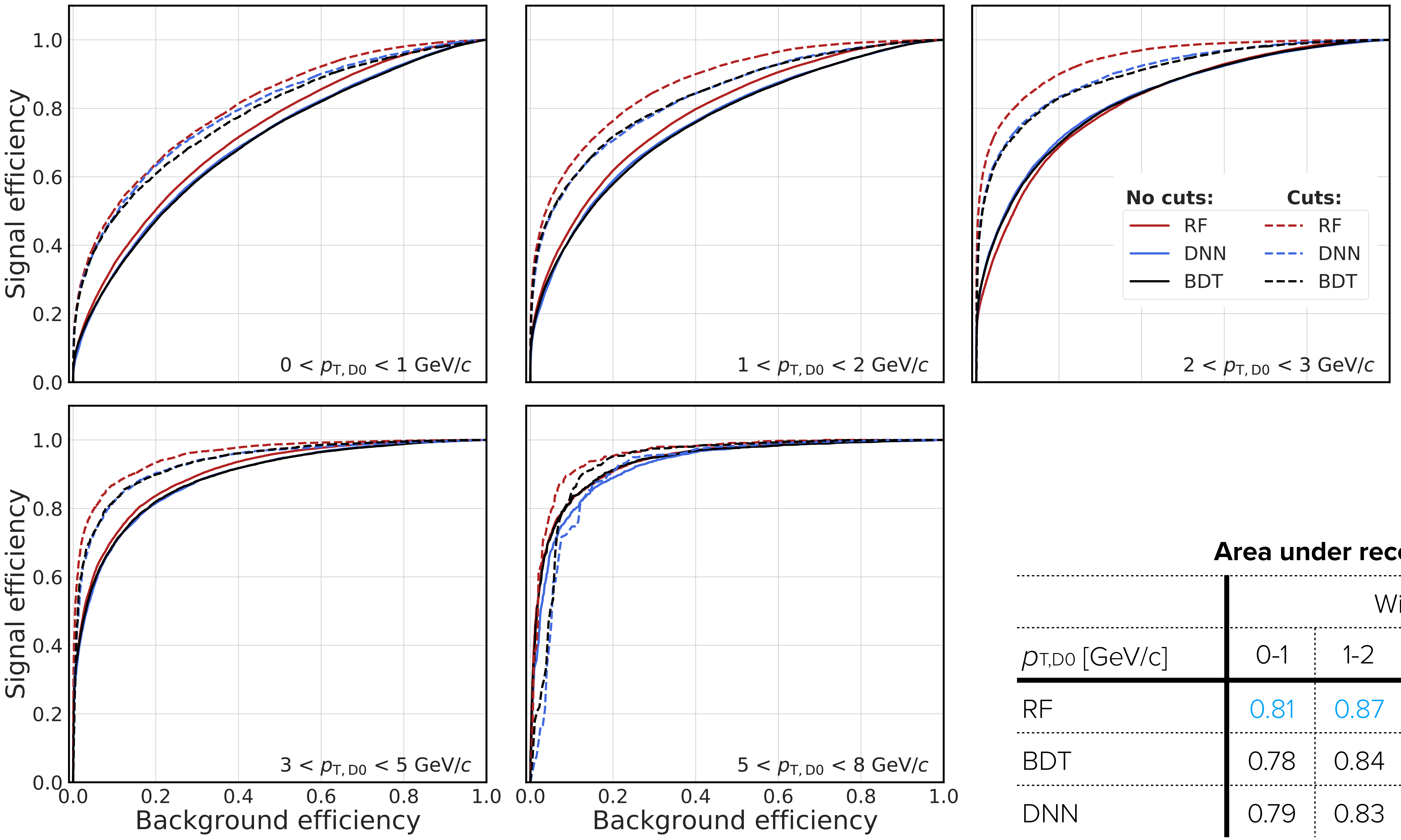
- **Dataset split:**
  - **training data:** 40% of the samples used for training and hyper-parameter optimization (pre-training 70 epochs)
  - **validation data:** 20% of the samples used to evaluate the pre-trained models (selection of best hyper-parameters combination)
  - **training data** are then used again to train model with optimized hyper-parameters, training is stopped with EarlyStopping rule only (validation ACC was not improved during last 60 epochs)
  - **test data:** 20% of the samples used to evaluate the final model

Optimized parameters of deep neural network

	With precuts					Without precuts				
$p_{T,D0}$ [GeV/c]	0-1	1-2	2-3	3-5	5-8	0-1	1-2	2-3	3-5	5-8
Activation a	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	tanh
Init. Learning rate [$\cdot 10^{-5}$]	1	1	1	1	1	1	1	1	1	1
Dropout rate p	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Kernel init.	He Normal	He uniform	Lecun unif.	Lecun unif.	Lecun unif.	He Normal	Lecun unif.	Lecun unif.	Lecun unif.	Lecun unif.
Batch size	64	64	64	64	64	64	64	64	64	64
Neurons in hidden layer w	256	256	256	256	256	256	256	256	256	25ž

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Evaluation of receiver operating characteristics over test set



- Classification ability increases with higher $p_{T,D0}$
- For $p_{T,D0}$ < 5 GeV/c, classifiers trained over the data after preselection criteria (cuts) application achieve higher AUC values as those trained without their application.
- RF classification has significantly better performance than other methods

Area under receiver operating characteristic curve (AUC)

	With precuts					Without precuts				
$p_{T,D0}$ [GeV/c]	0-1	1-2	2-3	3-4	4-5	0-1	1-2	2-3	3-4	4-5
RF	0.81	0.87	0.94	0.95	0.96	0.73	0.79	0.82	0.90	0.94
BDT	0.78	0.84	0.90	0.93	0.93	0.70	0.76	0.83	0.89	0.94
DNN	0.79	0.83	0.90	0.93	0.91	0.70	0.76	0.83	0.89	0.92

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Conclusion

- Various machine learning methods (RF, BDT, DNN) were compared in order to obtain efficient classifier for finding signal in two-body decay of D^0 meson
- HIJING full-event simulation of 2016 d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV was used to study $D^0 \to K\,\pi$ decay channel
- Tight cut preselection and the option with no preselection criteria was applied to topological variables
- Despite the fact, that RF method is not common in high-energy physics, it shows notable enhancement of signal efficiency compared to the other presented methods.
- DNN in presented setup does not show significant improvement compared to the widely used BDT.
  - However, their performance is close and the DNN might help to improve signal significance in similar analyses

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

Backup

Signal and background efficiencies vs. reconstructed transverse momentum of Kπ pairs