

Application of science-informed AI in experimental particle physics and neuroscience (Poster #85)

Péter Lévai and Gergő Orbán
Wigner Research Centre for Physics,
Hungarian Research Network (HUN-REN)
EuCAIFCon, Amsterdam, 30 April 2024

Department of High Energy Physics:

CERN LHC CMS, ALICE and CERN FCC

LVK (LIGO/VIRGO/KAGRA) and Einstein Telescope (ET)

Department of Computational Science:

Natural and Artificial Intelligence, Neuroscience

Brain Research, AI/ML/Deep Learning

Wigner Datacenter: CERN Grid, GPU, HPC, Quantum comp.



KFKI Budapest → KFKI RMKI →
→ HUN-REN Wigner RCP

**Institute for Particle and
and Nuclear Physics (1952)**

Particle and nuclear physics, gravity,
Detectors/accelerators, neutron stars

Functional materials, space science
ML/AI/Brain res. & IT (GPU, HPC)

Quantum comp., QTech/Qcommun.

Basic problem of neuroscience and brain research:

- Brain sciences lack a unifying mathematical theory of brain function, theories for individual faculties are not available either.**

Innovative answers from neuroscientists:

- Generative modelling framework has been widely adopted to explore mathematical principles in a data-driven manner: generative models are probabilistic models that assume that observations are a result of a (nonlinear) combination of latent factors that correspond to relevant quantities (e.g. physical measures);**
- Neuroscience has adopted an open-ended approach to learn about the mathematical principles by constraining them with an ever-more-complex approach to data: instead of constraining data to more-and-more controlled settings, rich and little-constrained data is flexibly interpreted with AI-borrowed deep generative models;**
- Diffusion models, variational autoencoders, contrastive learning methods provide a spectrum of opportunities to integrate complex nonlinear generative models with physical intuitions as inductive biases, or generalize interpretations across existing experiments, as well as to new experiments.**

Basic problem of HEP research:

- A unified mathematical theory (Standard Model) describes experimental data with high precision, but we do not understand the origin of the „free parameters” and do not see hints for Beyond SM phenomena – although numerous candidates exist to become the winner mathematical model.

Basic problems of Nuclear Physics research:

- The World of strongly interacting many-body systems is very rich, no unified mathematical theory describes experimental data with reasonable precision;
- Numerous phenomenological description has been invented, but the connection between these models are weak in many cases.

Basic problems of Astroparticle (and Astro)Physics:

- Enormous amount of data arrive from the new instruments (telescopes, detectors) and the understanding of these data is focusing into a very narrow target direction;
- On the other hand the usual expectation is to understand the multimessenger data in a unified frame, answering basic questions about the investigated objects;
- Numerous phenomenological description has been invented, but the connection between these models and connection to HEP is not well established (see e.g. the problem of „dark matter” and „dark energy”).

How does Artificial Intelligence and Machine Learning could help? (What could we learn from neuroscientists?)

- Probabilistic approaches already receive wider support in HEP
(see e.g. the separation of gluon-jets and quark-jets during their study and detection)
- ML applications became part of the usual routine protocol in data analyses,
self-improving cycles are capable to increase the precision
(although the request of CPU-time is enormous, it is limiting the applications)
- Generative model applications are widely adopted to explore mathematical
principles in a data-driven manner: the identification and exploration of latent factors
and their non-linear combination is in the focus of recent analysis to discover new
knowledge element hiding behind the phenomenological descriptions;
- During the analyses of huge and complex datasets (including high-resolution pictures and
time evolution with small timesteps --- see e.g. multimessenger astrophysics)
the application of diffusion models, variational autoencoders, contrastive learning
methods could provide a spectrum of opportunities to integrate complex nonlinear
generative models with physical intuitions as inductive biases, and improve
the interpretations and understanding of existing data (see e.g. LSST mission).

Cross-fertilization between different fields and disciplines could become very useful !

Useful Links:

HUN-REN Wigner Research Centre for Physics:

<https://wigner.hun-ren.hu/en/>

14. GPU Days: Meeting on Massive Parallel Computing

Date: 30-31 May 2024, Budapest, Hungary

<https://gpuday.com/>

6. HEPTECH AIME on AI/ML and Quantum Computing

Date: 18-19 November 2024, Budapest, Hungary

5. AIME: <https://indico.wigner.hu/event/1523/>

Wigner Datacenter at Wigner RCP

<https://wignerdc.wigner.hu/home>

Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

¹TTK RWTH Aachen University ²IEP Universität Hamburg ³CDCS Universität Hamburg ⁴NHETC Rutgers University

Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

Motivation

- To find new physics, improve largely model agnostic searches, e.g., resonance searches
- Use pattern recognition capability of machine learning in high dimensional feature space to gain higher sensitivity
- Problem: Currently many papers use only 4 high level features ("baseline" feature set) on one benchmark dataset (LHCO R&D dataset [1])
- For more model agnostic setup need to be able to use more features

Goal: Improve classifier setup for more high level features and low level features

Weakly supervised anomaly detection

Classification Without Labels (CWoLa) [2]

- Classifier between mixed datasets $p_S(x) = f_1 p_S(x) + (1 - f_1) p_B(x)$ with signal fractions f_i
- $R_{\text{mixed}} = \frac{f_1 R_{\text{signal}}(x) + (1 - f_1) R_{\text{background}}(x)}{f_1 R_{\text{signal}}(x) + (1 - f_1) R_{\text{background}}(x)}$ where $R_{\text{optimal}}(x) = \frac{p_S(x)}{p_B(x)}$
- is the optimal classifier between signal and background distributions $p_{S/B}$
- Mathematically equivalent as R_{mixed} monotonous in R_{optimal}

Application to resonance searches

- Divide data into signal region (SR) and sideband (SB), where $p_S(x) = p_S(x|m \in \text{SR}) + p_B(x|m \in \text{SB})$ and $p_B(x) = p_B(x|m \in \text{SB})$
- for classification features x .
- Construct "background template" from SB, ideally with $p(x) = p_B(x|m \in \text{SB})$
- Here, we use idealized case to study classifier only

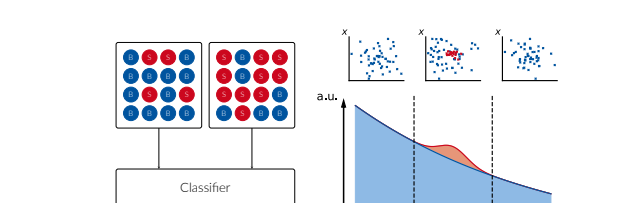


Figure 1: Left: Sketch of weakly supervised classification setup. Right: Division of data into SB and SR for a resonance search.

BDTs for high level features [3]

Machine Learning background

Boosted Decision Trees (BDTs) are known to be very effective on tabular data, especially for small datasets [4].

- Few signal events → small effective dataset
- High level features → tabular data

Classifier Setup

- NN: Ensemble of N fully connected neural networks
- BDT: Ensemble of N gradient boosted decision trees

Study: Uninformative features

We study the classifiers by introducing uninformative features (features drawn from Gaussian noise), which the NN is particularly sensitive to. The BDT's performance is very robust, meaning that we can add more features to an analysis.

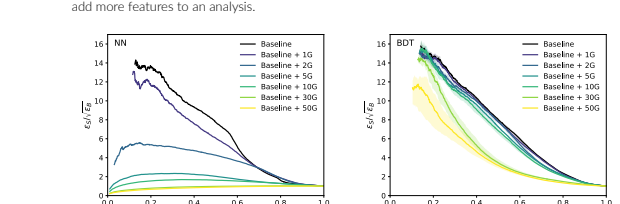


Figure 2: ROC curves of MAD NN/BDT classifiers employing four baseline features and additional Gaussian features. For 30 and 50 Gaussian features, ensembling of BDT increased to $A_{\text{ROC}} = 38\%$ otherwise $A_{\text{ROC}} = 36\%$.

Study: Additional physics-motivated features

We study datasets with more subtlety-based features.

- Extended set 1: 10 features (baseline + 6 additional), some largely uninformative
- Extended set 2: 12 features, all slightly informative
- Extended set 3: 56 features, all slightly informative

BDT robustness against uninformative features translates to being well-behaved with additional features. Not present for NN.

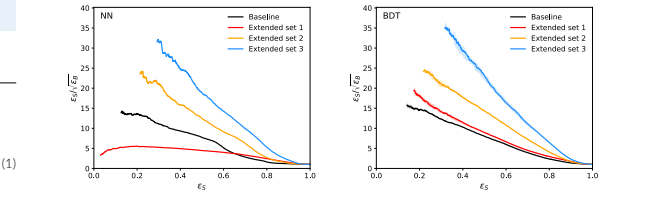


Figure 3: ROC curves of MAD NN/BDT classifier with 4 feature baseline dataset and three extended feature sets.

Graphs for low level features

Machine Learning background

Graph Neural Networks can represent IHEP data in a permutation invariant manner. Architectures can incorporate symmetries directly.

- Very successful on top tagging tasks

Study: Top tagger on LHCO dataset

State of the art top taggers were studied on the LHCO R&D dataset:

- Modified LovelaceNet architecture [5] found to result in the best performance.
- Performance drops sooner than observed for high level features.

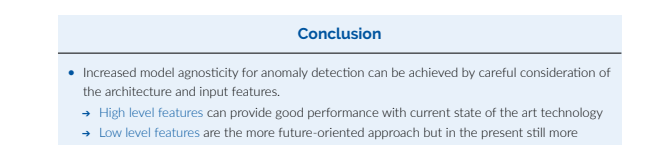


Figure 4: ROC curves for supervised classifier and MAD on low level features.

Conclusion

- Increased model agnosticity for anomaly detection can be achieved by careful consideration of the architecture and input features.
- High level features can provide good performance with current state of the art technology
- Low level features are the more future-oriented approach but in the present still more difficult to achieve

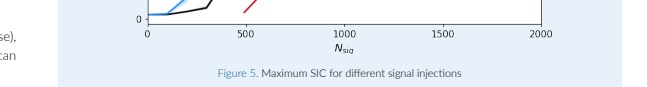


Figure 5: Maximum SIC for different signal injections

References

[1] G. Kasieczka et al., "The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics," *Phys. Rev. D*, vol. 102, no. 3, p. 034003, 2020.

[2] E. M. Abdellatif, B. Kniehm, and J. Thaler, "Classification without labels: Learning from mixed samples in high-energy physics," *JHEP*, vol. 05, p. 174, 2017.

[3] Y. Geng, D. Jiang, J. Jiang, R. Liu, C. Liu, S. Chen, W. Du, Z. M. Ma, and T. Y. Li, "An efficient context-aware graph neural network for jet tagging," *JHEP*, vol. 07, p. 030, 2022.

[4] "Tree-based algorithms for weakly supervised anomaly detection," *Phys. Rev. D*, vol. 102, no. 3, p. 034003, 2020.

[5] L. G. Corcoran, E. Chinellato, and G. W. Moore, "Why do the best models still outperform deep learning on tabular data?"

Weakly supervised anomaly detection can be applied to resonance searches to find BSM physics.

Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Anomaly Detection

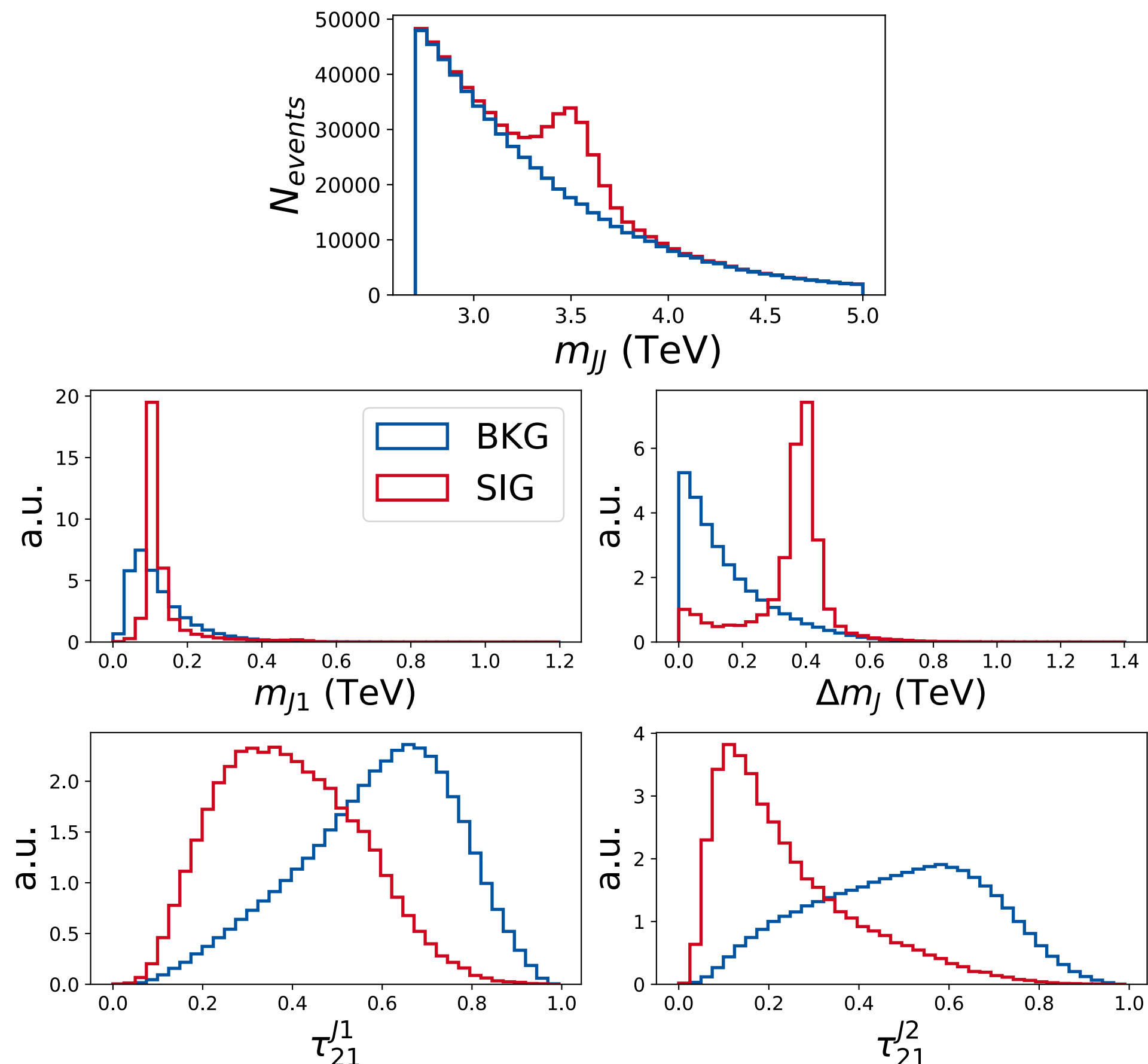
Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

¹TTK RWTH Aachen University ²IEP Universität Hamburg ³CDCS Universität Hamburg ⁴NHETC Rutgers University

LHCO R&D dataset

Background: QCD dijets
Signal:

Usual features for weak supervision papers
⇒ not very model agnostic



Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

Motivation

- To find new physics, improve largely model agnostic searches.
- Use pattern recognition capability of machine learning in high dimensional feature space to gain higher sensitivity.
- Problem: Currently many papers use only 4 high level features ("baseline" feature set) on one benchmark dataset (LHCO R&D dataset [1]).
- For more model agnostic setup need to be able to use more features.

Goal: Improve classifier setup for more high level features and low level features

Weakly supervised anomaly detection

Classification Without Labels (CWoLa) [2]

- Classifier between mixed datasets $p_S(x) = f_1 p_S(x) + (1 - f_1) p_B(x)$ with signal fractions f_i
- $R_{\text{mixed}} = \frac{f_1 R_{\text{optimal}}(x) + (1 - f_1)}{f_1 R_{\text{optimal}}(x) + (1 - f_1)}$ where $R_{\text{optimal}}(x) = \frac{p_S(x)}{p_B(x)}$ [2]
- is the optimal classifier between signal and background distributions p_S, p_B .
- Mathematically equivalent as R_{mixed} monotonous in R_{optimal} .

Application to resonance searches

- Divide data into signal region (SR) and sideband (SB), where
- $p_S(x) = p_S(x|m \in \text{SR}) + p_B(x|m \in \text{SR})$ and $p_B(x) = p_B(x|m \in \text{SB})$ [2]
- for classification features x .
- Construct "background template" from SB, ideally with $p(x) = p_B(x|m \in \text{SB})$
- Here, we use idealized case to study classifier only

BDTs for high level features [3]

Machine Learning background

Boosted Decision Trees (BDTs) are known to be very effective on tabular data, especially for small datasets [4].

- Few signal events → small effective dataset
- High level features → tabular data

Classifier Setup

- NN: Ensemble of N fully connected neural networks
- BDT: Ensemble of N gradient boosted decision trees

Study: Uninformative features

We study the classifiers by introducing uninformative features (features drawn from Gaussian noise), which the NN is particularly sensitive to. The BDT's performance is very robust, meaning that we can add more features to an analysis.

Study: Additional physics-motivated features

We study datasets with more subtlety-based features.

- Extended set 1: 10 features (baseline + 6 additional), some largely uninformative
- Extended set 2: 12 features, all slightly informative
- Extended set 3: 56 features, all slightly informative

BDT robustness against uninformative features translates to being well-behaved with additional features. Not present for NN.

Graphs for low level features

Machine Learning background

Graph Neural Networks can represent HEP data in a permutation invariant manner. Architectures can incorporate symmetries directly.

- Very successful on top tagging tasks

Study: Top tagger on LHCO dataset

State of the art top taggers were studied on the LHCO R&D dataset.

- Modified LovaszNet architecture [5] found to result in the best performance.
- Performance drops sooner than observed for high level features.

Conclusion

- Increased model agnosticity for anomaly detection can be achieved by careful consideration of the architecture and input features.
- High level features can provide good performance with current state of the art technology
- Low level features are the more future-oriented approach but in the present still more difficult to achieve

References

[1] G. Kasieczka et al., "The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics," Phys. Rev. D, vol. 102, no. 3, p. 034003, 2020.

[2] E. M. Abdalla, B. Fuhrmann, and J. Thaler, "Classification without labels: learning from mixed samples in high-energy physics," JHEP, vol. 05, p. 174, 2017.

[3] Y. Geng, D. Jiang, J. Jiang, R. Liu, C. Liu, S. Chen, W. Du, Z. M. Ma, and T. Li, "An efficient control variate graph neural network for jet tagging," JHEP, vol. 07, p. 030, 2022.

[4] "Tree-based algorithms for weakly supervised anomaly detection," Phys. Rev. D, vol. 107, no. 3, p. 034003, 2023.

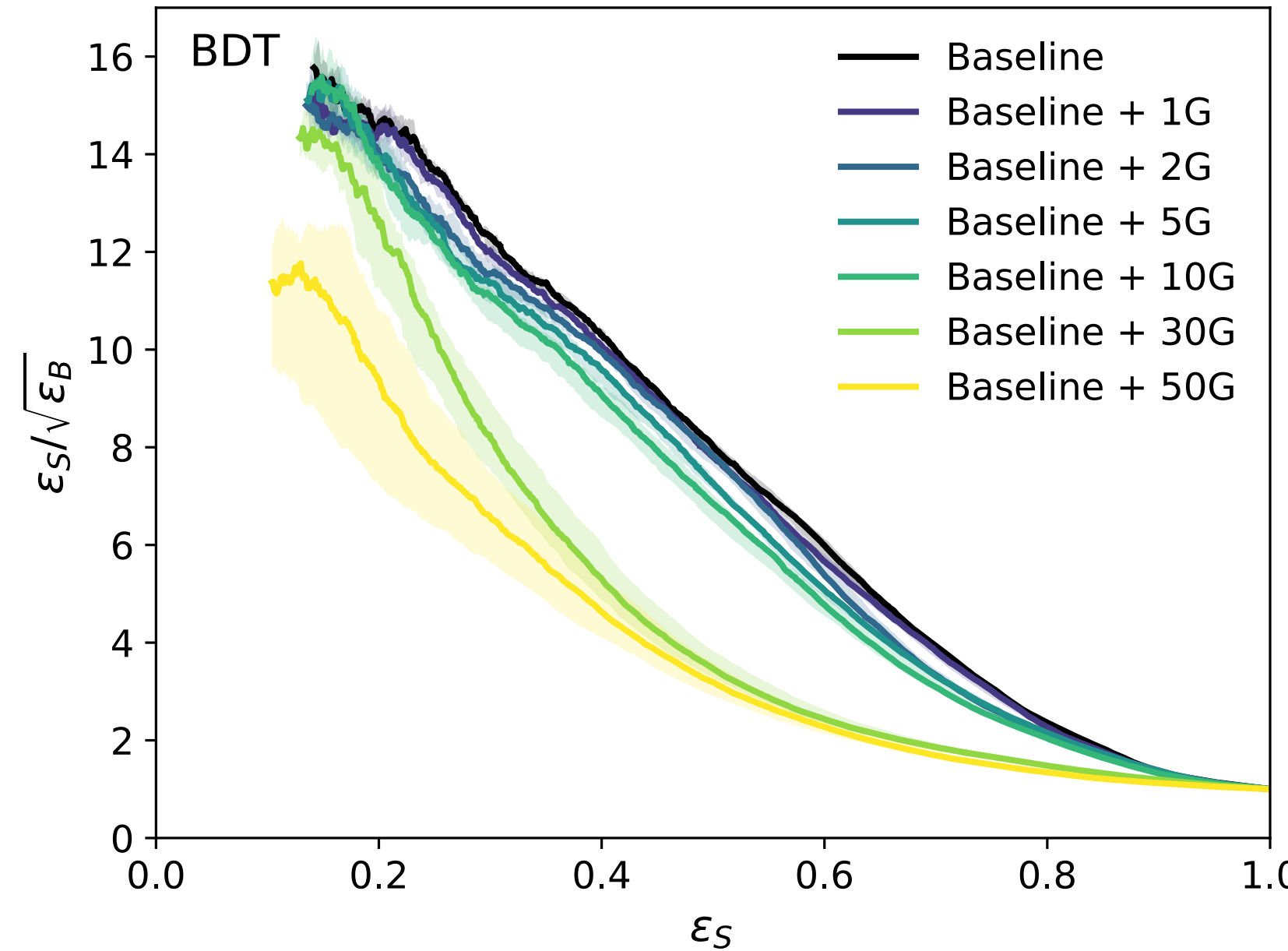
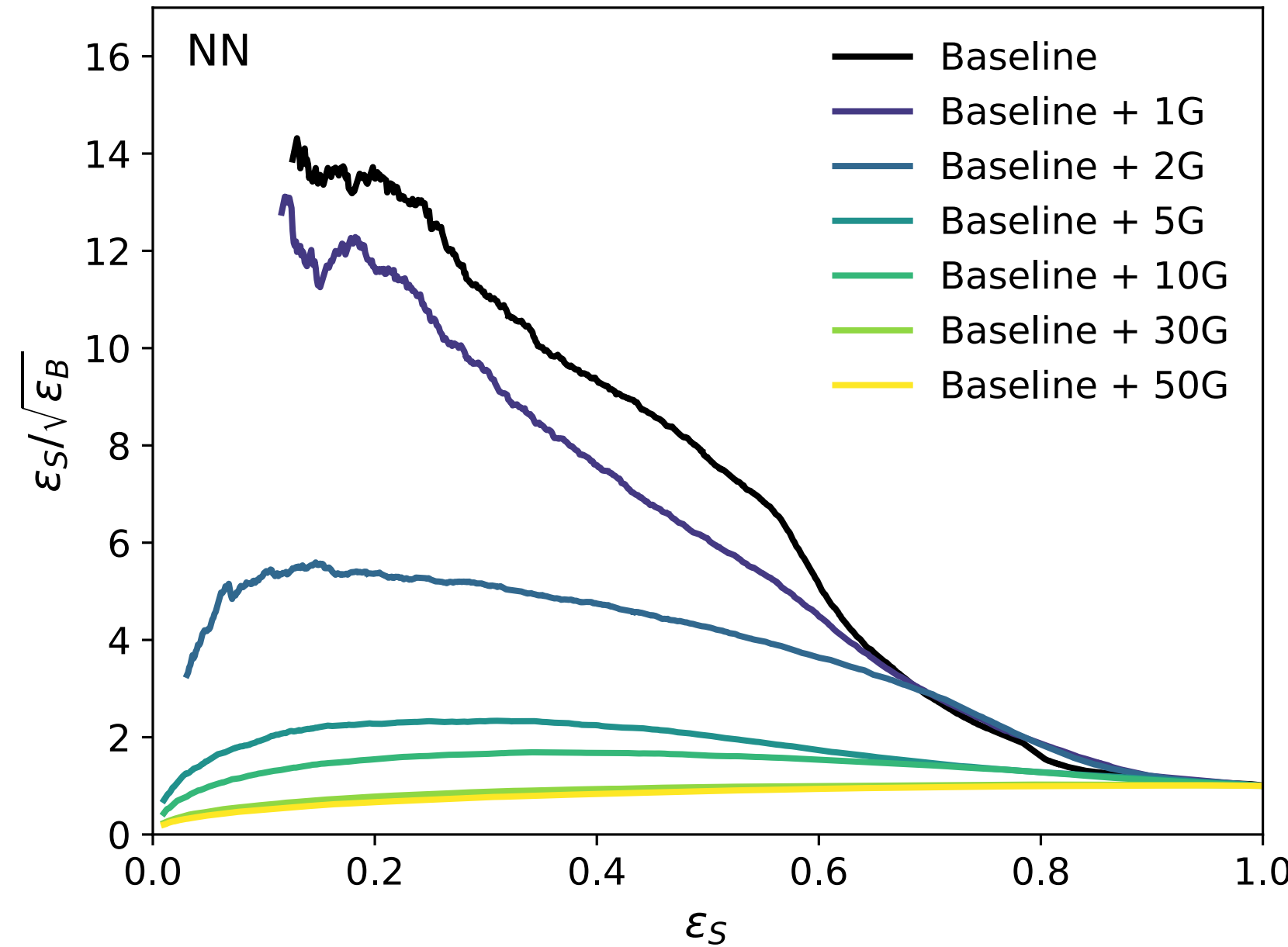
[5] L. Grottel, E. Chinellato, and G. Wenzel, "Who do the best models still outperform deep learning on tabular data?," arXiv:2305.17427, 2023.

[6] T. Fink, M. Hein, G. Kasieczka, M. Krämer, A. Mück, P. Prangchaikul, T. Quadfasel, D. Shih, and M. Sommerhalder, "The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics," Phys. Rev. D, vol. 102, no. 3, p. 034003, 2020.

Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

¹TTK RWTH Aachen University ²IEP Universität Hamburg ³CDCS Universität Hamburg ⁴NHETC Rutgers University



To include more features, robustness against uninformative features is necessary, which is not present for NNs.

Increasing the Model Agnosticity of Weakly Supervised Anomaly Detection

Thorben Finke¹ Joep Geuskens¹ Marie Hein¹ Gregor Kasieczka^{2,3} Michael Krämer¹ Alexander Mück¹ Parada Prangchaikul² Tobias Quadfasel² David Shih⁴ Manuel Sommerhalder²

TTK RWTH Aachen University ¹IEP Universität Hamburg ²CDCS Universität Hamburg ⁴NHETC Rutgers University

Motivation

- To find new physics, improve largely model agnostic searches, e.g., resonance searches
- Use pattern recognition capability of machine learning in high dimensional feature space to gain higher sensitivity
- Problem: Currently many papers use only 4 high level features ("baseline" feature set) on one benchmark dataset (LHCO R&D dataset [1])
- For more model agnostic setup need to be able to use more features
- Goal: Improve classifier setup for more high level features and low level features

Weakly supervised anomaly detection

Classification Without Labels (CWoLa) [2]

- Classifier between mixed datasets $p_S(x) = f_1 p_S(x) + (1 - f_1) p_B(x)$ with signal fractions f_i
- $R_{mixed} = \frac{f_1 R_{optimal}(x) + (1 - f_1)}{f_1 R_{optimal}(x) + (1 - f_1)}$ where $R_{optimal}(x) = \frac{p_S(x)}{p_B(x)}$ [1]
- Is the optimal classifier between signal and background distributions p_S, p_B
- Mathematically equivalent as R_{mixed} monotonous in $R_{optimal}$

Application to resonance searches

- Divide data into signal region (SR) and sideband (SB), where
- $p_S(x) = p_S(x|m \in SR) + p_B(x|m \in SR)$ and $p_B(x) = p_B(x|m \in SB)$ [2]
- for classification features x .
- Construct "background template" from SB, ideally with $p(x) = p_B(x|m \in SR)$
- Here, we use idealized case to study classifier only

BDTs for high level features [3]

Machine Learning background

Boosted Decision Trees (BDTs) are known to be very effective on tabular data, especially for small datasets [4].

1. Few signal events → small effective dataset
2. High level features → tabular data

Classifier Setup

- NN: Ensemble of N fully connected neural networks
- BDT: Ensemble of N gradient boosted decision trees

Study: Uninformative features

We study the classifiers by introducing uninformative features (features drawn from Gaussian noise), which the NN is particularly sensitive to. The BDT's performance is very robust, meaning that we can add more features to an analysis.

Study: Additional physics-motivated features

We study datasets with more subjectivity-based features.

- Extended set 1: 10 features (baseline + 6 additional), some largely uninformative
- Extended set 2: 12 features, all slightly informative
- Extended set 3: 56 features, all slightly informative

BDT robustness against uninformative features translates to being well-behaved with additional features. Not present for NN.

Graphs for low level features

Machine Learning background

Graph Neural Networks can represent HEP data in a permutation invariant manner. Architectures can incorporate symmetries directly.

- Very successful on top tagging tasks

Study: Top tagger on LHCO dataset

State of the art top taggers were studied on the LHCO R&D dataset:

- Modified LovaszNet architecture [5] found to result in the best performance.
- Performance drops sooner than observed for high level features.

Conclusion

- Increased model agnosticity for anomaly detection can be achieved by careful consideration of the architecture and input features.
- High level features can provide good performance with current state of the art technology
- Low level features are the more future-oriented approach but in the present still more difficult to achieve

References

[1] G. Kasieczka et al., "The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics," *Phys. Rev. D*, vol. 102, no. 3, p. 034003, 2020.

[2] E. M. Abdurrobbil, R. Bernabini, and J. Thaler, "Classification without labels: Learning from mixed samples in high-energy physics," *JHEP*, vol. 05, p. 174, 2017.

[3] Y. Gao, D. Jiang, J. Wang, R. Du, C. Li, S. Chen, W. Du, Z. M. Ma, and T. Y. Lu, "An efficient context-aware graph neural network for jet tagging," *JHEP*, vol. 07, p. 030, 2022.

[4] "Tree-based algorithms for weakly supervised anomaly detection," *Phys. Rev. D*, vol. 102, no. 3, p. 034003, 2020.

[5] L. G. Gerber, E. Chinellato, and G. W. Moore, "Who do the best models still outperform deep learning on tabular data?" *arXiv preprint arXiv:2104.00001*, 2021.

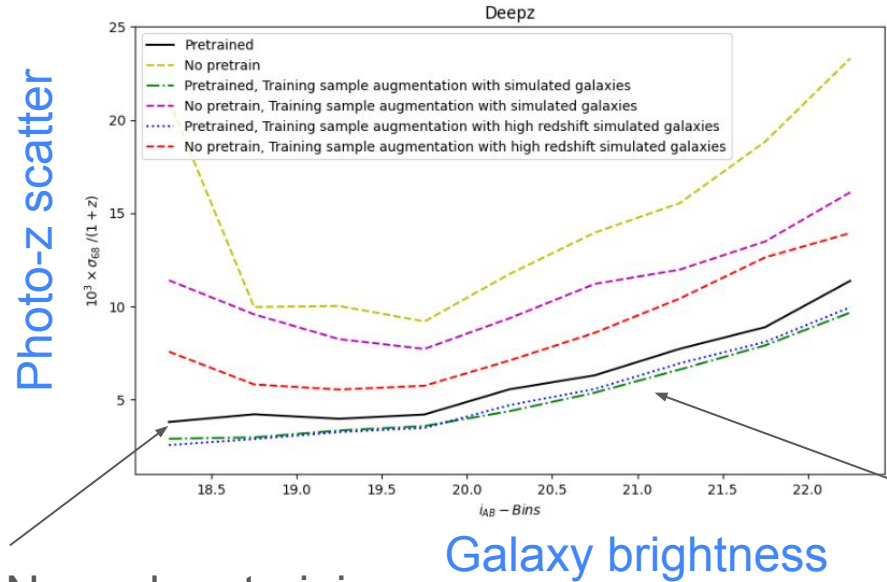
Galaxy redshift estimations with transfer and multi-task learning

M. Eriksen [eriksen@pic.es], L.Cabayol, H.Guo - IFAE-PIC, Barcelona

- Cosmology requires redshift estimations for large number of galaxies.
- Image galaxies in different bands and determine redshift as an inverse problem.
- **Challenge**: Inferring galaxy distances with small and biased training samples.



Transfer learning from simulations



Normal pretraining

Galaxy brightness

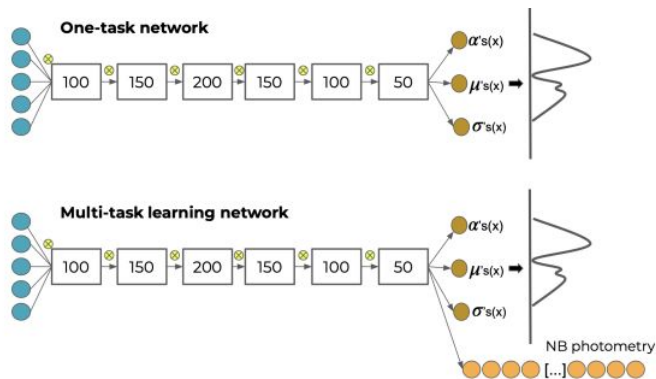
No simulated data

- Deepz is a deep neural network for photo-z estimation.
- Achieved state-of-the-art results on narrow-band photometry.
- Combining simulated data is key.

Better transfer learning scheme

arXiv: 2004.07979

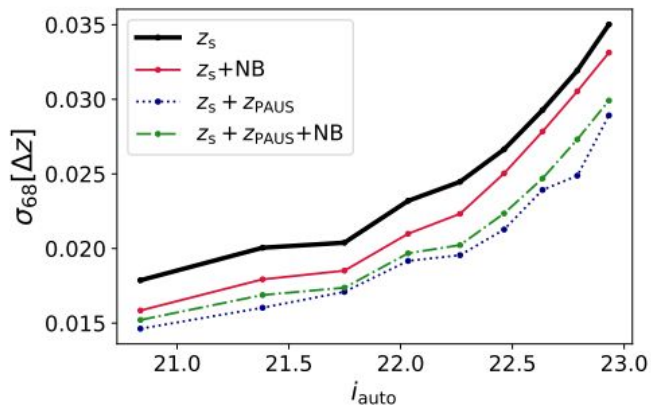
Multi-task learning



Problem: How to benefit from PAUS NB, which only covers 0.3% sky-area of *Euclid*.

Solution: Multi-task learning, predicting PAUS narrow bands (top plot).

Result: Reduces the photo-z scatter for all galaxies (bottom plot).



arXiv: 2209.10161

Gradient-Annihilated PINNs for Solving Riemann Problems: Application to Relativistic Hydrodynamics

Antonio Ferrer-Sánchez

IDAL, Electronic Engineering Department,
ETSE-UV, University of Valencia.

Valencian Graduate School and Research
Network of Artificial Intelligence (ValgrAI),
Spain.

(Antonio.Ferrer-Sanchez@uv.es)

José D. Martín-Guerrero

IDAL, Electronic Engineering Department,
ETSE-UV, University of Valencia.

Valencian Graduate School and Research
Network of Artificial Intelligence (ValgrAI),
Spain.

(jose.d.martin@uv.es)

Roberto Ruiz de Austri

Instituto de Física Corpuscular CSIC-UV.

([rruiz@ific.uv.es](mailto:r Ruiz@ific.uv.es))

José A. Font

Department of Astronomy and Astrophysics,
University of Valencia.

(j.antonio.font@uv.es)

Alejandro Torres-Forné

Department of Astronomy and Astrophysics,
University of Valencia.

(alejandrotorres@uv.es)

VNIVERSITAT
ID VALÈNCIA

Department of
Astronomy and Astrophysics



Keywords

Riemann problem \longrightarrow Problems with discontinuous initial conditions

Neural networks \longrightarrow Physics-Informed Neural Networks

Euler equations

Relativistic hydrodynamics

Problems in Astrophysics:
New application

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot f(\mathbf{u}) = S(\mathbf{u}) \longrightarrow$$

$$\left. \frac{\partial}{\partial t} \begin{pmatrix} D \\ S \\ \tau \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} Du \\ Su + p \\ S - Du \end{pmatrix} = 0 \right\} \rho, \mathbf{u}, p$$

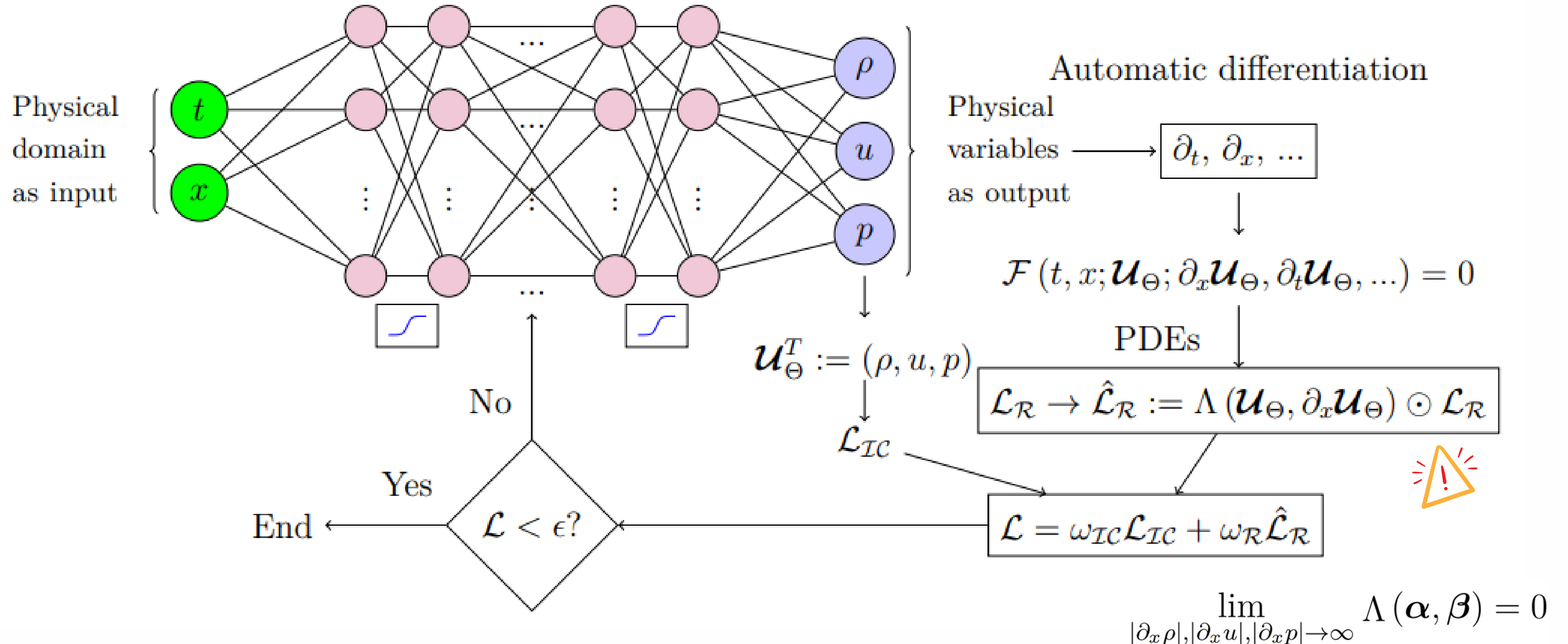
Relativistic densities of mass, momentum and energy, respectively.

Density, velocity and pressure of the fluid: primitive variables.



METHODOLOGY PROPOSED

Diagram and algorithm



SOME RESULTS

Riemann problems in Relativistic Hydrodynamics

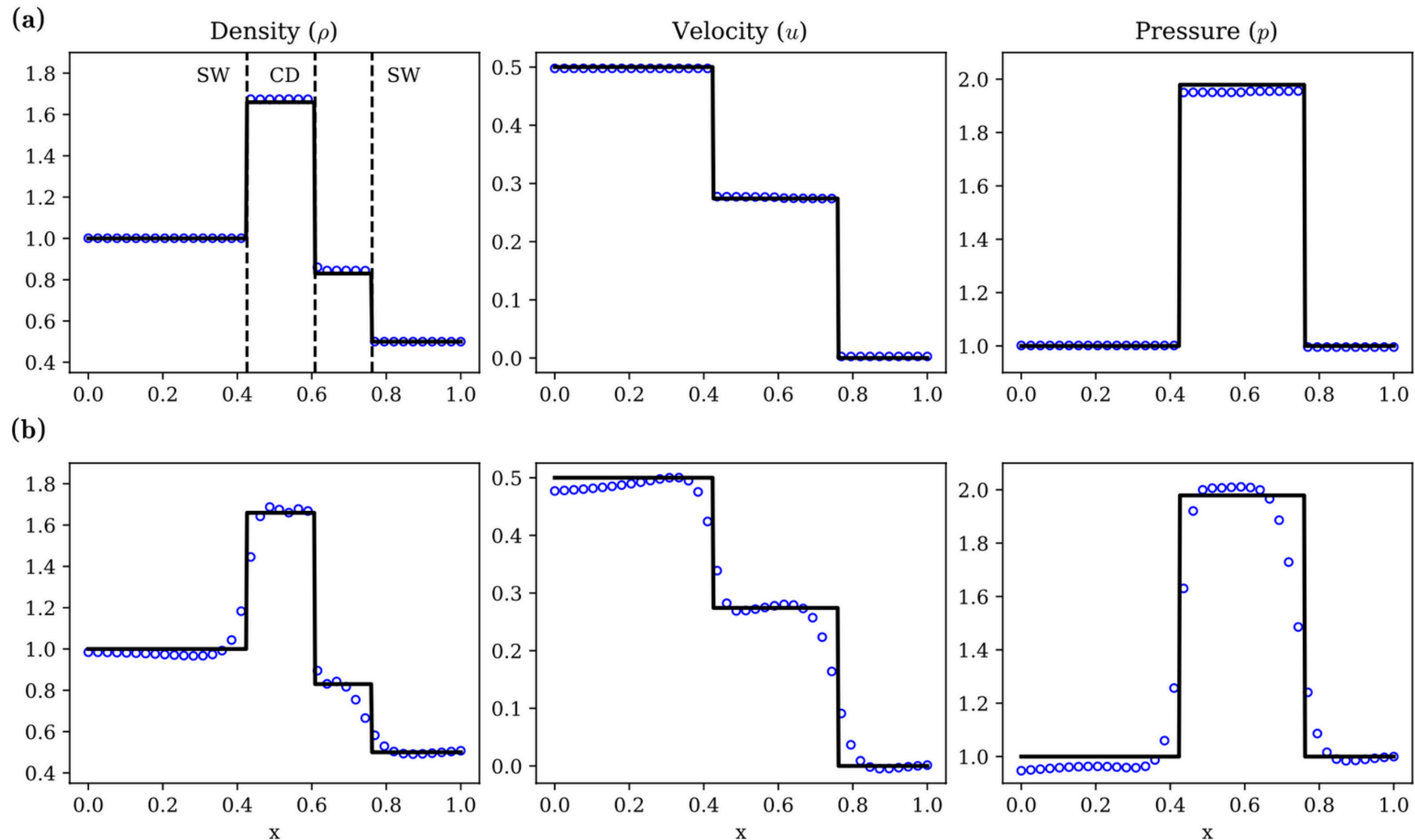


Figure 2: Final profiles for the primitive variables (density, velocity and pressure) with respect to the analytical solution (black solid line) obtained by the GA-PINN in (a) and by a vanilla PINN model in (b).