

# Advancing Physics Data Analysis through Machine Learning and Physics-Informed Neural Networks

Vasileios Vatellis,<sup>1,2</sup>

<sup>1</sup>*Institute of Informatics and Telecommunications, National Centre for Scientific Research "Demokritos",  
15310 Aghia Paraskevi, Athens, Greece*

<sup>2</sup>*School of Rural, Surveying and  
Geoinformatics Engineering, National Technical University of Athens,  
Zografou 157 73, Greece*

## Abstract

In an era increasingly focused on green computing and explainable AI, revisiting traditional approaches in theoretical and phenomenological particle physics is paramount. This project evaluates various machine learning (ML) algorithms—including Nearest Neighbors, Decision Trees, Random Forest, AdaBoost, Naive Bayes, Quadratic Discriminant Analysis (QDA), and XGBoost—alongside standard neural networks and a novel Physics-Informed Neural Network (PINN) for physics data analysis. We apply these techniques to a binary classification task that distinguishes the experimental viability of simulated scenarios based on Higgs observables and essential parameters. Through this comprehensive analysis, we aim to showcase the capabilities and computational efficiency of each model in binary classification tasks, thereby contributing to the ongoing discourse on integrating ML and Deep Neural Networks (DNNs) into physics research. In this study, XGBoost emerged as the preferred choice among the evaluated machine learning algorithms for its speed and effectiveness, especially in the initial stages of computation with limited datasets. However, while standard Neural Networks and Physics-Informed Neural Networks (PINNs) demonstrated superior performance in terms of accuracy and adherence to physical laws, they require more computational time. These findings underscore the trade-offs between computational efficiency and model sophistication.

*Keywords:* Machine Learning, Artificial Intelligence, Computer Science

*DOI:* 10.31526/JMLFS.2025.549

## 1. INTRODUCTION

Particle physics involves complex phenomena, such as neutrino masses and dark matter, which challenge existing theoretical frameworks. To explore these phenomena, it is essential to test various theoretical models against experimental data. This study focuses on enhancing the efficiency of this testing process by integrating machine learning (ML) techniques. Specifically, we introduce an ML filter to streamline the identification of experimentally viable parameter sets, improving computational efficiency and model validation in particle physics simulations.

The initial stage of the physics analysis involves randomly initializing the free parameters of our theoretical model and computing the tree-level masses. These computed values are then input into the SPheno [1, 2] code, which calculates decay modes for supersymmetric particles and Higgs bosons, as well as production cross-sections in electron-positron annihilation scenarios. SPheno also integrates with other tools, such as HiggsBounds [3] and MadGraph5\_aMC@NLO [4], to evaluate the compatibility of Higgs boson observations with theoretical predictions and simulate high-energy particle collisions. This integrated approach enables a comprehensive examination of the model, as shown in Fig.1. While the method offers high accuracy, it comes at a significant computational cost. To improve efficiency, we introduce a machine learning (ML) filter at the initialization stage, which preliminarily assesses whether the randomly initialized variables are likely to be experimentally viable. By applying this filter, we can focus computational resources on promising parameter sets, thus optimizing the overall process. Although the detailed physics computations are integral to our work, they are not the primary focus of this study. For a deeper understanding of the physics calculations, we refer readers to Ferreira et al. [5]. Our main emphasis is on leveraging machine learning to enhance the efficiency of identifying experimentally viable parameter sets.

## 2. DATASET DESCRIPTION AND PREPARATION

The dataset employed in this study was derived from the comprehensive phenomenological analysis conducted by Ferreira et al. [5, 6], focusing on a Branco-Grimus-Lavoura (BGL) model. This model, a variant within the two Higgs doublet framework augmented by a single scalar singlet, stands out for its rich phenomenology in particle physics. The original dataset encompasses both experimental outcomes and the corresponding set of free parameters instrumental in computing these values.

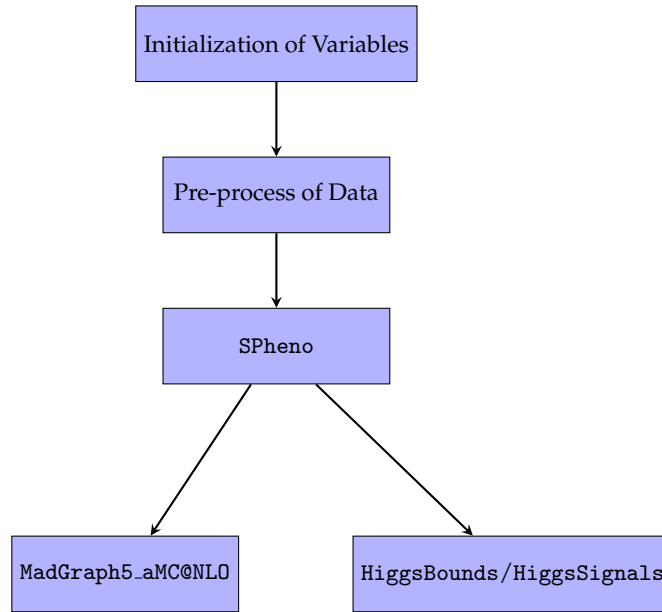


FIGURE 1: Process Flowchart

For our analysis, we conducted an extensive data-cleaning process, in which we retained only a subset of variables relevant to our work from the initial dataset. This refinement involved selecting the essential free parameters and specific physics values (denoted by  $\mathbf{P}$ ) relevant for predictive modeling. The selected points were tested based on electroweak precision observables (or oblique parameters), Higgs observables, and quark flavor violation (QFV) constraints included in  $\mathbf{P}$ , categorizing each instance as either experimentally viable or nonviable according to predefined experimental criteria. This preprocessing step was crucial to ensure that our models focused on the free parameters, thereby enhancing the accuracy and interpretability of the analysis within the framework of the BGL model. Notably, our dataset contains approximately 16% viable points, indicating an imbalance. We represent the dataset as  $\mathbf{X}$ , where  $\mathbf{X} = x_1, \dots, x_N$ , and  $N$  is the number of simulated data points. A supplementary dataset contains the physics observables, represented as  $\mathbf{P}$ , such that for each data point  $x_i$  with  $i \in 1, \dots, N$ , there is an associated set of physics observables  $\mathbf{p}_i = p_1^i, p_2^i, p_3^i$ , with  $i$  ranging from 1 to  $N$ . Further details regarding the restrictions and the physics can be found in Ferreira et al. [5].

### 3. METHODOLOGY

As part of our study, we conducted a comprehensive evaluation of various machine learning algorithms to determine their effectiveness in predicting outcomes based on our dataset. The assessment process began with the preparation of the dataset, followed by the application of standard scaling to normalize the feature space. This preprocessing step is crucial for models that are sensitive to feature magnitude. We implemented a stratified k-fold cross-validation approach with five folds for evaluating the ML algorithms. This method ensures a robust assessment while maintaining the original dataset's class proportion within each fold, thus addressing potential biases in model performance due to imbalanced class distributions. Our selection of ML models ranged from simpler algorithms, such as Gaussian Naive Bayes and Decision Trees, to more complex ensemble methods including Random Forest, AdaBoost, and XGBoost [7]. Each model was rigorously tested to gain a comprehensive understanding of its performance, focusing on metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve (ROC AUC). These metrics were selected for their capacity to provide a detailed insight into model performance, particularly in the classification "valid point" or "no-valid point." tasks where the inherent imbalance in our dataset needs to be addressed, as balancing true positive and false positive rates is essential. Creating synthetic data would introduce bias, and therefore, it was not employed in our study. We systematically recorded the evaluation results, noting not only the mean and standard deviation of each metric across the folds but also the time required for each model's training and evaluation. This method allowed us to make nuanced comparisons between the models, considering both their predictive performance and computational efficiency. The hyperparameters used for training each of these machine learning models are summarized in Table 2.

For the deep neural networks segment of our study, we adopted a focused approach to assess the training procedure, primarily utilizing the validation loss as a key indicator of performance. This method involves dividing our dataset into training, validation, and test subsets, with the latter serving to further validate our results and as a proxy to estimate the model's generalization ability on unseen data. Throughout the training epochs, we closely monitored the validation loss, aiming to minimize it as an objective measure of the model's prediction error. A distinctive emphasis was placed on exploring the capabilities of Physics-Informed

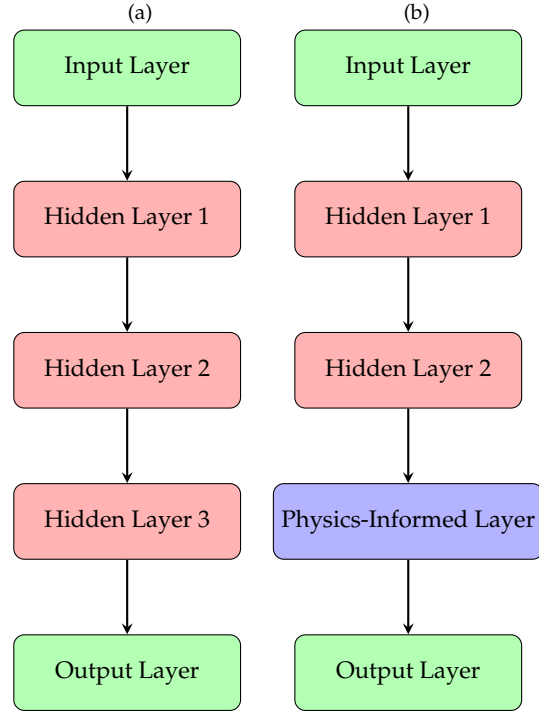


FIGURE 2: Comparison of a standard Neural Network Figure 2a and a Physics-Informed Neural Network (PINN) Figure 2b.

Neural Networks (PINNs) [8] alongside traditional Neural Networks (NNs) to address the unique challenges presented by our physics-based dataset. In subsection 3.1, we delve into the structural details of these models and their operational dynamics.

### 3.1. Physics-Informed Neural Networks (PINNs)

The cornerstone of Physics-Informed Neural Networks (PINNs) lies in their innovative integration of physical laws and information directly into the learning process. This integration is meticulously encapsulated within our dual-output model architecture. This architecture, visualized in Fig 2, encapsulates the essence of PINNs and illustrates the differences with traditional neural networks (NNs). The PINN enables the simultaneous prediction of physics-informed outputs  $\hat{\mathbf{P}}$  alongside the primary task-specific predictions  $\hat{\mathbf{Y}}$ , ensuring compliance with underlying physical principles:

$$\begin{aligned}\hat{\mathbf{P}} &= \mathcal{G}_w(\mathbf{X}) \Leftrightarrow \hat{p}_i = \mathcal{G}_w(x_i), \\ \hat{\mathbf{Y}} &= \mathcal{F}_w(\mathbf{X}, \hat{\mathbf{P}}) \Leftrightarrow \hat{y}_i = \mathcal{F}_w(x_i, \hat{p}_i).\end{aligned}\quad (1)$$

The critical innovation in our PINN approach (Fig 3) occurs during the training phase (Alg 1), particularly in the structuring of the loss function. In addition to the conventional loss term for measuring the accuracy of primary task predictions (e.g., classification accuracy or mean squared error for regression), an additional term is introduced to ensure that the predictions made by the physics-informed layer—P1, P2, P3 as shown in Fig 3—closely match the actual physics quantities derived from the data or known physical laws. To effectively balance the contribution of this physics-based loss term with the primary task loss, we include a normalization constant,  $\lambda$ . In this study,  $\lambda$  is set to one, simplifying the integration process of these terms and maintaining a direct influence of the physics-informed constraints on the overall loss function.

$$\begin{aligned}\mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{classification}} + \lambda \mathcal{L}_{\text{physics}}, \\ &= \mathcal{L}_{\text{classification}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \lambda \mathcal{L}_{\text{physics}}(\mathbf{P}, \hat{\mathbf{P}}), \\ &= \frac{1}{N} \sum_i^N \mathcal{L}_{\text{classification}}(y_i, \mathcal{F}_w(x_i, \mathcal{G}_w(x_i))) + \lambda \mathcal{L}_{\text{physics}}(\mathbf{p}_i, \mathcal{G}_w(x_i)).\end{aligned}\quad (2)$$

Here,  $\mathcal{L}_{\text{classification}}$  is given by the Binary Cross-Entropy with Logits Loss function, while  $\mathcal{L}_{\text{physics}}$  is the mean-squared-error between the outputs of the physics-informed layer and the physical variables as computed in Ferreira et al. [5]. By incorporating this additional physics-based loss term, the PINN is explicitly trained not only to achieve accuracy in its primary tasks but also to approximate physical values and make decisions informed by physics. This approach enhances both the interpretability and reliability of the model's outputs, bridging the gap between complex data analysis and fundamental physical principles.

The training process diverges from the evaluation process, by focusing exclusively on the classification loss during the evaluation phase. This strategic omission of the physics loss term during model evaluation allows for a direct and clear comparison

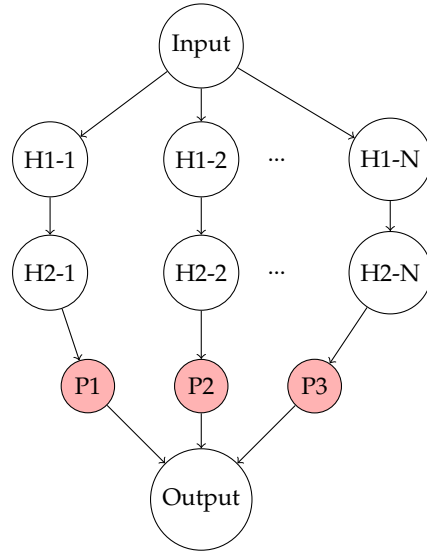


FIGURE 3: Architecture of the Physics-Informed Neural Network (PINN). The network comprises an input layer, followed by multiple fully connected hidden layers (H1 and H2), and a specialized physics-informed layer (highlighted in red), concluding with an output layer. While the illustration shows selective connections for clarity, it is important to note that each layer is fully connected to its subsequent layer. The physics-informed layer (P1, P2, P3) is designed to encode specific physical principles or constraints relevant to the problem domain, facilitating the integration of domain knowledge directly into the learning process.

between the performance metrics of standard Neural Networks (NNs) and Physics-Informed Neural Networks (PINNs), ensuring our evaluations are objective and insightful.

---

**Algorithm 1:** Training a Physics-Informed Neural Network (PINN)

---

**Data:** Training data  $\mathbf{X} = \{x_1, \dots, x_N\}$ , Training labels  $\mathbf{Y} = \{y_1, \dots, y_N\}$ , Physics data  $\mathbf{P} = \{p_1, \dots, p_N\}$

**Training:** Tuning of  $w$  based on  $\hat{y}_i = f_w(x_i, p_i)$  and  $\hat{p}_i = \mathcal{G}_w(x_i)$ ;

// Start of training step

**for all**  $x_i \in X$  with  $i \in \{1, \dots, N\}$  **do**

Forward pass to compute predictions  $\hat{y}_i$  and physics-informed outputs  $\hat{p}_i$ ;

Compute classification loss  $\mathcal{L}_{\text{classification}}(y_i, \hat{y}_i) = \mathcal{L}_{\text{classification}}(y_i, f_w(x_i, \hat{p}_i))$ ;

Compute physics loss  $\mathcal{L}_{\text{physics}}(p_i, \hat{p}_i) = \mathcal{L}_{\text{physics}}(p_i, \mathcal{G}_w(x_i))$ ;

Compute total loss  $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{classification}} + \lambda \mathcal{L}_{\text{physics}}$ ;

Backpropagation;

**end**

---

## 4. RESULTS AND DISCUSSION

In our evaluation of machine learning models for physics data analysis which can be seen in Table 1, we focused on critical metrics such as Precision, Recall, and F1 Score, which are essential for accurately identifying true positive scenarios. XGBoost excelled among the models tested, achieving the highest F1 Score ( $0.8817 \pm 0.0037$ ). This score demonstrates an optimal balance between precision and recall, pivotal for minimizing classification errors in our study. Moreover, XGBoost also led with the best ROC AUC score ( $0.9882 \pm 0.0009$ ), underscoring its exceptional ability to differentiate between classes across various threshold settings. It additionally recorded the highest precision ( $0.8990 \pm 0.0121$ ). While the Naive Bayes model displayed the highest recall, it did so at the expense of precision, highlighting the trade-offs involved in model selection. Overall, XGBoost proved to be the most effective model for our specific analytical requirements, offering high reliability, robust discrimination capabilities, and an outstanding balance of performance metrics, requiring also 9.24 sec for training.

Our analysis of the neural network models which can be seen in Table 3, examined under various learning rates, reveal small yet crucial differences in performance metrics that impact their efficacy in scenario classification. The model with a learning rate of 0.0001 outperforms others, achieving the highest F1 Score (0.9219) and Precision (0.9404), alongside an impressive Recall (0.9040), indicating its strong capability to balance between identifying true positives and minimizing false outcomes. Its ROC AUC of 0.9938 also demonstrates superior discriminatory power.

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Time (seconds)
Nearest Neighbors	0.8359 ± 0.0026	0.5160 ± 0.0091	0.4461 ± 0.0074	0.4785 ± 0.0033	0.7924 ± 0.0066	629.70
Decision Tree	0.8781 ± 0.0007	0.6851 ± 0.0039	0.5140 ± 0.0060	0.5874 ± 0.0037	0.9015 ± 0.0032	149.38
Random Forest	0.8333 ± 0.0063	0.3493 ± 0.8596	0.0130 ± 0.0401	0.0249 ± 0.0761	0.8463 ± 0.0090	27.13
AdaBoost	0.8576 ± 0.0026	0.6488 ± 0.0243	0.3409 ± 0.0270	0.4466 ± 0.0202	0.8804 ± 0.0021	580.41
Naive Bayes	0.6059 ± 0.0045	0.2953 ± 0.0024	<b>0.9636 ± 0.0042</b>	0.4521 ± 0.0029	0.8174 ± 0.0048	<b>2.01</b>
QDA	0.6183 ± 0.0058	0.3012 ± 0.0034	0.9562 ± 0.0100	0.4581 ± 0.0045	0.8173 ± 0.0034	8.01
XGBoost	<b>0.9608 ± 0.0014</b>	<b>0.8990 ± 0.0121</b>	0.8651 ± 0.0105	<b>0.8817 ± 0.0037</b>	<b>0.9882 ± 0.0009</b>	9.24

TABLE 1: Performance Metrics of Various Machine Learning Models in Physics Data Analysis. This table presents a comparative overview of the accuracy, precision, recall, F1 score, ROC AUC, and computation time for each model. Top-performing values in each category are emphasized in bold.

Model	Hyperparameters
Nearest Neighbors	n_neighbors = 3
Decision Tree	max_depth = 5
Random Forest	max_depth = 5, n_estimators = 10, max_features = 1
AdaBoost	Default hyperparameters
Naive Bayes	Default hyperparameters
QDA	Default hyperparameters
XGBoost	use_label_encoder = False, eval_metric = 'logloss'

TABLE 2: Hyperparameters used for training machine learning models of Table 1.

Learning Rate	Accuracy	Precision	Recall	F1 Score	ROC AUC
0.0001	<b>0.9740</b>	0.9404	0.9040	<b>0.9219</b>	<b>0.9938</b>
0.0002	0.9736	0.9531	0.8879	0.9193	0.9936
0.0005	0.9716	0.9217	<b>0.9099</b>	0.9158	0.9937
8e-05	0.9682	<b>0.9549</b>	0.8530	0.9011	0.9924

TABLE 3: Performance Metrics of Top Neural Network Models Under Different Learning Rates. This table showcases the results for neural network configurations that demonstrated the best overall performance in our analysis. Values in bold represent the highest scores achieved in each respective performance category across all tested models, indicating the most effective model settings for each metric."

LR	Act. Func.	Phys. Act. Func.	Accuracy	Precision	Recall	F1 Score	ROC AUC
0.0009	GELU	GELU	<b>0.9711</b>	<b>0.9145</b>	0.9152	<b>0.9148</b>	<b>0.9933</b>
0.004	GELU	GELU	0.9684	0.8853	<b>0.9351</b>	0.9095	0.9924
0.005	ReLU	LeakyReLU	0.9694	0.9178	0.9003	0.9089	0.9894
0.004	ReLU	LeakyReLU	0.9706	0.9013	0.9286	0.9147	0.9924
0.006	ReLU	ELU	0.9619	0.8546	0.9345	0.8928	0.9913
0.003	ReLU	ELU	0.9608	0.8705	0.9032	0.8866	0.9910
0.005	GELU	GELU	0.9630	0.8605	0.9331	0.8953	0.9904
0.003	GELU	GELU	0.9658	0.8746	0.9319	0.9024	0.9926
0.002	GELU	GELU	0.9675	0.8894	0.9235	0.9061	0.9899
0.006	ReLU	LeakyReLU	0.9676	0.9172	0.8894	0.9031	0.9911

TABLE 4: Performance Comparison of PINN Models with Varying Activation Functions. This table displays the accuracy, precision, recall, F1 scores, and ROC AUC for Physics-Informed Neural Networks (PINNs) across different learning rates and activation function configurations. The best values in each performance category are highlighted in bold, emphasizing the models' top performances in our evaluation.

In our study, while XGBoost offers a balanced performance across all metrics, it is outpaced by almost all the neural network models tested. Notably, the neural network with a learning rate of 0.0001 not only demonstrated superior performance with the highest F1 Score and Precision but also achieved an impressive Recall (0.9040), indicating its exceptional capability in balancing true positive identification and minimizing false outcomes. Its ROC AUC of 0.9938 further underscores its superior discriminatory power over XGBoost. Although these neural networks generally require more training time than XGBoost, their superior performance across almost all metrics makes them highly effective for applications where model accuracy is prioritized over training speed.

In our expanded analysis of Physics-Informed Neural Network (PINN) models detailed in Table 4, we assessed various configurations to gauge how learning rates and activation functions impact essential metrics like Precision, Recall, and the F1 Score. The top PINN model, with a learning rate of 0.0009 and GELU as both the activation and physics activation function, stands out by achieving the highest F1 Score (0.9148). This score reflects an optimal balance, with equally high Precision (0.9145) and Recall (0.9152), indicating effective integration of physical principles into the learning process. The superior performance of GELU can be attributed to its Gaussian Distribution-Based Smoothing, which weights inputs through a probabilistic lens, enhancing the model's ability to handle the complex, non-linear relationships inherent in physics data. Additionally, GELU supports smoother gradient flow during backpropagation compared to other activation functions. This characteristic prevents issues related to vanishing or exploding gradients, facilitating more stable and effective learning across deeper network architectures. These properties make GELU particularly suitable for PINNs, where capturing subtle physical behaviors accurately is crucial.

Comparatively, this top PINN model outperforms the best XGBoost configuration, which had an F1 Score of 0.8817, demonstrating superior balance and precision-recall trade-off. When matched against the top-performing neural network model, which reached an F1 Score of 0.9219, the leading PINN model shows competitive precision but a slightly lower overall F1 score. Despite this, the unique strength of PINNs in incorporating physical laws directly into their framework provides an explainability advantage against both XGBoost and top-performing neural network.

These results highlight the effectiveness of PINNs in optimizing both sensitivity and specificity, crucial for reducing false negatives and positives in complex systems and the ability to make decisions based on physics information without losing effectiveness. Such capabilities make PINNs particularly valuable in physics data analysis, where ensuring high accuracy in prediction under physical laws is paramount.

In summary, for the initial stages of computation, where only a small portion of the dataset is available, XGBoost is highly recommended as the machine learning filter. Its speed and relative accuracy make it ideal for quick, preliminary filter. On the other hand, for long-term simulations where comprehensive datasets are collected, Physics-Informed Neural Networks (PINNs) are advisable. PINNs leverage the underlying physical laws to enhance the predictive capability of traditional neural networks, integrating crucial physics-based insights into the otherwise non-transparent ML models. This approach ensures that the models are not only accurate but also fulfil the established physical principles, making them particularly valuable for in-depth, physics-intensive investigations.

## 5. FUTURE WORK

Looking ahead, the exploration of several intriguing research directions promises to enrich the domain of physics data analysis further. One of the most compelling prospects involves investigating the transferability of the top-performing neural network and PINN models to other physical systems with similar Lagrangians. This endeavour would not only highlight the generalizability and adaptability of these models but also deepen our comprehension of their fundamental workings. Moreover, determining the minimal dataset size necessary to achieve approximately 90% accuracy emerges as a fascinating challenge, especially in the case of XGBoost, which we recommended as a preliminary filter. This inquiry would entail a comparison between the time savings

from reduced simulation requirements and the computational demands of model training, aiming to optimize data usage efficiency. Additionally, the integration of uncertainty quantification within PINNs could significantly enhance their predictive power by offering probabilistic bounds on predictions, thereby bolstering confidence in model outputs and facilitating more informed decision-making. The adoption of ensemble methods that fuse the strengths of diverse models, including PINNs and traditional machine learning algorithms, may also unveil synergistic benefits, potentially elevating overall model performance and robustness. Lastly, applying PINNs in applications, spanning from particle physics experiments to cosmological data analysis, would furnish a comprehensive testing ground for these advanced computational tools, pushing the frontiers of current technology and knowledge in a concerted effort to unravel the complexities of our universe.

## CONFLICTS OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] Werner Porod. Spheno, a program for calculating supersymmetric spectra, susy particle decays and susy particle production at e+ e- colliders. *Computer Physics Communications*, 153(2):275–315, 2003.
- [2] W. Porod and F. Staub. Spheno 3.1: extensions including flavour, cp-phases and models beyond the mssm. *Computer Physics Communications*, 183(11):2458–2469, November 2012.
- [3] Philip Bechtle, Daniel Dercks, Sven Heinemeyer, Tobias Klingl, Tim Stefaniak, Georg Weiglein, and Jonas Wittbrodt. Higgsbounds-5: testing higgs sectors in the lhc 13 tev era. *The European Physical Journal C*, 80(12), December 2020.
- [4] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.
- [5] Pedro M Ferreira, Felipe F Freitas, João Gonçalves, Antonio P Morais, Roman Pasechnik, and Vasileios Watellis. Phenomenology of a flavored multiscalar branco-grimus-lavoura-like model with three generations of massive neutrinos. *Physical Review D*, 106(7):075017, 2022.
- [6] Pedro M Ferreira, João Gonçalves, Antonio P Morais, António Onofre, Roman Pasechnik, and Vasileios Watellis. Collider phenomenology of new neutral scalars in a flavored multi-higgs model. *Physical Review D*, 107(9):095041, 2023.
- [7] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. ACM, August 2016.
- [8] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.