



Geometric Learning in Black-Box Optimization: A GNN Framework for Algorithm Performance Prediction

Ana Kostovska
Jožef Stefan Institute
Ljubljana, Slovenia

Carola Doerr
Sorbonne Université, CNRS, LIP6
Paris, France

Sašo Džeroski
Jožef Stefan Institute
Ljubljana, Slovenia

Panče Panov
Jožef Stefan Institute
Ljubljana, Slovenia

Tome Eftimov
Jožef Stefan Institute
Ljubljana, Slovenia

Abstract

Automated algorithm performance prediction in numerical black-box optimization often relies on problem characterizations, such as exploratory landscape analysis features. These features are typically used as inputs to machine learning models and are represented in a tabular format. However, such approaches often overlook algorithm configurations, a key factor influencing performance. The relationships between algorithm operators, parameters, problem characteristics, and performance outcomes form a complex structure best represented as a graph.

This work explores the use of heterogeneous graph data structures and graph neural networks to predict the performance of optimization algorithms by capturing the complex dependencies between problems, algorithm configurations, and performance outcomes. We focus on two modular frameworks, modCMA-ES and modDE, which decompose two widely used derivative-free optimization algorithms: the covariance matrix adaptation evolution strategy (CMA-ES) and differential evolution (DE). We evaluate 324 modCMA-ES and 576 modDE variants on 24 BBOB problems across six runtime budgets and two problem dimensions.

Achieving up to 36.6% improvement in *MSE* over traditional tabular-based methods, this work highlights the potential of geometric learning in black-box optimization.

CCS Concepts

• **Computing methodologies** → **Continuous space search; Supervised learning by regression; Neural networks.**

Keywords

algorithm performance prediction, graph neural networks, numerical black-box optimization

ACM Reference Format:

Ana Kostovska, Carola Doerr, Sašo Džeroski, Panče Panov, and Tome Eftimov. 2025. Geometric Learning in Black-Box Optimization: A GNN Framework for Algorithm Performance Prediction. In *Genetic and Evolutionary Computation Conference (GECCO '25 Companion)*, July 14–18, 2025, Malaga, Spain.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
GECCO '25 Companion, July 14–18, 2025, Malaga, Spain
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1464-1/2025/07
<https://doi.org/10.1145/3712255.3726696>

Spain. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3712255.3726696>

1 Introduction

Numerical black-box optimization has long been an active area of research, resulting in the development of numerous optimization algorithms. Consequently, meta-learning tasks such as automated algorithm performance prediction [18], algorithm configuration [15], and algorithm selection [8] have become increasingly important for identifying the most suitable algorithm for a given problem.

Over the years, many novel algorithmic ideas have been introduced, along with incremental improvements to well-known derivative-free optimization algorithms such as the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [6] and Differential Evolution (DE) [17]. These advancements have further diversified the landscape of optimization algorithms, making it more challenging to systematically compare and select among them.

To address this, modular optimization frameworks such as modCMA-ES [3] and modDE [20] decompose CMA-ES and DE into modular components, enabling researchers to analyze how different configurations and parameter settings affect performance. By capturing detailed information about algorithmic structure and interactions, these frameworks offer a rich resource for algorithm analysis. However, despite their potential, they remain underutilized in meta-learning tasks.

Traditional approaches to algorithm performance prediction typically rely on tabular representations of problem features, such as Exploratory Landscape Analysis (ELA) [14]. While effective, these methods fail to capture the complex inter-dependencies among problems, algorithms, configurations, and performance outcomes. Representing these entities as nodes in a graph, with edges capturing their relationships, naturally reflects the relational structure of the data and supports richer modeling.

Geometric learning, a branch of machine learning that handles data in non-Euclidean domains like graphs and manifolds [1], provides a suitable framework for this task. Its approaches range from shallow node embeddings to more advanced Graph Neural Networks (GNNs), all leveraging relational and topological information within the data.

An early demonstration of this concept involved modeling performance prediction as a link prediction task. Specifically, in our previous work [11], we employed shallow node embeddings to infer “missing” performance links between modular optimization algorithms and black-box problems, labeling a link as solved if

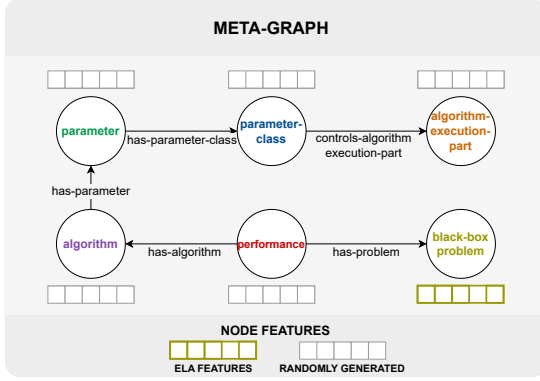


Figure 1: The meta-graph for the heterogeneous graph.

an algorithm reached a predefined precision threshold within a fixed budget of function evaluations. While effective, this approach framed performance prediction as a binary classification task rather than a more expressive regression problem. Moreover, shallow embeddings restricted the framework to transductive learning, limiting generalization to unseen problems.

GNNs address these limitations by capturing higher-order dependencies and enabling inductive learning. While not yet applied to modular optimization algorithms, GNNs have been successfully used in related domains. For example, Lukasik et al. [13] employed GNNs to predict neural architecture performance on NAS-Bench-101 [22], and Singh et al. [16] modeled deep networks as graphs to improve runtime prediction. Chai et al. [2] proposed PerfSAGE, a GNN-based model that predicts inference latency, energy usage, and memory footprint for neural networks on edge devices.

Our contributions: In this study, we apply GNNs for predicting the performance of modular optimization algorithms. First, we propose a heterogeneous graph representation that encodes data from *modCMA-ES* and *modDE*, capturing problems, algorithm components, parameters, and performance in a unified graph structure. Second, building on this representation, we introduce a message-passing GNN framework designed to operate on the heterogeneous graph, performing node regression on performance nodes to predict an algorithm’s performance on a given problem. Finally, we present experiments demonstrating that incorporating relational structure via graph-based models improves predictive performance compared to conventional tabular approaches.

Outline: The remainder of the paper is structured as follows: Section 2 details the methodology employed in our study, including the graph construction and GNN design. Section 3 outlines the experimental setup. Section 4 presents the study’s results, and Section 5 concludes with insights and directions for future work.

Code and Data: All related code and data used in this study are available at: <https://github.com/KostovskaAna/GNN4PerformancePrediction>.

2 Methodology

This section outlines our approach to predicting algorithm performance using GNNs on benchmarking data from modular optimization frameworks. We describe the heterogeneous graph representation, followed by the GNN model architecture.

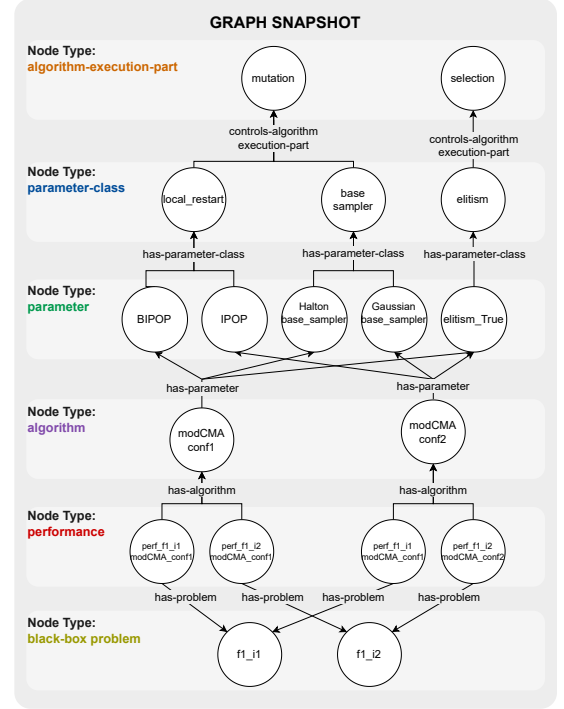


Figure 2: An illustration of an instantiation of the meta-graph, showing a snapshot of the heterogeneous graph.

2.1 Graph representation

We use a heterogeneous graph to capture the black-box optimization problems, modular optimization algorithms, their configuration, and corresponding performance scores.

A *heterogeneous graph* (HG) is defined as a graph $HG = \{\mathcal{V}, \mathcal{E}, \mathcal{R}, \mathcal{T}\}$, where \mathcal{V} represents the set of nodes, \mathcal{E} represents the set of edges, \mathcal{R} represents the set of relation types, and \mathcal{T} represents the set of node types. Each node $v \in \mathcal{V}$ is associated with a node type through a mapping function $T(v) : \mathcal{V} \rightarrow \mathcal{T}$, and each edge $e \in \mathcal{E}$ is associated with a relation type through a mapping function $R(e) : \mathcal{E} \rightarrow \mathcal{R}$. For the graph to be considered heterogeneous, the sum of distinct node and relation types must be greater than two, i.e., $|\mathcal{T}| + |\mathcal{R}| > 2$. In addition to type mappings, each node $v \in \mathcal{V}$ is associated with a feature vector $\mathbf{x}_v \in \mathbb{R}^d$, where d denotes the dimensionality of the feature space. These node features encode the properties of the nodes.

In Figure 1, we illustrate the meta-graph of the heterogeneous graph that we propose. This *meta-graph* serves as a higher-level template for our *HG*, representing \mathcal{T} (the set of node types) as nodes and \mathcal{R} (the set of relation types) as edges between them. It includes six node types: *parameter*, *parameter class*, *algorithm execution part*, *algorithm*, *performance*, and *black-box optimization problem* and five edge (relation) types: *has-parameter*, *has-parameter-class*, *controls-algorithm-execution-part*, *has-algorithm*, and *has-problem*.

In Figure 2, we illustrate a concrete instantiation of the meta-graph for a subset of \mathcal{V} and \mathcal{E} , providing a snapshot of the corresponding heterogeneous graph. It shows two black-box problems, two algorithm variants, and their associated performance nodes and parameters. The parameters belong to *local_restart*, *base_sampler*,

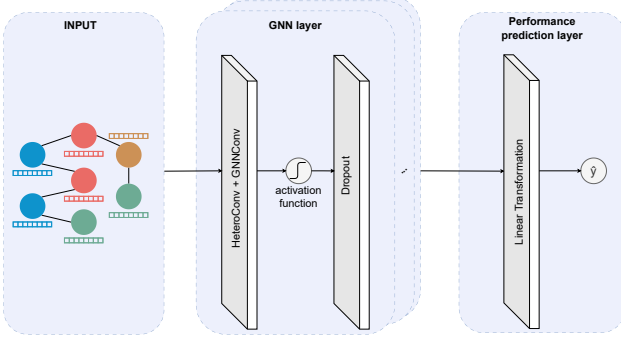


Figure 3: An overview of the message-passing GNN architecture for predicting algorithm performance using heterogeneous graphs.

and *elitism* classes, affecting different parts of algorithm execution. We construct one graph per unique combination of problem dimensionality, runtime budget, and algorithm type, resulting in multiple graph instances.

2.2 GNN architecture design

We employ a heterogeneous message-passing GNN architecture to handle multiple node and relation types in our graph. In essence, message passing in a GNN involves each node gathering “messages” from its neighbors, transforming and combining them to update its own representation. In a heterogeneous setting, different node and edge types require distinct message-passing functions or aggregation procedures, a well-established approach in the literature [25]. Consequently, each relation type r dictates how messages flow from source nodes of type t_u to destination nodes of type t_v . After aggregating messages for each relation, we consolidate the outputs across all relevant relation types to form the updated node embedding. Repeating this process over multiple layers enables nodes to capture information from increasingly distant parts of the graph.

Our architecture supports stacking multiple message-passing layers (see Figure 3). We use GraphSAGE [4] for its inductive capabilities, though other message-passing layers are also compatible. Each layer performs relation-specific aggregation followed by cross-relation aggregation, with non-linearity introduced via an activation function and dropout applied to prevent overfitting.

The learning task is node regression, where the goal is to predict the performance of an algorithm on a specific problem instance. Each performance node is associated with a numerical score, and the GNN learns to predict it by leveraging the structure and features of the heterogeneous graph. The final GNN embeddings are passed through a linear layer to predict the performance value \hat{y} .

Note that, such trained model is capable of generalizing across different algorithm variants, predicting the performance of *all modular algorithm variants* for a given runtime budget and problem dimensionality. Also, although the input graph (Figure 2) is directed, we add reverse edges, effectively treating the graph as undirected. This enables bidirectional information flow and improves representation learning.

3 Experimental setup

3.1 Data collection

3.1.1 Problem instance portfolio and landscape data. We use the 24 single-objective, noiseless, continuous black-box problems from the BBOB benchmark suite from the COCO environment [5]. Each problem includes multiple instances generated through transformations such as offsets and scalings. We select the first five instances pre-generated by COCO for each problem at dimensions $D = 5$ and $D = 30$. To represent the problem landscapes, we use 46 ELA features available through the *OPTION KB* [10].

3.1.2 Algorithm instances and configuration. We consider two modular algorithms: *CMA-ES* and *DE*. For *CMA-ES*, we use the modCMA-ES framework [3], which includes variations in sampling distributions, weighting schemes, and restart strategies. For *DE*, we employ the modDE package [20], which supports a range of mutation operators, base selection options, crossover strategies, and update mechanisms inspired by state-of-the-art DE variants. Full details on the modules and parameter spaces are provided in [11], from which this data has been reused.

In total, we analyze 324 modCMA-ES and 576 modDE variants. Each is evaluated on six function evaluation budgets, $B \in \{50D, 100D, 300D, 500D, 1000D, 1500D\}$, where $D \in \{5, 30\}$. The best precision (i.e., distance to the optimum) achieved at each budget is recorded.

All data used in this study is reused from Kostovska et al. [11] and accessed through the *OPTION KB* [10].

3.2 Model training and evaluation

We implement the GNN architecture using the Deep Graph Library (DGL) [21] to predict the performance of modular optimization algorithms. The model uses SageConv layers, based on the GraphSAGE algorithm [4], for per-relation message aggregation. These embeddings are then combined using HeteroConv layers for cross-relation aggregation. Mean aggregation is applied within relations, while summation is used across relations.

We use a 4-layer GNN to capture all relevant algorithmic information. Problem nodes are represented by 46 ELA features, while features of other node types are initialized using Kaiming uniform distribution. We use GELU [7] to introduce non-linearity.

We perform nested cross-validation to tune dropout rates (0.1, 0.2, 0.3) and embedding sizes (32, 64, 128), repeating each experiment 10 times. Models are trained for 200 epochs using L1 loss, optimized with Adam [9] (learning rate 0.1, halved every 20 stagnant epochs). We adopt the same *leave-instance-out nested cross-validation* protocol as in our previous work [12], where we train Random Forest (RF) regressors on the 46 ELA features. These serve as baselines for evaluating our GNN models.

4 Results

Building on the setup above, we conducted experiments to predict the performance of modular algorithms. Table 1 presents MSE scores across two dimensionalities and six budgets, comparing GraphSAGE models with RF baselines using only tabular ELA features from Kostovska et al. [12].

GNNs generally outperform RF across all settings. For modCMA and 5D problem dimensionality, GNN improves MSE by 0.03–0.06

for lower budgets (50D, 100D), with the margin growing for larger budgets (e.g., 5.19 → 4.57 at 1500D). At 30D problem dimensionality, gains are even more pronounced for mid to high budgets—for instance, at 100D, MSE drops from 0.27 to 0.19, and at 1000D, the relative improvement reaches 36.6%.

For modDE at 5D, GNN shows lower MSE than RF for most budgets, with the largest improvement at 300D and 500D. At 30D, GNN consistently outperforms RF. The best gain is at 100D, where MSE drops from 0.25 to 0.19, a relative improvement of 24%. At 500D, the gain is 15.8%.

Table 1: The MSE scores of the GNN and RF regression models for predicting the performance of modular algorithm variants for the BBOB problem instances in 5 and 30 dimensions.

Budget	CMA-ES 5D		CMA-ES 30D		DE 5D		DE 30D	
	GNN	RF	GNN	RF	GNN	RF	GNN	RF
50D	0.75	0.78	0.15	0.15	0.36	0.37	0.21	0.26
100D	1.16	1.22	0.19	0.27	0.39	0.43	0.19	0.25
300D	3.68	3.98	0.75	1.03	0.62	0.81	0.27	0.30
500D	4.39	4.85	0.85	1.27	1.00	1.04	0.32	0.38
1000D	4.38	5.22	1.09	1.72	2.08	1.95	0.49	0.54
1500D	4.57	5.19	1.34	1.88	2.26	2.34	0.58	0.63

5 Conclusions

In this study, we explored the use of GNNs for predicting the performance of modular optimization algorithms in numerical black-box optimization. Our results show that GNNs outperform tabular RF models by capturing relationships between problems, algorithm configurations, and performance outcomes. To our knowledge, this is the first work to apply heterogeneous graph structures and GNNs to performance prediction in this domain, offering a promising direction for enhancing meta-learning tasks such as algorithm configuration and selection.

Promising directions for future research include exploring alternative GNN architectures beyond GraphSAGE, such as graph attention networks [19] or graph transformers [24], which could further enhance predictive performance. Explainability methods like GNNExplainer [23] can provide insights into which nodes, relations, and features most influence predictions. Leveraging pre-trained GNNs through fine-tuning on our heterogeneous graphs may also improve generalization. A key challenge ahead is extending this approach to non-modular algorithms, which requires a standardized vocabulary for algorithmic components and their configurations – a task that calls for broader community collaboration.

Acknowledgments

We acknowledge the support of the Slovenian Research and Innovation Agency via program grants No. P2-0098 and P2-0103, and project grants No.J2-4460 and GC-0001. This work is also funded by the European Union under Grant Agreement 101187010 (HE ERA Chair AutoLearn-SI).

References

- [1] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.
- [2] Yuji Chai, Devashree Tripathy, Chuteng Zhou, Dibakar Gope, Igor Fedorov, Ramon Matas, David Brooks, Gu-Yeon Wei, and Paul Whatmough. 2023. Perfsage: Generalized inference performance predictor for arbitrary deep learning models on edge devices. *arXiv preprint arXiv:2301.10999* (2023).
- [3] Jacob de Nobel, Diederick Vermetten, Hao Wang, Carola Doerr, and Thomas Bäck. 2021. Tuning as a means of assessing the benefits of new ideas in interplay with existing algorithmic modules. In *Proc. of Genetic and Evolutionary Computation Conference (GECCO, Companion Material)*. ACM, 1375–1384.
- [4] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).
- [5] Nikolaus Hansen, Anne Auger, Raymond Ros, Olaf Mersmann, Tea Tušar, and Dimo Brockhoff. 2020. COCO: A platform for comparing continuous optimizers in a black-box setting. *Optimization Methods and Software* 36 (2020), 114–144. doi:10.1080/10556788.2020.1808977
- [6] Nikolaus Hansen and Andreas Ostermeier. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proc. of IEEE Congress on Evolutionary Computation*. IEEE, 312–317.
- [7] Dan Hendrycks and Kevin Gimpel. 2016. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415* (2016).
- [8] Pascal Kerschke, Holger H Hoos, Frank Neumann, and Heike Trautmann. 2019. Automated algorithm selection: Survey and perspectives. *Evolutionary computation* 27, 1 (2019), 3–45.
- [9] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [10] Ana Kostovska, Diederick Vermetten, Carola Doerr, Sašo Džeroski, Panče Panov, and Tome Eftimov. 2023. OPTION: OPTImization Algorithm Benchmarking Ontology. *IEEE Transactions on Evolutionary Computation* 27, 6 (2023), 1618–1632. doi:10.1109/TEVC.2022.3232844
- [11] Ana Kostovska, Diederick Vermetten, Sašo Džeroski, Panče Panov, Tome Eftimov, and Carola Doerr. 2023. Using Knowledge Graphs for Performance Prediction of Modular Optimization Algorithms. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*. Springer, 253–268.
- [12] Ana Kostovska, Diederick Vermetten, Peter Korošec, Sašo Džeroski, Carola Doerr, and Tome Eftimov. 2024. Using Machine Learning Methods to Assess Module Performance Contribution in Modular Optimization Frameworks. *Evolutionary computation* (2024), 1–28.
- [13] Jovita Lukasič, David Friede, Heiner Stuckenschmidt, and Margret Keuper. 2021. Neural architecture performance prediction using graph neural networks. In *Pattern Recognition: 42nd DAGM German Conference, DAGM GCPR 2020, Tübingen, Germany, September 28–October 1, 2020, Proceedings* 42. Springer, 188–201.
- [14] O. Mersmann, B. Bischl, H. Trautmann, M. Preuss, C. Weihs, and G. Rudolph. 2011. Exploratory Landscape Analysis. In *Genetic and Evolutionary Computation Conference (GECCO)*. ACM, 829–836.
- [15] Elias Schede, Jasmin Brandt, Alexander Tornede, Marcel Wever, Viktor Bengs, Eyke Hüllermeier, and Kevin Tierney. 2022. A survey of methods for automated algorithm configuration. *Journal of Artificial Intelligence Research* 75 (2022), 425–487.
- [16] Shikhar Singh, Benoit Steiner, James Hegarty, and Hugh Leather. 2021. Using graph neural networks to model the performance of deep neural networks. *arXiv preprint arXiv:2108.12489* (2021).
- [17] Rainer Storn and Kenneth Price. 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization* 11, 4 (1997), 341–359.
- [18] Risto Trajanov, Stefan Dimeski, Martin Popovski, Peter Korošec, and Tome Eftimov. 2021. Explainable landscape-aware optimization performance prediction. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 01–08.
- [19] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat* 1050, 20 (2017), 10–48550.
- [20] Diederick Vermetten, Fabio Caraffini, Anna V. Kononova, and Thomas Bäck. 2023. Modular Differential Evolution. In *Proc. of the Genetic and Evolutionary Computation Conference (GECCO '23)*. ACM, 864–872. doi:10.1145/3583131.3590417
- [21] Minjie Yu Wang. 2019. Deep graph library: Towards efficient and scalable deep learning on graphs. In *ICLR workshop on representation learning on graphs and manifolds*.
- [22] Chris Ying, Aaron Klein, Eric Christiansen, Esteban Real, Kevin Murphy, and Frank Hutter. 2019. NAS-bench-101: Towards reproducible neural architecture search. In *International conference on machine learning*. PMLR, 7105–7114.
- [23] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. GNNExplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems* 32 (2019).
- [24] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2019. Graph transformer networks. *Advances in neural information processing systems* 32 (2019).
- [25] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 793–803.