TabPFN-Assisted CMA-ES for Expensive Black-Box Optimization

Vojtech Blaha[1,*], Steven Adriaensen[2], Vladyslav Furda[3] and Martin Holena[4]

[1]Department of Theoretical Computer Science and Mathematical Logic, Charles University, Prague, Czechia
[2]University of Freiburg
[3]Faculty of Mathematics and Physics, Charles University, Prague, Czechia
[4]Institute of Computer Science, Czech Academy of Sciences, Prague, Czechia

Abstract

Expensive black-box optimization problems require optimization methods capable of achieving strong performance under limited evaluation budgets. The state-of-the-art Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is a widely used approach for such problems. Surrogate-assisted variants of CMA-ES further address this challenge by replacing a portion of expensive objective evaluations with surrogate predictions. In this work, we investigate the use of TabPFN, a pretrained transformer for tabular prediction, as a regression surrogate assisting CMA-ES. The proposed approach combines TabPFN-based surrogate modeling with a hybrid training-set selection strategy and top-fraction evolution control. The method is evaluated on the noiseless COCO BBOB benchmark suite and compared against several representative surrogate-assisted CMA-ES variants. Experimental results show that the proposed approach performs particularly well on structured multimodal functions, where it consistently achieves the best performance across all tested dimensions. On the broader multimodal function group, the method remains competitive with state-of-the-art surrogate-assisted CMA-ES variants. These findings suggest that pretrained transformer surrogates constitute a promising direction for expensive evolutionary black-box optimization.

Keywords

CMA-ES, TabPFN, black-box optimization, surrogate models, expensive optimization

1. Introduction

Many real-world optimization problems involve objective functions that are expensive to evaluate and available only as black-box procedures. Typical examples include engineering design simulations, e.g., for aerodynamic design optimization, chemical compound discovery and hyperparameter optimization [1, 2, 3, 4]. In such applications, a single objective function evaluation may require computationally expensive simulations, laboratory experiments, or physical measurements. Consequently, optimization under low evaluation budgets is of substantial practical importance.

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [5, 6] is among the most successful evolutionary algorithms for continuous black-box optimization. However, the algorithm may require a large number of objective function evaluations, which limits its applicability in expensive optimization scenarios.

To reduce the number of expensive evaluations, surrogate-assisted variants of CMA-ES employ predictive models approximating the objective function using previously evaluated candidate solutions [7]. Existing approaches differ in the employed surrogate models and evolution-control strategies, including ranking-based, local approximation, Gaussian-process-based, and ensemble-based methods [8, 9, 10, 11]. These methods commonly evaluate only a subset of candidate solutions using the true objective function while the remaining candidates receive surrogate-predicted fitness values.

ITAT'26: Information Technologies – Applications and Theory, September 25–29, 2026, Biele Karpaty, Slovakia
*Corresponding author.
vojtech.blaha600@student.cuni.cz (V. Blaha)
0009-0006-7409-3598 (V. Blaha); 0009-0008-8544-0852 (S. Adriaensen); 0009-0003-4722-1169 (V. Furda); 0000-0002-2536-9328 (M. Holena)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-09

CEUR-WS.org/Vol-4323/paper3.pdf

Recently, transformer-based prior-data fitted networks (PFNs) [12] and PFN-based foundation models for tabular data have attracted increasing attention due to their strong performance in low-data supervised learning tasks [13]. These models approximate Bayesian posterior predictive inference and use training on large collections of synthetically generated tasks [12]. In particular, TabPFN is a pretrained PFN designed for efficient tabular prediction on small datasets [14]. Due to having been pretrained on a huge amount of data, TabPFN can perform accurate inference without expensive task-specific training.

PFN-based models have also recently been explored in optimization and Bayesian optimization settings [15], where pretrained transformer surrogates demonstrated promising performance in low-data optimization regimes. However, their use as surrogate models within evolutionary black-box optimization algorithms such as CMA-ES remains unexplored.

In this work, we investigate the use of TabPFN as a regression surrogate assisting CMA-ES. In each generation, the surrogate is fitted on previously truly evaluated candidate solutions and predicts objective values for the current CMA-ES population. Only a selected fraction of the most promising candidates is evaluated using the true objective function, while the remaining candidates retain surrogate-predicted fitness values. Furthermore, we employ a hybrid training-set selection strategy combining recently evaluated samples with spatially nearest samples in order to maintain both temporal adaptivity and local relevance during optimization.

The proposed approach is evaluated on the noiseless COCO BBOB benchmark suite [16] across multiple dimensions and evaluation budgets. Experimental results indicate that the proposed method performs particularly well on structured multimodal benchmark functions under low-to-medium evaluation budgets, for which reducing the number of expensive objective evaluations is especially important.

The main contributions of this paper are as follows:

- We propose a TabPFN-based surrogate-assisted CMA-ES framework for expensive black-box optimization.
- We introduce a hybrid training-set selection strategy combining local and recent evaluation history for surrogate construction.
- We evaluate the proposed method on the COCO BBOB noiseless benchmark suite across multiple dimensions and function groups.
- We analyze the behavior of pretrained transformer-based surrogate modeling across different BBOB function groups and evaluation budgets.

2. Background and Related Work

2.1. Covariance Matrix Adaptation Evolution Strategy

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is a stochastic population-based optimization algorithm designed for continuous black-box optimization problems [5, 6, 17]. The algorithm belongs to the family of evolution strategies and is particularly effective on non-convex, non-separable, and non-differentiable optimization landscapes.

In each generation, CMA-ES samples a population of candidate solutions from a multivariate normal distribution parameterized by a mean vector $m^{(g)}$, covariance matrix $C^{(g)}$, and global step size $\sigma^{(g)}$:

$$x_k^{(g+1)} \sim m^{(g)} + \sigma^{(g)} \mathcal{N}(0, C^{(g)}),$$

where g denotes the generation index [17].

After evaluating the sampled population, candidate solutions are ranked according to their objective values and the distribution mean is updated using the best-performing individuals. CMA-ES further adapts the covariance matrix and global step size using evolution paths that accumulate information about successful search directions over multiple generations [5, 6]. These mechanisms enable the algorithm to learn correlations between decision variables and efficiently adapt the search distribution to

the local structure of the optimization landscape while preserving its characteristic invariance properties, namely invariance with respect to monotonic objective transformations and affine transformations of the search space [6].

2.2. Surrogate-Assisted CMA-ES

An important aspect of surrogate-assisted CMA-ES is evolution control, i.e., the strategy determining which candidate solutions are evaluated using the true objective function and which are assigned surrogate-predicted fitness values [7]. Consequently, surrogate-assisted CMA-ES methods are defined by two main design choices: the surrogate model and the evolution-control strategy. Existing approaches combine Ranking-SVM, local quadratic, Gaussian-process, and random-forest surrogate models with ranking-based preselection, top-k evaluation, uncertainty-aware sampling, or model-lifelength strategies [8, 9, 10, 11].

Different surrogate models provide different trade-offs between approximation quality, uncertainty estimation, computational complexity, and scalability with respect to dimensionality and training-set size [7]. Gaussian processes typically provide high-quality uncertainty estimates but may become computationally expensive for larger training sets [10]. Local quadratic models are computationally efficient but may struggle on highly multimodal landscapes [9]. Random forests can model complex nonlinear relationships but may provide less smooth approximations of the objective landscape [11].

Despite substantial progress in surrogate-assisted evolutionary optimization, most existing approaches rely on surrogate models specifically trained for the current optimization task. In contrast, recent transformer-based prior-data fitted networks enable in-context prediction using pretrained models without expensive task-specific training [12, 13]. This motivates the investigation of pretrained transformer surrogates within CMA-ES.

2.3. Prior-Data Fitted Networks and TabPFN

Prior-data fitted networks (PFNs) are transformer-based models trained to approximate Bayesian posterior predictive inference [12]. During pretraining, supervised learning tasks are repeatedly sampled from a predefined prior, together with context observations and query points. The PFN is trained to predict the query targets conditioned on the observed context, thereby amortizing inference across many synthetic tasks into the model parameters. At inference time, a new dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ and query x^* are provided directly as input, and the pretrained model approximates the predictive distribution

$$p(y^* \mid x^*, \mathcal{D})$$

without updating its parameters [12]. Thus, adaptation to a new task is performed through in-context prediction rather than task-specific gradient-based training.

TabPFN applies the PFN principle to tabular prediction [13]. It is pretrained on synthetically generated tabular tasks and jointly processes labeled context samples and unlabeled query samples to produce predictions for the latter. While the original TabPFN focused primarily on small-data classification, subsequent versions extended its capabilities and scalability.

3. Proposed Method

The proposed method combines CMA-ES with a TabPFN v3 [18], a pretrained transformer-based regression surrogate. In each generation, CMA-ES samples a population of candidate solutions from the current search distribution. The surrogate model then predicts objective values of the sampled candidates using previously truly evaluated points as contextual training data.

3.1. TabPFN-Assisted CMA-ES

Let

$$\mathscr{P}^{(g)} = \{x_i^{(g)}\}_{i=1}^{\lambda}$$

denote the population sampled by CMA-ES in generation g, where λ is the population size. Furthermore, let

$$\mathscr{H}^{(g)} = \{(x_i, y_i)\}_{i=1}^{N}$$

denote the history of previously truly evaluated candidate solutions and their objective values.

If the number of previously truly evaluated candidate solutions is smaller than a predefined threshold $N_{\min}$, all sampled candidates are evaluated using the true objective function. Otherwise, a subset

$$\mathscr{D}^{(g)} \subseteq \mathscr{H}^{(g)}, \qquad |\mathscr{D}^{(g)}| \leq N_{\max},$$

where $N_{\max}$ denotes the maximum surrogate training-set size, is selected and used as contextual training data for the TabPFN surrogate model.

The surrogate predicts objective values conditioned on that set

$$\hat{y}_i^{(g)} = f_{\mathrm{PFN}}(x_i^{(g)} \mid \mathscr{D}^{(g)})$$

for all candidate solutions in the current CMA-ES population. Before prediction, objective values in $\mathscr{D}^{(g)}$ are standardized using the mean and standard deviation of the selected training set.

Candidate solutions are subsequently ranked according to predicted objective values and only the best fraction of candidates is evaluated using the true objective function. Let $\rho \in (0, 1]$ denote the fraction of candidates selected for true evaluation. Then

$$k = \lceil \rho\lambda \rceil$$

candidate solutions with the best predicted objective values are selected for true evaluation.

The remaining candidate solutions retain surrogate-predicted fitness values. Consequently, the CMA-ES update step operates on a mixed population consisting of both truly evaluated and surrogate-evaluated candidate solutions.

Unlike conventional surrogate models, TabPFN performs prediction in a purely in-context manner without iterative task-specific optimization during runtime [13]. This makes the approach particularly suitable for low-data optimization regimes where only a limited number of truly evaluated candidate solutions is available. The complete surrogate-evaluation and evolution-control procedure for one CMA-ES generation is summarized in Algorithm 1.

3.2. Training Set Construction

The proposed method employs a bounded hybrid training-set selection strategy combining temporal recency and spatial locality. If the number of available truly evaluated candidate solutions does not exceed a predefined maximum training-set size, all available points are used for surrogate construction. Otherwise, only a subset of previously evaluated points is selected.

The selected subset consists of:

- recently evaluated candidate solutions,
- candidate solutions nearest to the current CMA-ES distribution mean in Euclidean distance.

The relative proportion of recent samples is controlled by the hyperparameter $r \in [0, 1]$. The hybrid selection is intended primarily to keep the surrogate context relevant to the current search region. Older observations that are neither recent nor spatially close to the current candidate population may

Algorithm 1 One Generation of TabPFN-Assisted CMA-ES

Require: population $\mathcal{P}^{(g)}$, archive $\mathcal{A}^{(g)}$, top fraction ρ, recency fraction r, $N_{\min}$, $N_{\max}$
if $|\mathcal{A}^{(g)}| < N_{\min}$ **then**
    evaluate all candidates using the true objective function
else
    construct training set $\mathcal{D}^{(g)}$ by selecting a fraction r of the most recently evaluated samples and filling the remaining capacity up to $N_{\max}$ with samples nearest to the current candidate population
    fit TabPFN on $\mathcal{D}^{(g)}$
    predict $\hat{y}_i^{(g)}$ for each $x_i^{(g)}$, $i = 1, \dots, \lambda$
    sort candidate indices i according to increasing $\hat{y}_i^{(g)}$
    select $k = \lceil \rho\lambda \rceil$ best candidates
    evaluate selected candidates using the true objective function
    assign predicted objective values $\hat{y}_i^{(g)}$ to all non-evaluated candidates $x_i^{(g)}$
end if
update CMA-ES using the resulting mixed fitness population
update archive $\mathcal{A}^{(g)}$

lie far from the region currently explored by CMA-ES and can therefore be less informative for local prediction.

In addition, we also investigated uncertainty-based evaluation of uncertain candidate solutions, preliminary experiments showed that uncertainty-based exploration did not improve optimization performance.

4. Experimental Setup

4.1. Benchmark and Evaluation Protocol

The proposed method was evaluated on the noiseless COCO BBOB benchmark suite [16], which consists of 24 continuous black-box optimization functions with different landscape characteristics, including separable, unimodal with low or moderate conditioning, unimodal with high conditioning, multi-modal with adequate global structure, multi-modal with weak global structure [19].

Experiments were performed on problem dimensions

$$d \in \{2, 3, 5, 10, 20, 40\}.$$

The BBOB suite distinguishes between function IDs, which define the underlying benchmark functions, and instance IDs, which correspond to different transformed realizations of the same function. The proposed method was tuned on instances 16–30 of all considered BBOB functions, while the final evaluation used the disjoint instances 1–15. Thus, tuning and testing were performed on the same function classes but on different problem instances, preventing direct overfitting to the reported test instances. Hyperparameters were selected using sequential grid searches over individual parameters on the tuning instances.

The optimization budget was chosen according to the expensive optimization setting considered in this work. In particular, the analysis primarily focuses on low-to-medium evaluation budgets corresponding approximately to 500 objective evaluations per dimension, where reducing the number of expensive true evaluations is especially important.

Performance was evaluated using empirical cumulative distribution function (ECDF) curves provided by the COCO postprocessing framework [16]. Additional analysis was performed separately for different BBOB function groups.

4.2. Compared Methods

The proposed method was compared against standard CMA-ES as well as several representative surrogate-assisted CMA-ES variants for which reference results are available in the COCO benchmarking framework. The compared methods include:

- CMA-ES [9],
- BIPOP-saACM-k [8],
- lq-CMA-ES [9],
- DTS-CMA-ES [10],
- RF-based surrogate-assisted CMA-ES [11].

These methods represent different surrogate modeling strategies, including ranking-based surrogates, local quadratic approximation, Gaussian-process-based surrogate modeling, and random-forest-based surrogate modeling.

4.3. Parameter Configuration

The resulting configuration used in all reported experiments is summarized in Table 1.

Table 1
Parameter configuration of the proposed method.

Parameter	Value
ρ	0.6
$N_{\min}$	5
$N_{\max}$	100
r	0.1

Preliminary experiments indicated that uncertainty-based evolution control did not improve optimization performance. Consequently, uncertainty-based candidate selection was not included in the final experimental configuration.

4.4. Computational Cost

The primary performance measure in this work is the number of true objective evaluations rather than wall-clock time. This choice reflects the intended application setting, in which objective evaluations correspond to expensive simulations or physical experiments that may take hours or days, whereas surrogate inference is performed computationally. In such settings, the computational overhead of TabPFN is small compared with the cost of acquiring a new true evaluation.

5. Results and Discussion

The proposed method was evaluated on the COCO BBOB noiseless benchmark suite and compared against the above-mentioned surrogate-assisted CMA-ES variants. Performance is reported using ECDF target ratios provided by the COCO postprocessing framework, which measure the fraction of optimization targets reached within a given evaluation budget.

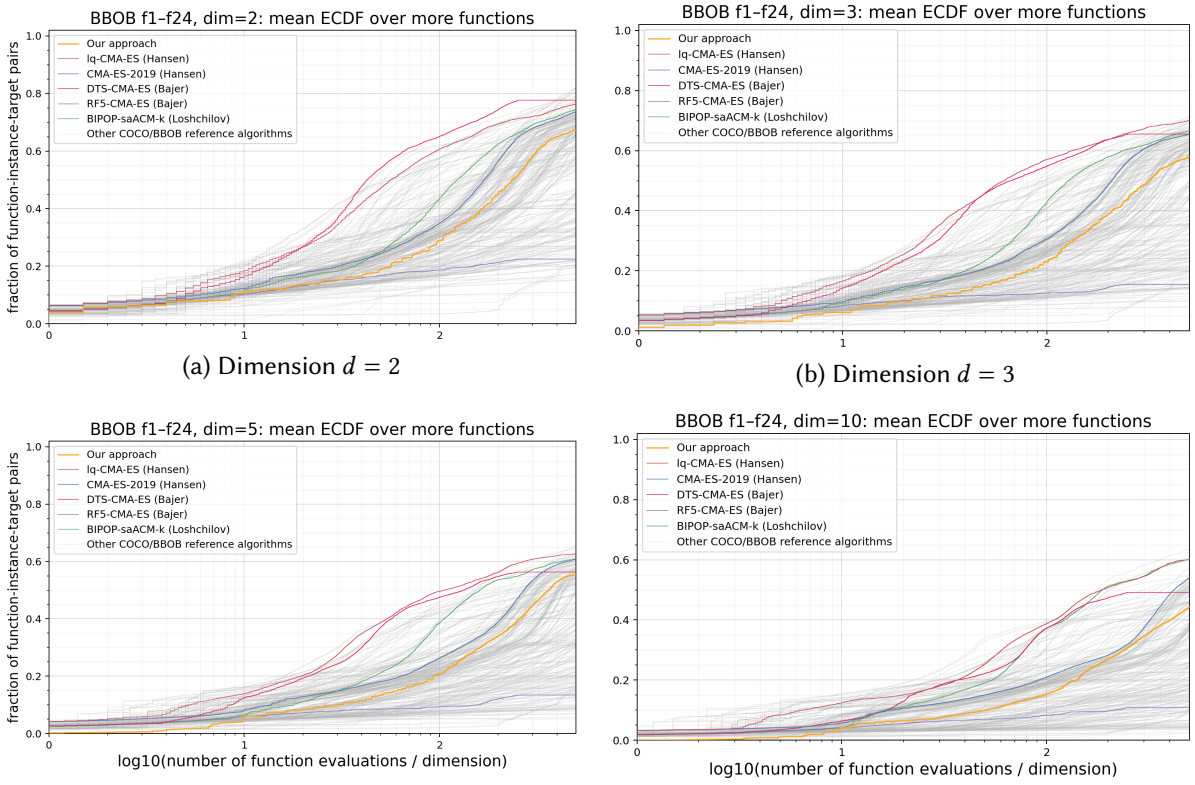
5.1. Overall Performance

When aggregated across all 24 BBOB functions, the proposed method does not outperform state-of-the-art surrogate-assisted CMA-ES variants. In particular, BIPOP-saACM-k, DTS-CMA-ES, and lq-CMA-ES achieve consistently higher ECDF target ratios across most dimensions. This suggests that a pretrained

TabPFN surrogate is not universally superior to specialized surrogate models designed specifically for black-box optimization.

Figure 1 illustrates the aggregate ECDF comparison on the complete BBOB benchmark for dimensions $d = 2$, $d = 3$, $d = 5$, and $d = 10$. Across all four dimensions, the proposed method is clearly outperformed by the strongest reference algorithms over most of the evaluation budget, confirming that its advantage does not generalize across the complete set of BBOB functions.

However, the overall benchmark aggregates function classes with substantially different landscape characteristics. To better understand the strengths of the proposed approach, we therefore analyze performance on individual BBOB function groups.



(a) Dimension $d = 2$

(b) Dimension $d = 3$

(c) Dimension $d = 5$

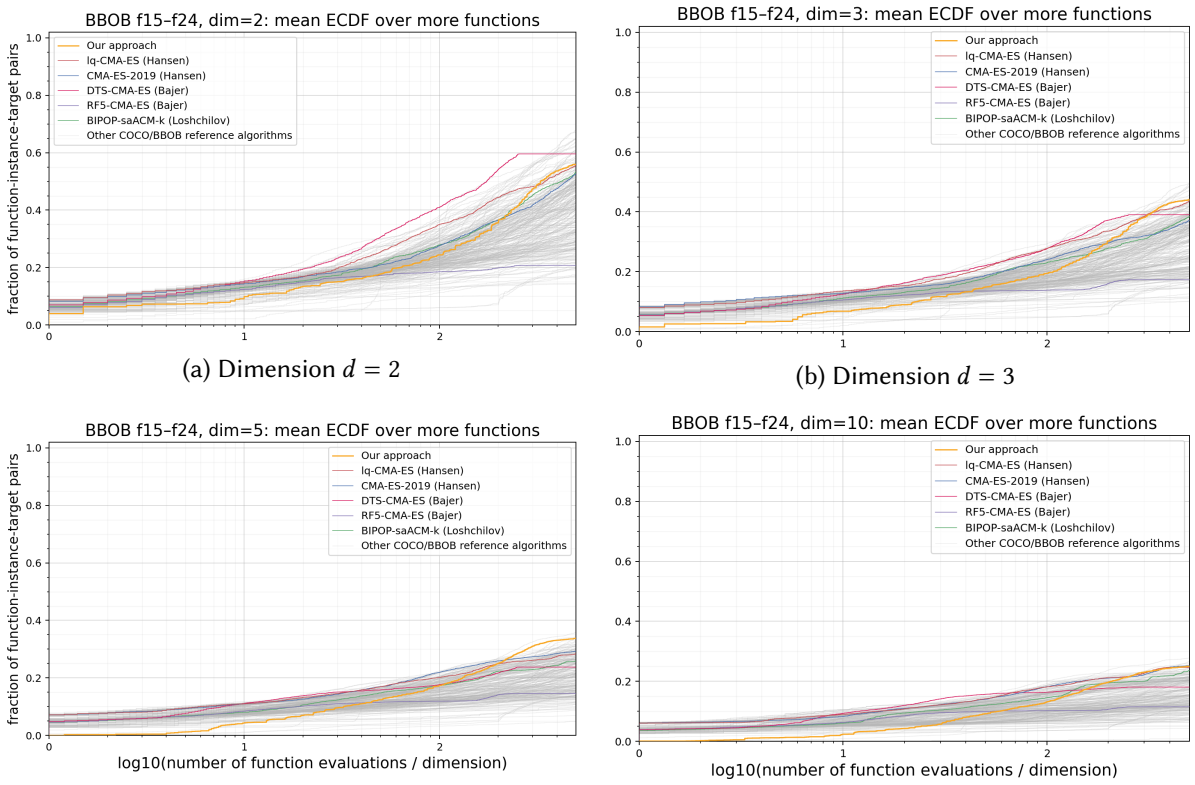
(d) Dimension $d = 10$

Figure 1: Aggregate ECDF comparison on all BBOB functions (f1–f24) for dimensions $d = 2$, $d = 3$, $d = 5$, and $d = 10$, where d denotes the problem dimension.

5.2. Multimodal Functions

On the multimodal function group (f15–f24), the proposed method achieves competitive performance, particularly in lower-dimensional settings. Figure 2 and Table 2 show that the proposed method achieves the best aggregate performance in dimensions $d = 3$ and $d = 5$ and remains competitive with the strongest baseline methods in higher dimensions. The values reported in Table 2 correspond to an evaluation budget of $500d$.

These results suggest that the TabPFN surrogate is capable of exploiting structural regularities present in multimodal landscapes, even when only a limited number of truly evaluated samples is available.



(a) Dimension $d = 2$

(b) Dimension $d = 3$

(c) Dimension $d = 5$

(d) Dimension $d = 10$

Figure 2: Aggregate ECDF comparison on multimodal functions (f15–f24) for dimensions $d = 2$, $d = 3$, $d = 5$, and $d = 10$, where d denotes the problem dimension.

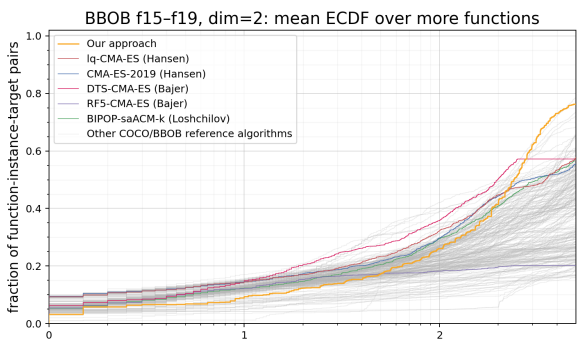
Table 2
Final ECDF target ratios on multimodal functions (f15–f24) at an evaluation budget of $500d$, where d denotes the problem dimension. Higher values are better.

Dimension	2	3	5	10	20	40
Our approach	0.559	**0.441**	**0.338**	0.250	0.202	0.161
CMA-ES	0.527	0.372	0.293	**0.252**	0.181	0.167
lq-CMA-ES	0.559	0.437	0.284	0.247	**0.211**	**0.199**
DTS-CMA-ES	**0.596**	0.390	0.238	0.181	0.154	–
BIPOP-saACM-k	0.533	0.385	0.259	0.237	0.193	0.159
RF-CMA-ES	0.208	0.173	0.146	0.115	0.097	–

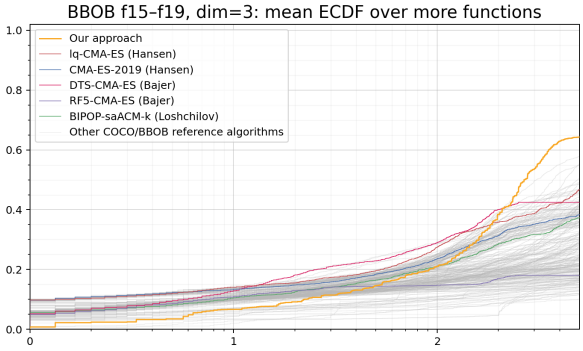
5.3. Structured Multimodal Functions

The strongest results are obtained on the BBOB function group f15–f19 (Multi-modal with adequate global structure). Figure 3 and Table 3 show that the proposed method outperforms all considered baseline methods across all tested dimensions. At an evaluation budget of $500d$, it achieves the highest ECDF target ratio in every dimension, with the largest improvements in lower-dimensional settings, while the ECDF curves further indicate faster target attainment over the low-to-medium evaluation budget range.

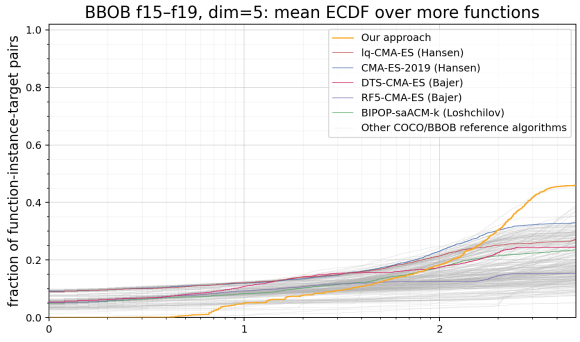
One possible explanation is that the structured multimodal landscapes of functions f15–f19 contain regularities that can be effectively captured by the pretrained TabPFN surrogate. Together with the proposed local-recent training-set selection, this may improve the identification of promising candidates in these landscapes. However, the present experiments do not isolate whether this advantage is caused by the surrogate model, the training-set selection strategy, or their interaction with the evolution-control scheme.



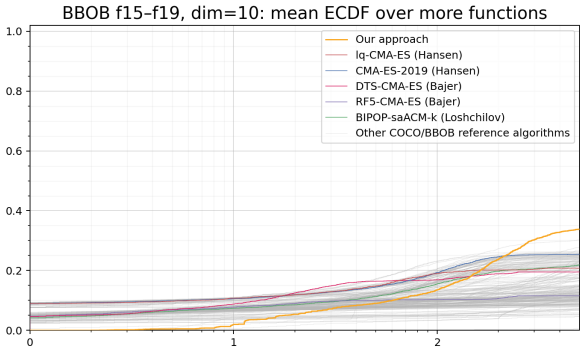
(a) Dimension $d = 2$



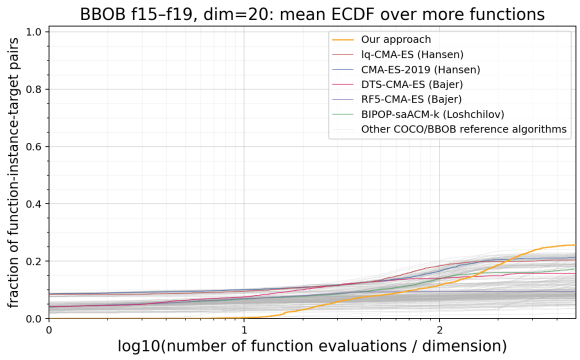
(b) Dimension $d = 3$



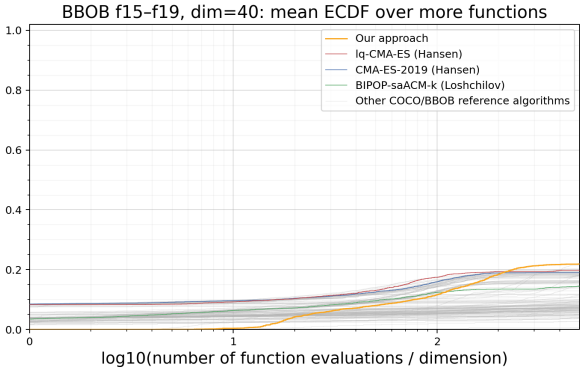
(c) Dimension $d = 5$



(d) Dimension $d = 10$



(e) Dimension $d = 20$



(f) Dimension $d = 40$

Figure 3: Aggregate ECDF comparison on multimodal functions (f15–f19) for dimensions $d = 2$, $d = 3$, $d = 5$, $d = 10$, $d = 20$, and $d = 40$, where d denotes the problem dimension.

Table 3
Final ECDF target ratios on structured multimodal functions (f15–f19) at an evaluation budget of $500d$, where d denotes the problem dimension. Higher values are better.

Dimension	2	3	5	10	20	40
Our approach	**0.762**	**0.643**	**0.460**	**0.338**	**0.257**	**0.219**
CMA-ES	0.555	0.384	0.331	0.254	0.213	0.190
lq-CMA-ES	0.578	0.468	0.273	0.208	0.204	0.198
DTS-CMA-ES	0.572	0.425	0.244	0.195	0.158	–
BIPOP-saACM-k	0.575	0.374	0.236	0.219	0.173	0.145
RF-CMA-ES	0.202	0.181	0.153	0.115	0.094	–

5.4. Discussion

The experimental results show that pretrained transformer surrogates can be successfully integrated into surrogate-assisted CMA-ES. While the proposed approach does not outperform specialized surrogate-assisted CMA-ES variants on the complete BBOB benchmark, it performs particularly well on structured multimodal functions, where it achieves the best aggregate performance across all tested dimensions.

The present experiments evaluate optimization performance but do not directly characterize the behavior of the TabPFN surrogate. An important next step is therefore to analyze prediction errors and, in particular, disagreement between predicted and true candidate rankings during optimization. Such diagnostics could help explain why the method performs strongly on some landscape classes but not on others.

Overall, the results suggest that pretrained tabular foundation models are a promising alternative to conventional surrogates in selected expensive optimization regimes. Further work is needed to improve robustness across broader classes of optimization landscapes and higher-dimensional problems.

6. Conclusion

This paper investigated the use of TabPFN as a pretrained transformer surrogate within CMA-ES for expensive black-box optimization. The proposed approach combines TabPFN-based regression, a hybrid training-set selection strategy, and top-fraction evolution control to reduce the number of expensive objective evaluations.

Experimental evaluation on the COCO BBOB noiseless benchmark suite showed that the proposed method is particularly effective on structured multimodal functions (f15–f19), where it consistently outperformed all considered baseline methods across all tested dimensions. On the broader multimodal function group (f15–f24), the method remained competitive and achieved the best aggregate performance in several dimensions. In contrast, the proposed approach did not consistently outperform state-of-the-art surrogate-assisted CMA-ES variants when results were aggregated across all benchmark functions.

The obtained results suggest that pretrained tabular foundation models can serve as effective surrogate models in evolutionary black-box optimization, particularly in low-data regimes and on landscapes with exploitable structure. Future work will investigate the internal prediction and ranking behavior of TabPFN, uncertainty-driven and adaptive evolution control, surrogate-guided generation or preselection of additional candidate points, alternative target representations, and adaptation of PFNs to families of related optimization problems.

Acknowledgments

The authors acknowledge the computational resources provided by MetaCentrum and the e-INFRA CZ infrastructure. This work was supported by the Grant Agency of Charles University, project No. 82126. Martin Holeňa's research was partially supported by the German Research Foundation (DFG) project funding no. 441926934.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5.0 in order to: Grammar and spelling correction, stylistic improvements of the manuscript, and generation of auxiliary LaTeX code for figures, tables, equations, and formatting. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] W. Kumagai, K. Yasuda, Black-Box Optimization and Its Applications, 2023, pp. 81–100. doi:10.1007/978-981-19-7776-3_6.

[2] K. Terayama, M. Sumita, R. Tamura, K. Tsuda, Black-box optimization for automated discovery, Accounts of Chemical Research (2021). doi:10.1021/acs.accounts.0c00713.

[3] M. Baerns, M. Holeňa, Combinatorial Development of Solid Catalytic Materials. Design of High-Throughput Experiments, Data Analysis, Data Mining, Imperial College Press / World Scientific, London, 2009.

[4] D. Stenger, M. Ay, D. Abel, Robust parametrization of a model predictive controller for a CNC machining center using bayesian optimization, 2020. doi:10.1016/j.ifacol.2020.12.2778.

[5] N. Hansen, A. Ostermeier, Completely derandomized self-adaptation in evolution strategies., Evol. Comput 11 (2003) 159–195.

[6] N. Hansen, The CMA evolution strategy: A comparing review, volume 192, 2007, pp. 75–102. doi:10.1007/3-540-32494-1_4.

[7] Y. Jin, Surrogate-assisted evolutionary computation: Recent advances and future challenges, Swarm and Evolutionary Computation 1 (2011) 61–70. doi:10.1016/j.swevo.2011.05.001.

[8] I. Loshchilov, M. Schoenauer, M. Sebag, BI-population CMA-ES algorithms with surrogate models and line searches, 2013. doi:10.1145/2464576.2482696.

[9] N. Hansen, A global surrogate assisted CMA-ES, 2019, pp. 664–672. doi:10.1145/3321707.3321842.

[10] Z. Pitra, L. Bajer, J. Repický, M. Holeňa, Comparison of ordinal and metric gaussian process regression as surrogate models for CMA evolution strategy, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1764–1771. URL: https://doi.org/10.1145/3067695.3084206. doi:10.1145/3067695.3084206.

[11] L. Bajer, Z. Pitra, M. Holeňa, Benchmarking gaussian processes and random forests surrogate models on the BBOB noiseless testbed, 2015. doi:10.1145/2739482.2768468.

[12] S. Müller, N. Hollmann, S. Pineda, J. Grabocka, F. Hutter, Transformers can do bayesian inference, 2021. doi:10.48550/arXiv.2112.10510.

[13] N. Hollmann, S. Müller, K. Eggensperger, F. Hutter, TabPFN: A transformer that solves small tabular classification problems in a second (2023). URL: https://arxiv.org/abs/2207.01848. arXiv:2207.01848.

[14] N. Hollmann, S. Müller, L. Purucker, et al., Accurate predictions on small data with a tabular foundation model, Nature 637 (2025) 319–326. doi:10.1038/s41586-024-08328-6.

[15] S. Müller, M. Feurer, N. Hollmann, F. Hutter, PFNs4BO: In-context learning for bayesian optimization, in: Proceedings of the 40th International Conference on Machine Learning (ICML 2023), 2023.

[16] N. Hansen, A. Auger, O. Mersmann, T. Tusar, D. Brockhoff, COCO: A platform for comparing continuous optimizers in a black-box setting, Optimization Methods and Software 36 (2016). doi:10.1080/10556788.2020.1808977.

[17] N. Hansen, The CMA evolution strategy: a tutorial (2010). doi:10.48550/arXiv.1604.00772.

[18] L. Grinsztajn, K. Flöge, O. Key, F. Birkel, et al., TabPFN-3: Technical report (2026). doi:10.48550/arXiv.2605.13986.

[19] N. Hansen, S. Finck, R. Ros, A. Auger, Real-parameter black-box optimization benchmarking 2009: Noisy functions definitions, 2009.

A. Online Resources

The source code and complete results are publicly available in the project repository:

- TabPFN-CMA-ES GitHub repository.