

CIRP Conference on Manufacturing Systems (CIRP CMS)

Foundation Models for Process Parameters Optimization: Evaluating TabPFN for Bayesian Optimization in HMLV Manufacturing

Lucas Greif ^a, Tim Berti ^{a,*}, Anne Meyer ^a^a*Institute for Information Management in Engineering, Karlsruhe Institute of Technology, Zirkel 2, Karlsruhe 76131, Germany** Corresponding author. Tel.: +49 721 608-46616 E-mail address: tim.berti@kit.edu

Abstract

High-mix, low-volume (HMLV) manufacturing relies on expensive physical experiments to optimize machine and process parameters, making sample-efficient optimization strategies essential. Bayesian optimization (BO) addresses this need by combining a probabilistic surrogate model with an acquisition function to identify promising settings under strict evaluation budgets. While Gaussian Processes (GPs) and Random Forests (RFs) are widely used as surrogates, recent tabular foundation models such as TabPFN offer a new alternative by providing fast, probabilistic predictions trained on large corpora of synthetic data. However, their suitability as BO surrogates for experimental manufacturing data remains unexplored. This work addresses the gap by benchmarking TabPFN as a surrogate model for BO on six low-dimensional, real-world experimental data sets from materials and process engineering. We emulate costly experimentation by reporting checkpoint-wise best-so-far performance up to 100 evaluations, reflecting typical evaluation budgets in HMLV settings. We compare methods using improvement-over-random (IoR), normalized area under the IoR curve, and the evaluation budget required to close 90% of the random-to-optimum gap. The results show that GP achieves the highest median IoR-AUC, whereas TabPFN attains the best average rank across data sets and the lowest median N@90, and dominates the dense rank statistics for budgets $N \geq 35$. RF and random search are consistently non-competitive. Overall, the study demonstrates that TabPFN is a viable and often advantageous surrogate for BO in HMLV-style experimental optimization, particularly in the low-budget regime.

© 2026 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Peer-review under responsibility of the scientific committee of the 2025 CIRP CMS

Keywords: Bayesian optimization; surrogate models; TabPFN; foundation models; high-mix low-volume manufacturing; process parameter optimization; design of experiments;

1. Introduction

Driven by increasing expectations for customized products, manufacturers are often operating in high-mix, low-volume (HMLV) settings characterized by many product variants, small batch sizes, and frequent design changes [7]. These demands require greater manufacturing flexibility and have catalyzed progress on two complementary fronts. In terms of production processes, digitally controlled and quickly reconfigurable techniques, like additive manufacturing [7], reduce changeover efforts and broaden the possible design space by allowing swift, mostly tool-free production. On the optimization side, data-efficient methods such as Bayesian optimization (BO) accelerate machine and process parameter optimization by modeling expensive black-box responses and identifying high-

performing settings with few experiments [15]. Within the field of BO for machine and process parameter optimization, several decisive factors dominate the procedure's efficiency: Firstly, the initial Design of Experiments (DoE) is crucial [9], as the priors significantly influence the acquisition function, typically guiding in-depth exploration in specific domains of the design space. The surrogate model then plays the central role, as it approximates the real-world process and serves as the basis for the acquisition strategy. Traditionally, Gaussian Processes (GPs) are used for this purpose, as they inherently provide an uncertainty estimate in the form of predictive variance [22]. However, increasingly, ensemble-based methods, such as Random Forests (RFs), are also gaining traction, offering advantages in handling high-dimensional or categorical data and robustness to irrelevant features, where uncertainty can be estimated via the variance of predictions across the ensemble's tree [15]. In contrast, other high-performance regression models, such as XGBoost, are not suitable for classic BO because their underly-

ing boosting mechanism does not allow for a sound estimation of the predictive uncertainty required for the acquisition strategy. Recently, a promising foundation model for tabular data, TabPFN, was introduced [11]. TabPFN utilizes a transformer-based architecture and demonstrates performance that often outperforms conventional state-of-the-art methods, particularly in low-dimensional scenarios with very limited sample sizes. A decisive advantage of TabPFN for BO is its ability to generate a full probability distribution for each predicted value. With TabPFN, there is now a model that has both state of the art performance and provides well motivated uncertainty estimates, making it an ideal candidate for use as a surrogate model. Therefore, the objective of this study is to systematically investigate the suitability of TabPFN as a surrogate model for the BO of process and machine parameters. The focus is explicitly on low-dimensional problems with strictly limited data budgets, as is typical in the HMLV context. This study examines the effectiveness of TabPFN as a viable surrogate for BO in comparison with conventional techniques such as GPs and RFs.

2. Theoretical Foundation and Related Works

2.1. Bayesian Optimization

Optimization of industrial processes often involves objective functions that are expensive, noisy, and non-convex. Evaluations may require time-consuming physical experiments, high-fidelity simulations, or costly production trials. In such settings, direct gradient-based optimization or exhaustive parameter sweeps are infeasible. BO is a widely adopted framework for optimizing such black-box functions while minimizing the number of evaluations [6]. BO starts by constructing a probabilistic surrogate model that approximates the underlying objective function. This surrogate provides both a mean prediction and an estimate of the predictive uncertainty. Then, an acquisition function, such as Expected Improvement (EI), Probability of Improvement (PI), or Upper Confidence Bound (UCB), uses these quantities to determine the next candidate point to evaluate. Each iteration therefore balances exploration, by sampling where uncertainty is high, and exploitation, by refining regions predicted to yield good performance. This iterative and data-efficient procedure is particularly well suited for industrial optimization tasks where only a limited number of experiments can be performed [9].

Applications of BO in engineering include materials discovery [15], chemical process design [16], aerodynamic shape optimization [20], machine parameter optimization [9], and hyperparameter tuning of control systems [5]. The efficiency of BO critically depends on the choice and quality of the surrogate model.

2.2. Surrogate Models

Surrogate models in BO must meet specific requirements: they must be sample-efficient, which means that they can learn from very few data points, and they must provide a well-

calibrated quantification of predictive uncertainty. The need for an uncertainty estimate significantly restricts the number of available models. The following section introduces the most commonly used surrogates.

2.2.1. Gaussian Processes

GPs are the traditional surrogate model in BO [6]. A GP defines a distribution over functions, fully characterized by a mean function and a covariance kernel. Conditioning this prior on observed data yields closed-form posterior predictive distributions, providing both a mean prediction and a principled uncertainty measure. GPs perform well for smooth, low-dimensional functions with continuous inputs. However, they suffer from several practical limitations. The computational cost of GP inference scales cubically with the number of observations, limiting their applicability when more than a few thousand samples are available. Moreover, GPs rely on stationary kernel assumptions that may poorly capture non-stationary or discontinuous industrial processes [21].

2.2.2. Random Forests

RFs offer a good alternative, particularly when data is heterogeneous or high-dimensional. They are ensembles of decision trees that approximate the target function by averaging over randomly constructed sub-models. RFs scale linearly with data size and handle categorical and mixed data types natively, making them attractive for industrial applications involving diverse sensor readings or process parameters [1]. In the context of BO, RFs have been employed in algorithms such as SMAC, which have proven competitive on many real-world problems [13]. The primary limitation of RFs is their uncertainty estimate. Typically, uncertainty is approximated by the variance of predictions across individual trees. While simple, this heuristic can produce under-calibrated uncertainties, occasionally leading to overly exploitative acquisition behavior [24]. Additionally, RFs are often beaten by Boosting models in terms of predictive accuracy [19].

2.2.3. Bayesian Neural Networks

Bayesian Neural Networks (BNNs) extend Neural Networks (NNs) by placing probability distributions over their weights to model predictive uncertainty [14]. In principle, they offer flexible, uncertainty-aware surrogates for Bayesian Optimization. In practice, however, conventional NNs perform poorly on tabular data, especially in small, noisy datasets typical of industrial processes. They require extensive tuning and large datasets [10]. The strong performance of neural networks in other domains has motivated the development of specialized architectures and training regimes for tabular data, leading to the emergence of tabular foundation models such as TabPFN.

2.3. TabPFN

TabPFN (Tabular Prior-Data Fitted Network) is a recently proposed paradigm for learning on tabular data that leverages

the concept of PFNs, neural networks trained offline to approximate Bayesian inference for a specified synthetic data-generating prior. Instead of fitting model parameters to a new dataset during inference, a PFN performs in-context learning: it receives the training samples of a new task directly as part of the input sequence and outputs posterior predictive distributions for the query points in a single forward pass. This design makes learning at inference time computationally fast and shifts modeling to the offline pretraining stage [18]. TabPFN extends the PFN framework with a Transformer-based architecture specifically tailored for small tabular datasets [12]. The model is trained on 9216000 synthetic datasets drawn from a mixture of structural causal model and BNN priors, each having a fixed number of 1024 data points. These priors encode distributions over possible causal structures, feature dependencies, nonlinear interactions, and noise processes, thereby defining a broad hypothesis space against which the PFN learns to approximate the corresponding posterior predictive distribution. After this offline prior-fitting phase, TabPFN can be used on real datasets without further gradient-based training. Given a set of observed samples, the model treats each (x,y)-pair as a token and predicts class probabilities for new feature vectors via attention over the provided training instances. Because the model directly models a full probabilistic predictive distribution rather than point estimates, it naturally yields well-structured uncertainty estimates.

2.4. Research Gap and Contribution

Recent tabular foundation models offer fast, probabilistic, retraining-free inference, yet their suitability as surrogates for low-dimensional, real-world experimental optimization remains unexplored.

PFNs4BO demonstrates that PFNs can approximate GP posteriors and act as flexible BO surrogates, but evaluations are limited to synthetic priors and hyperparameter-optimization tasks rather than physical experiments, with noted sensitivity to prior mismatch and non-accuracy objectives [18]. TabPFN itself is benchmarked mainly on classification tasks, without studying its regression uncertainty or sequential-decision behavior in BO settings [12]. GIT-BO investigates TabPFN v2 in BO, but exclusively in high-dimensional synthetic and computational engineering benchmarks, leaving its behavior unaddressed in the low-dimensional, small-sample regime [25]. As a result, it remains unknown whether TabPFN constitutes a viable surrogate for HMLV-style experimental optimization, how it compares to classical surrogates under realistic evaluation budgets, and in which budget regimes it is advantageous.

This study addresses this gap through the following contributions:

1. We provide the first systematic study of TabPFN as a surrogate model in BO for real-world, low-dimensional experiments from materials and process engineering under strict evaluation budgets.
2. We assess performance using evaluation metrics specifically designed for settings with expensive experiments.

3. We analyze surrogate model behavior over a range of evaluation budgets and assess if there are regimes in which TabPFN consistently outperforms classical surrogate models.

3. Methods

We benchmark BO under cost-aware conditions designed to emulate expensive real-world experimentation by reporting checkpoint-wise performance at iterations $t \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100\}$, where the result at each checkpoint is the best objective value observed up to t . Throughout, we restrict attention to short BO campaigns with at most 100 function evaluations per method, reflecting typical evaluation budgets in HMLV settings. All tasks were treated as minimization problems. Four surrogate models were used: GP with an isotropic kernel, GP with automatic relevance determination (ARD), a RF regressor, and a TabPFN regressor together with a Random Search (RS) baseline drawn uniformly over the design domain.

The next recommendation was obtained by maximizing EI. Let b denote the current incumbent, i.e., the best objective value observed so far. For minimization, we define the improvement at $\mathbf{x}$ as $I(\mathbf{x}) = \max\{0, b - Y(\mathbf{x})\}$. Letting $m(\mathbf{x})$ and $s(\mathbf{x})$ denote the predictive mean and standard deviation, respectively, and assuming the objective value $Y(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), s^2(\mathbf{x}))$, the standardized improvement is

$$z = \frac{b - m(\mathbf{x})}{s(\mathbf{x})}. \quad (1)$$

The acquisition function EI is then

$$\text{EI}(\mathbf{x}) = \begin{cases} (b - m(\mathbf{x})) \Phi(z) + s(\mathbf{x}) \phi(z), & s(\mathbf{x}) > 0, \\ 0, & s(\mathbf{x}) = 0, \end{cases} \quad (2)$$

with cumulative and probability density function Φ and ϕ . At each iteration the next candidate was the deterministic maximizer

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{EI}(\mathbf{x}). \quad (3)$$

For the GP surrogate, EI is evaluated in closed form under this Gaussian predictive model. For the RF surrogate, we approximate the predictive distribution by a Gaussian with mean and variance given by the empirical mean and variance of the individual tree predictions and apply the same EI formula. For the TabPFN surrogate, EI is computed numerically as the expected improvement under its full predictive distribution.

To use the pre-existing experimental data, the recommended point was determined by mapping to the nearest available sam-

ple in the corresponding data set via the Euclidean distance.

$$\mathbf{x}^{\text{real}} = \arg \min_{\mathbf{x}_i \in \mathcal{D}} \|\mathbf{x}_i - \mathbf{x}^{\star}\|_2, \quad (4)$$

The objective value associated with $\mathbf{x}^{\text{real}}$ was used for the evaluation. The same realization rule was consistently applied to all surrogates and the RS baseline. Six publicly available data sets spanning diverse experimental domains, fabrication and characterization pipelines, and optimization targets served as the test bed. Their key properties are summarized in Table 1, where d denotes the number of decision variables and size the number of unique experiments.

Table 1. Data sets and their properties.

Dataset	d	Domain	Fabrication and Characterization	Objective	Size
AgNP [17]	5	Nanoparticle synthesis	Droplet-based microfluidic synthesis; in-line hyperspectral absorption measurement	Match nanoprisms absorption spectrum	120
AutoAM [3]	4	Autonomous AM	Direct-write printing; image capture and geometry analysis	Minimize geometry error	100
Crossed Barrel [8]	4	Additive meta-material structures	FDM fabrication of lattice specimens; uniaxial compression testing	Maximize toughness	1800
P3HT [4]	5	Organic semiconductors	Spin-coated thin films; OFET characterization to extract field-effect mobility	Maximize field-effect mobility	233
Perovskite [23]	3	Hybrid halide perovskites	Spin-coating of thin films; optical monitoring under accelerated aging for stability assessment	Maximize compositional stability	139
Airfoil [2]	5	Aero-acoustics	Wind-tunnel measurements of S809 airfoil; microphone SPL across operating conditions	Minimize scaled sound pressure level	1503

For robustness, we performed $R = 10$ independent repetitions (random seeds) and report, at each checkpoint t , the mean and standard deviation of the best-so-far objective per method and dataset, where $b_{m,d}(t)$ denotes the best objective value observed up to t averaged over the R seeds.

To compare progress across datasets, we report Improvement-over-Random (IoR) relative to the empirical per-dataset optimum $f_d^{\star}$ and the RS baseline $b_{\text{RS},d}(t)$, using a small constant $\delta > 0$ added to the denominator to avoid

division by zero:

$$\text{IoR}_{m,d}(t) = 1 - \frac{b_{m,d}(t) - f_d^{\star}}{b_{\text{RS},d}(t) - f_d^{\star} + \delta}. \quad (5)$$

The efficiency of a surrogate over the budget range is summarized by the normalized area under the best-so-far IoR curve:

$$\text{AUC}_{m,d} = \frac{1}{t_T - t_1} \sum_{k=1}^{T-1} \frac{\widehat{\text{IoR}}_{m,d}(t_k) + \widehat{\text{IoR}}_{m,d}(t_{k+1})}{2} (t_{k+1} - t_k). \quad (6)$$

We also report the budget required to close a fraction ε of the RS $\rightarrow f_d^{\star}$ gap,

$$T_{m,d}(\varepsilon) = \min\{t \in \{t_1, \dots, t_T\} : \widehat{\text{IoR}}_{m,d}(t) \geq \varepsilon\}, \quad (7)$$

using $N@90 := T_{m,d}(0.90)$ and its median across datasets where finite, together with the reach rate $\#\{d : T_{m,d}(0.90) < \infty\}/|\mathcal{D}|$.

4. Results

In the supplementary material, the checkpoint-wise results across the five data sets for RF, GP (isotropic), GP (anisotropic), TabPFN and the RS baseline are reported. Fig. 1 summarizes cross-dataset performance as a function of checkpoint N by aggregating all available data sets into a single normalized score. For each data set, the mean objective of each method is first linearly rescaled so that the values are in $[0,1]$; the y-axis plots are then the average of these normalized means between the data sets at each checkpoint. From about $N \approx 10$ onward, the GP variants and TabPFN dominate with a slight edge to the GP variants. At 15 iterations and higher the difference between the GP variants and TabPFN becomes negligible.

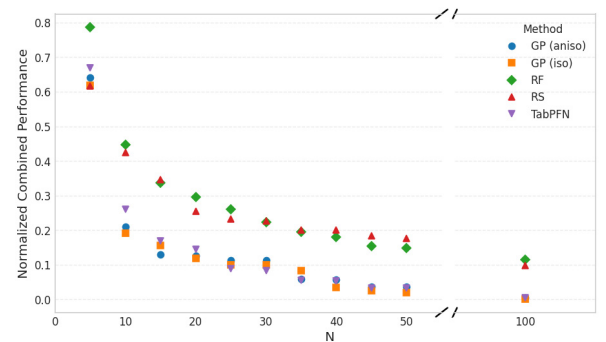


Fig. 1. Normalized combined-performance per method.

As summarized in Table 2, GP (isotropic) attains the highest aggregated efficiency with a median IoR-AUC of 0.870 and reaches $\text{IoR} \geq 0.90$ on 5/6 datasets with a median $N@90$ of 40. GP (anisotropic) and TabPFN form a closely competitive tier

(median IoR-AUC 0.757 and 0.753, respectively; both 5/6 reach rate), with TabPFN showing the best average rank (2.00) and the fastest time to high performance (median $N@90=35$). In contrast, RF trails substantially (median IoR-AUC 0.162; reach 2/6; median $N@90=20$), and the RS baseline is close to zero (median IoR-AUC 0.000; reach 1/6; median $N@90=20$). Overall, GP (isotropic) leads in area-under-curve efficiency, while TabPFN offers the best rank consistency and the fastest approach to $\text{IoR} \geq 0.90$; RF and RS are non-competitive under these aggregates.

Table 2. Aggregated efficiency metrics: Median IoR-AUC, average rank across datasets, and sample-efficiency via Median $N@90$

Method	Median IoR-AUC	Avg Rank	Median $N@90$	Reached (k/6)
GP (isotropic)	0.870	2.50	40	5/6
GP (anisotropic)	0.757	2.17	40	5/6
TabPFN	0.753	2.00	35	5/6
RF	0.162	3.83	20	2/6
RS	0.000	4.50	20	1/6

Subsequently, the dense ranking of TabPFN was assessed relative to the evaluation budget N across each of the six benchmark datasets. For every checkpoint N and dataset, TabPFN is ranked against the four baseline surrogate models (RF, GP isotropic, GP anisotropic, RS) using dense ranking.

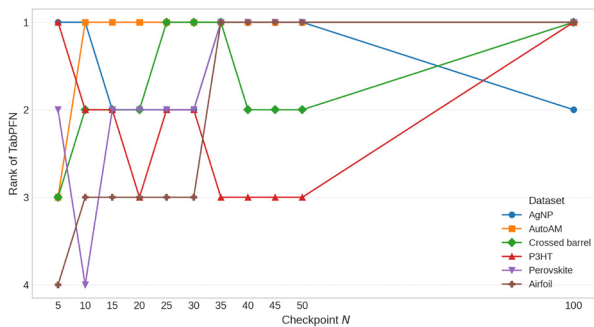


Fig. 2. Dense rank of TabPFN across checkpoints N on all six benchmark datasets.

As shown in Fig. 2, TabPFN quickly reaches rank 1 on several datasets for small budgets and, from $N \geq 35$, it is ranked first on five out of six benchmarks. At $N=100$, TabPFN remains the best or jointly the best method in all datasets except AgNP, where it is ranked second. Across checkpoints, TabPFN attains rank 1 on several datasets already at very small budgets and is mostly ranked 1 or 2 on the remaining datasets. From about $N \geq 25$, TabPFN consistently ranks 1 in Crossed barrel and Perovskite and does not drop below rank 2 in the other datasets. From $N=35$, TabPFN achieves rank 1 in five of six datasets, and this dominance is maintained up to $N=100$. At the largest budget $N=100$, TabPFN is ranked 1 on AutoAM, Crossed barrel, P3HT, Perovskite and Airfoil, and rank 2 on AgNP, indicating that TabPFN is the best or jointly best surrogate on almost all benchmarks once sufficient evaluations are available.

5. Discussion

Across all benchmarks, the results in Table 2 show that GP (isotropic) achieves the highest median IoR-AUC, indicating the most efficient overall utilization of the evaluation budget. However, TabPFN achieves the best average rank across data sets and the lowest median $N@90$, i.e., it reaches 90% of the improvement-over-random gap more quickly than any other surrogate. RF lags far behind both the GP variants and TabPFN, and the random search baseline is essentially non-competitive. These findings suggest that TabPFN is not uniformly superior in terms of area-under-curve efficiency, but is highly competitive and often faster in approaching near-optimal performance. The dense rank trajectories in Fig. 2 reveal a clear pattern between evaluation budgets. TabPFN has achieved rank 1 in several datasets for very small budgets ($N \geq 10$) and, for $N \geq 35$, is first in five of six benchmarks. This supports the view that TabPFN behaves particularly well in the low-sample regime, which is characteristic for HMLV manufacturing. However, in the largest budget considered ($N=100$), both TabPFN and GP-based surrogates frequently find the empirical minimum on a dataset. In this near-saturation regime, the methods become nearly indistinguishable and a direct comparison is no longer very informative; the resulting ties are reflected in Fig. 2 as dense rank-1 trajectories shared between models. In contrast, the higher IoR-AUC of GP (isotropic) in Table 2 indicates slightly better long-term efficiency once larger but still pre-saturation budgets are available. Practically, this suggests a trade-off: TabPFN is preferable when only a small to medium number of experiments can be afforded, whereas GP surrogates remain a strong choice for scenarios where substantially more evaluations are possible.

This study is subject to several limitations. First, the experiments are restricted to six low-dimensional, single-objective benchmarks; the conclusions may not directly transfer to high-dimensional, heavily constrained, or multi-objective settings. Second, we rely on the off-the-shelf TabPFN model without any domain-specific pretraining or fine-tuning. Given the sensitivity of BO to surrogate quality, adapting the synthetic pretraining distribution to manufacturing-specific priors could further amplify the observed benefits. Finally, only sequential single-fidelity BO was considered. Extending TabPFN-based surrogates to batch acquisition, multi-fidelity data, and robust optimization under distributional shift constitutes a promising direction for future research.

6. Conclusion and Outlook

This work has investigated whether a tabular foundation model, TabPFN, can serve as an effective surrogate for BO in HMLV process and machine parameter optimization. Using six low-dimensional real-world experimental datasets from materials and process engineering, we benchmarked TabPFN against GPs surrogates with isotropic and ARD kernels, a RF surrogate, and an RS baseline under strict evaluation budgets of up to 100 experiments. Across these benchmarks, GP (isotropic) achieved the highest median IoR-AUC, indicating slightly higher over-

all budget efficiency once a moderate number of evaluations are available. However, TabPFN achieved the best average rank in datasets and the lowest median N@90, which means that it closes 90% of the random-to-optimum gap faster than any other method. RF and RS were consistently non-competitive. The dense rank trajectories further showed that TabPFN dominates in the low to medium-budget regime (roughly $N \geq 10$ to $N \approx 35$), while, in the largest budget ($N = 100$), both TabPFN and GP surrogates frequently recover the empirical minimum, making the methods practically indistinguishable on many datasets. Taken together, these findings suggest a pragmatic design choice for HMLV-style optimization: when only a small to medium number of experiments can be afforded, TabPFN is an attractive surrogate due to its strong early performance and fast approach to near-optimal solutions. When more extensive experimentation is possible, classical GP surrogates remain a robust choice and can provide slightly better long-term efficiency in terms of area under the improvement-over-random curve. Finally, we plan to apply the proposed framework in ongoing industrial research, such as the research project AutoUni-Manu (Autonomous Universal Manufacturing), which aims to make laser-based manufacturing self-learning and autonomous through sensors, AI, digital twins, and model-based engineering. Such case studies will be crucial to validate the practical benefits and integration pathways of TabPFN-based Bayesian optimization in real HMLV manufacturing systems. Other avenues for future research include identifying how to most effectively transfer knowledge from related experiments and understanding how TabPFN behaves under various physical constraints.

Acknowledgments

This study was supported by the InnovationCampus Future Mobility (ICM), funded by the Ministry of Science, Research, and Arts of the Federal State of Baden-Württemberg.

References

- [1] Breiman, L., 2001. Random forests. *Machine Learning* 45, 532. doi:[10.1023/a:1010933404324](https://doi.org/10.1023/a:1010933404324).
- [2] Brooks, T.F., Pope, D.S., Marcolini, M.A., 2014. Airfoil self-noise. UCI Machine Learning Repository. doi:[10.24432/C5VW2C](https://doi.org/10.24432/C5VW2C).
- [3] Deneault, J.R., Chang, J., Myung, J., Hooper, D., Armstrong, A., Pitt, M., Maruyama, B., 2021. Toward autonomous additive manufacturing: Bayesian optimization on a 3d printer. *MRS Bulletin* 46, 566–575. doi:<https://doi.org/10.1557/s43577-021-00051-1>.
- [4] Dieterich, J.M., Hartke, B., 2012. Empirical review of standard benchmark functions using evolutionary global optimization. *Applied Mathematics* 3, 1552–1564. doi:<https://doi.org/10.4236/am.2012.330215>.
- [5] DrieSS, D., Englert, P., Toussaint, M., 2017. Constrained bayesian optimization of combined interaction force/task space controllers for manipulations, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 902–907. doi:[10.1109/ICRA.2017.7989111](https://doi.org/10.1109/ICRA.2017.7989111).
- [6] Frazier, P.I., 2018. A tutorial on bayesian optimization. URL: <https://arxiv.org/abs/1807.02811>, arXiv:1807.02811.
- [7] Gan, Z.L., Musa, S.N., Yap, H.J., 2023. A review of the high-mix, low-volume manufacturing industry. *Applied Sciences* 13, 1687. doi:[10.3390/app13031687](https://doi.org/10.3390/app13031687).
- [8] Gongora, A.E., Xu, B., Perry, W., Okoye, C., Riley, P., Reyes, K.G., Morgan, E.F., Brown, K.A., 2020. A bayesian experimental autonomous researcher for mechanical design. *Science advances* 6, eaaz1708. doi:<https://doi.org/10.1126/sciadv.aaz1708>.
- [9] Greif, L., Hübschle, N., Kimmig, A., Kreuzwieser, S., Martenne, A., Ovtcharova, J., 2025. Structured sampling strategies in bayesian optimization: evaluation in mathematical and real-world scenarios. *Journal of Intelligent Manufacturing*, 1–31doi:[10.1007/s10845-025-02597-2](https://doi.org/10.1007/s10845-025-02597-2).
- [10] Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data? URL: <https://arxiv.org/abs/2207.08815>, arXiv:2207.08815.
- [11] Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., Hutter, F., 2025. Accurate predictions on small data with a tabular foundation model. *Nature* 637, 319–326. doi:[10.1038/s41586-024-08328-6](https://doi.org/10.1038/s41586-024-08328-6).
- [12] Hollmann, N., Müller, S., Eggensperger, K., Hutter, F., 2023. Tabpf: A transformer that solves small tabular classification problems in a second. URL: <https://arxiv.org/abs/2207.01848>, arXiv:2207.01848.
- [13] Hutter, F., Hoos, H.H., Leyton-Brown, K., 2011. Sequential Model-Based Optimization for General Algorithm Configuration. Springer Berlin Heidelberg. p. 507523. doi:[10.1007/978-3-642-25566-3_40](https://doi.org/10.1007/978-3-642-25566-3_40).
- [14] Lampinen, J., Vehtari, A., 2001. Bayesian approach for neural network-review and case studies. *Neural Networks* 14, 257274. doi:[10.1016/s0893-6080\(00\)00098-8](https://doi.org/10.1016/s0893-6080(00)00098-8).
- [15] Liang, Q., Gongora, A., Ren, Z., Tiihonen, A., Liu, Z., Sun, S., Deneault, J., Bash, D., Mekki-Berrada, F., Khan, S., et al., 2021a. Benchmarking the performance of bayesian optimization across multiple experimental materials science domains. *npj Computational Materials* 7, 188. doi:[10.1038/s41524-021-00656-9](https://doi.org/10.1038/s41524-021-00656-9).
- [16] Liang, Q., Gongora, A.E., Ren, Z., Tiihonen, A., Liu, Z., Sun, S., Deneault, J.R., Bash, D., Mekki-Berrada, F., Khan, S.A., Hippalgaonkar, K., Maruyama, B., Brown, K.A., Fisher, III, J., Buonassisi, T., 2021b. Benchmarking the performance of bayesian optimization across multiple experimental materials science domains. *Npj Comput. Mater.* 7. doi:[10.1038/s41524-021-00656-9](https://doi.org/10.1038/s41524-021-00656-9).
- [17] Mekki-Berrada, F., Ren, Z., Huang, T., Wong, W.K., Zheng, F., Xie, J., Tian, I.P.S., Jayavelu, S., Mahfoud, Z., Bash, D., et al., 2021. Two-step machine learning enables optimized nanoparticle synthesis. *npj Computational Materials* 7, 55. doi:<https://doi.org/10.1038/s41524-021-00520-w>.
- [18] Müller, S., Feurer, M., Hollmann, N., Hutter, F., 2023. Pfns4bo: In-context learning for bayesian optimization. URL: <https://arxiv.org/abs/2305.17535>, arXiv:2305.17535.
- [19] Orzechowski, P., La Cava, W., Moore, J.H., 2018. Where are we now?: a large benchmark study of recent symbolic regression methods, in: Proceedings of the Genetic and Evolutionary Computation Conference, ACM. p. 11831190. doi:[10.1145/3205455.3205539](https://doi.org/10.1145/3205455.3205539).
- [20] Priem, R., Gagnon, H., Chittick, I., Dufresne, S., Diouane, Y., Bartoli, N., 2020. An efficient application of bayesian optimization to an industrial mdo framework for aircraft design., in: AIAA AVIATION 2020 FORUM, American Institute of Aeronautics and Astronautics. doi:[10.2514/6.2020-3152](https://doi.org/10.2514/6.2020-3152).
- [21] Remes, S., Heinonen, M., Kaski, S., 2017. Non-stationary spectral kernels. URL: <https://arxiv.org/abs/1705.08736>, arXiv:1705.08736.
- [22] Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2015. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE* 104, 148–175. doi:[10.1109/JPROC.2015.2494218](https://doi.org/10.1109/JPROC.2015.2494218).
- [23] Sun, S., Tiihonen, A., Oviedo, F., Liu, Z., Thapa, J., Zhao, Y., Hartono, N.T.P., Goyal, A., Heumueller, T., Batali, C., et al., 2021. A data fusion approach to optimize compositional stability of halide perovskites. *Matter* 4, 1305–1322. doi:<https://doi.org/10.1016/j.matt.2021.01.008>.
- [24] Wang, Z., Hutter, F., Zoghi, M., Matheson, D., de Freitas, N., 2016. Bayesian optimization in a billion dimensions via random embeddings. URL: <https://arxiv.org/abs/1301.1942>, arXiv:1301.1942.
- [25] Yu, R.T.Y., Picard, C., Ahmed, F., 2025. Git-bo: High-dimensional bayesian optimization with tabular foundation models. URL: <https://arxiv.org/abs/2505.20685>, arXiv:2505.20685.