TabPFN-2.5: A Preview

Prior Labs Team[1]

Abstract

The first tabular foundation model, TabPFN, and its successor TabPFNv2 have impacted tabular AI substantially, with dozens of methods building on it and hundreds of applications across different use cases. This paper previews TabPFN-2.5, the next generation of our tabular foundation model, built for datasets with up to 50,000 data points and 2,000 features, a $20\times$ increase in data cells compared to TabPFNv2. TabPFN-2.5 is now the leading method for the industry standard benchmark TabArena (which contains datasets with up to 100,000 training data points), substantially outperforming tuned tree-based models and matching the accuracy of AutoGluon 1.4, a complex four-hour tuned ensemble that even includes the previous TabPFNv2. Remarkably, default TabPFN-2.5 has a 100% win rate against default XGBoost on small to medium-sized classification datasets ($\leq$10,000 data points, 500 features) and a 87% win rate on larger datasets up to 100K samples and 2K features (85% for regression). For production use cases, we introduce a new distillation engine that converts TabPFN-2.5 into a compact MLP or tree ensemble, preserving most of its accuracy while delivering orders-of-magnitude lower latency and plug-and-play deployment. This new release will immediately strengthen the performance of the many applications and methods already built on the TabPFN ecosystem.

Full technical report: `https://arxiv.org/abs/2511.08667`
Try the model: `https://docs.priorlabs.ai`

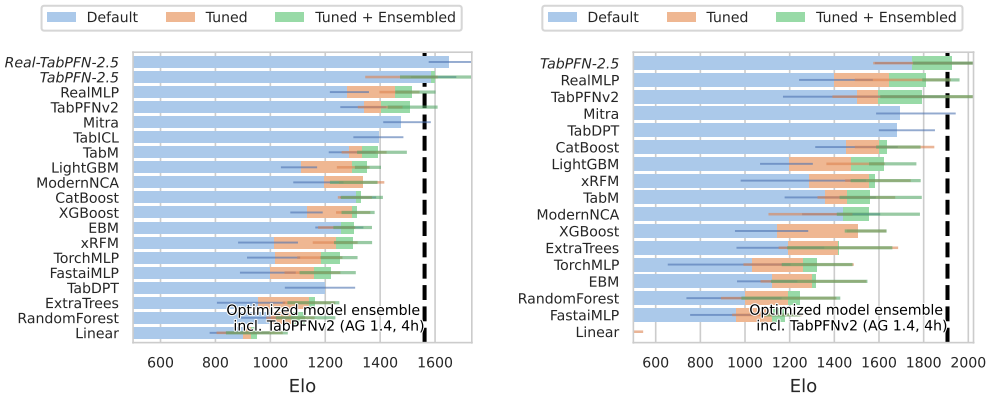


Figure 1: TabArena-Lite results on **classification** (left) and **regression** (right), restricted to datasets with less than **10K training samples and 500 features**. Note that tuning for TabPFN-2.5 is only based on 60 random configs compared to 200 for the baselines. The vertical dotted line stands for AutoGluon 1.4 extreme mode tuned for 4 hours, an ensemble of models including TabPFNv2 [39].

[1]The list of contributors can be found in Appendix A.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Tabular Data workshop at EurIPS 2025.

1 Introduction

Tabular data is ubiquitous, forming the backbone of decision-making in countless domains, from finance to healthcare. For decades, traditional tabular machine learning—built on gradient-boosted trees [26, 91, 57], random forests [1], and linear or additive models—has been the workhorse of applied data science. Yet these methods remain limited: they require extensive dataset-specific tuning, often provide uncalibrated or unreliable uncertainty estimates without significant modification, and lack the generalization and transferability of modern foundation models.

Tabular foundation models (TFMs) offer a new paradigm. They address these limitations by pretraining on large synthetic distributions of tabular tasks and performing inference via in-context learning instead of gradient descent. They are training-free predictors meta-trained to yield strong calibration, without the need for time-consuming and labor-intensive hyperparameter tuning necessary for gradient-boosted trees. Their strong generalization makes them particularly attractive for data-scarce domains.

Our initial release, TabPFNv1 [49] served as a proof-of-concept that a transformer could learn a Bayesian-like inference algorithm, though it was limited to small (up to 1k samples), clean, numerical-only data. Our successor, TabPFNv2 [50], scaled this idea into a practical model for datasets up to 10,000 samples. TabPFNv2 handles the messy and heterogeneous data seen in the real world—including categorical features, missing values, and outliers.

This paper describes the next release of TabPFN: TabPFN-2.5. Our key contributions are:

- **SOTA Performance:** In a forward pass, TabPFN-2.5 outperforms tuned tree-based models (like XGBoost and CatBoost) and matches the accuracy of AutoGluon 1.4 tuned for 4 hours—a complex ensemble that includes all previous methods, even TabPFNv2.
- **Improved Scalability:** We scale the power of in-context learning to datasets of up to 50,000 samples (5x increase over TabPFNv2) and 2,000 features (4x increase), making TFMs viable for a much wider range of real-world problems [1]. While TabPFN-2.5 was designed for up to 50,000 rows, we note that this limit is not strict and report strong results on benchmarks with up to 100,000 training samples.
- **Fast Inference:** We dramatically improve inference speed. We introduce TabPFN-as-MLP/TreeEns, a proprietary output engine, that yields an MLP or tree ensemble, combining most of TabPFN's accuracy with low-latency inference and easy deployment.

2 Model Overview

TabPFN-2.5 follows the same general design as TabPFNv2 but introduces deeper architectures, richer synthetic priors, and new calibration and inference modules. We summarize the key changes here.

Data. We improved our prior data generation substantially, broadened the set of distributions and scaled up to more data points and more features, while keeping the prediction tasks difficult. Like the original TabPFNv2, TabPFN-2.5 is trained purely on synthetically generated data. We also release a version that is fine-tuned on real data following Real-TabPFN [45]. It is trained on a curated corpus of **43 real-world tabular datasets** sourced from OpenML and Kaggle, deduplicated against all internal benchmarks and the full TabArena suite. We refer to this version as Real-TabPFN-2.5, and report strong improvement in Figures 1 and 2. See Appendix K for details on training and deduplication.

Architecture. We follow the alternating-attention transformer design of TabPFNv2, which attends across both data points and features to achieve permutation invariance, but introduces some changes:

- We increase the network depth from 12 to 18 layers for our regression model and 24 layers for our classification model.

[1] In exploratory runs, classification datasets up to $\sim$160k rows $\times$ 500 features and regression datasets up to $\sim$85k $\times$ 500 features fit into memory on an NVIDIA H100 (80 GB) using FP16 and FlashAttention-3. These configurations are outside our validated range and not included in reported benchmarks.

- We simultaneously increase the feature group size (the number of features being embedded together), which allows for faster training and inference. We use a group size of 3 for TabPFN-2.5, compared to 2 for TabPFNv2.
- For our regression models, we found a small improvement by replacing the linear encoder used in TabPFNv2 by a 2-layer MLP.
- Finally, we add 64 additional "thinking" rows to the input dataset of TabPFN-2.5, which are learned during pretraining. Inspired by results from the LLM literature [76, 46], these rows give additional computational capacity to the model and can also act as attention sinks to help the model ignore other rows [33].

Other core components from TabPFNv2—feature/sample dual attention, caching separation of training/test context, and positional feature embeddings—remain unchanged.

Preprocessing. We aggregate predictions across multiple dataset permutations and feature transformations to enhance robustness and generalization. In the updated TabPFN-2.5 configuration, additional feature transformations are introduced to enhance robustness against outlier-prone feature distributions and to increase the diversity among the individual estimators. Specifically, we combine robust scaling and soft clipping (following [51]) with quantile transformations and standard scaling to balance stability and sensitivity across features. Following TabPFNv2, we also include singular value decomposition (SVD) components as additional features in some of the estimators, capturing high-energy directions of variance that provide complementary global structure information.

Hyperparameter Tuning of TabPFN with TabPFN. TabPFN's hyperparameter space spans architectural, training, and prior-data parameters, making exhaustive grid search computationally infeasible. To explore this space efficiently, we adopted a surrogate-based optimization strategy. We first trained ≈ 100 models on a broad but sparse grid of hyperparameter configurations drawn from plausible prior ranges and evaluated them on a curated in-house validation suite, producing a compact set of hyperparameter–performance pairs. With ~ 50 hyperparameters and only 100 datapoints, direct interpolation was prone to overfitting. We therefore used a regression model well-suited for data-scarce structured prediction—our previous TabPFNv2 model—as a surrogate to predict validation performance over a denser grid of $10{,}000$ configurations. This self-referential "TabPFN-tunes-TabPFN" strategy efficiently surfaced promising regions of the search space for full, compute-intensive training runs.

Tuning custom metrics. TabPFN-2.5 adds new post-processing capabilities that enhance both calibration and metric-specific optimization. Our framework now supports tuning the classifier's decision threshold, enabling direct optimization of metrics beyond accuracy—such as the F1-score—by adjusting the operating point to the desired trade-off between precision and recall. For multiclass classification, it allows to apply temperature scaling to the final softmax outputs to improve probability calibration. This threshold tuning procedure can yield substantial performance improvements (see Appendix H). Unless otherwise noted, however, all classification results in this report are computed using uncalibrated, default scores, without temperature scaling or threshold tuning.

Reducing inference costs. Despite being a larger model than TabPFNv2, TabPFN-2.5 is between 1x and 2.3x faster thanks to optimized preprocessing and larger feature groups, as shown in Figure 19 in the appendix. We found additional speed gains in the adoption of FlashAttention-3 [102] and parallel evaluation across multiple GPUs.

Creating fast, deployable models. We also developed a proprietary distillation engine that can output a MLP or tree model that has very low latency and memory footprint for making predictions, and can be seamlessly integrated into existing production pipelines. See Appendix G.3.

3 Experimental Results

We demonstrate state-of-the-art performance on the industry standard benchmark TabArena [40]. We follow the paper's recommendation to benchmark on "TabArena-Lite", which is a cheaper but

representative version of the full benchmark using only one test fold. The benchmark contains a set of 51 datasets selected from 1053 to be representative of real-world tabular data.

Appendix F gives detailed results on TabArena-Lite, showing the pairwise win rates of the different models, and comparing TabPFN-2.5 to other foundation models like TabICL [92], TabDPT [72] or LimiX [128]. Appendices G to J report additional results using our own benchmarking framework, our advances to reduce inference latency, and new state-of-the-art performance for causal machine learning.

Pushing the limit on small to medium-sized datasets. Figure 1 shows results for TabPFN-2.5 on TabArena-Lite with up to 10,000 data points and 500 features, demonstrating that TabPFN-2.5, in a forward pass, outperforms the wide range of existing tabular prediction methods. On classification, TabPFN-2.5 in a forward pass outperforms AutoGluon 1.4, an ensemble tuned for four hours and including best other methods (even TabPFNv2). Using our Real-TabPFN-2.5 variant fine-tuned on real datasets (deduplicated from TabArena datasets) widens the lead even further. On the other hand, our regression model benefits much more from tuning and outperforms AutoGluon 1.4 after being tuned for 60 configurations.

Scaling to larger datasets. Figure 2 shows a similar experiment on all the TabArena datasets, with up to 100,000 data points and 2,000 features, clearly ranking TabPFN-2.5 as the best default model, and outperforming (for regression datasets) or approaching (for classification datasets) AutoGluon 1.4 (tuned for 4 hours) when tuned. Again, we highlight the very strong default performance of Real-TabPFN-2.5 on these larger classification datasets, beating in one forward pass any other tuned and ensembled model.

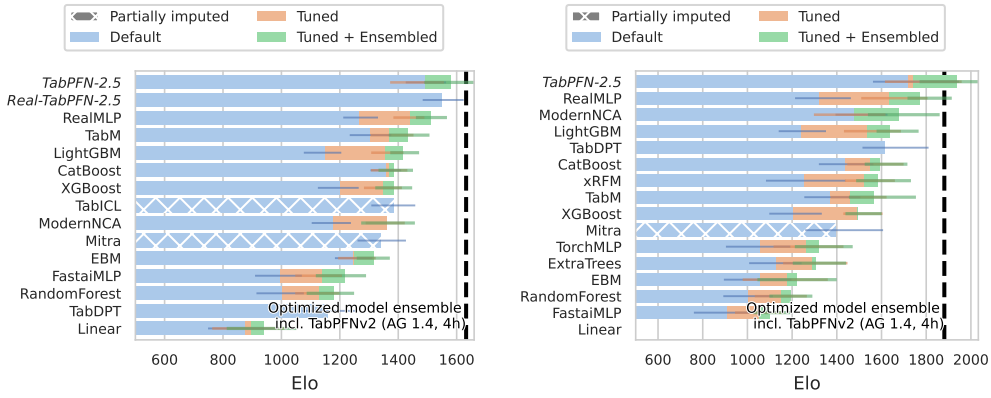


Figure 2: TabArena-Lite results on **classification** (left) and **regression** (right), evaluated on **all datasets**, going up to **100K training rows and 2K features**. Note that tuning for TabPFN-2.5 is only based on 60 random configs compared to 200 for the baselines, and that we removed the "dt-pfn" option from our tuning search space for the 4 largest datasets in the benchmark to reduce the tuning time. The vertical dotted line stands for AutoGluon 1.4 extreme mode tuned for 4 hours, an ensemble of models including TabPFNv2 [39].

4 Conclusion

We are excited about this release. Taken together, our experiments demonstrate that TabPFN-2.5 sets a new state-of-the-art for tuning-free tabular models. Built for datasets up to 50,000 rows and 2,000 features, TabPFN-2.5 matches the performance of complex 4-hour-tuned ensembles - ensembles that even include our previous TabPFNv2 - and in a forward pass outperforms any other tuned model on the unrestricted public TabArena benchmark (which contains datasets with up to 100,000 training data points).

5 Limitations

The next step is scaling to datasets with millions of rows. We are actively developing new techniques - including retrieval, fine-tuning, and novel architectures - and anticipate that systems based on Tabular Foundation Models (TFMs) will define state-of-the-art performance for datasets with millions of data points very soon.

Our broader vision beyond this release is to tackle the entire stack of problems with tabular-like data, including time series, multimodal tabular data, causal inference, unsupervised tasks, integration of domain knowledge and decision support, ultimately building the core intelligence engine for reasoning over structured and multimodal data.

References

[1] Random forests. 45(1):5–32, 2001. URL http://dx.doi.org/10.1023/A%3A1010933404324.

[2] Boosting pre-trained model with silica nanoparticles cellular toxicity prediction. *Research Square*, 2025. doi: 10.21203/rs.3.rs-7735307/v1. URL https://www.researchsquare.com/article/rs-7735307/v1. Preprint; uses TabPFN embeddings and in-context learning as the core model.

[3] Aarxshi. Rainfall_tabpfn: Post-processing rainfall forecasts with tabpfn. https://github.com/aarxshi/rainfall_tabpfn, 2024. Code repository for rainfall forecast post-processing with TabPFN.

[4] John Adeoye and Yu-Xiong Su. Artificial intelligence for predicting post-excision recurrence and malignant progression in oral potentially malignant disorders: a retrospective cohort study. *International Journal of Surgery*, 2025. doi: 10.1097/JS9.0000000000003592. Online ahead of print.

[5] Asif Adil et al. Advanced deep learning enables prediction of allogeneic stem cell mobilization success. bioRxiv preprint, 2025. URL https://www.biorxiv.org/content/10.1101/2025.09.17.676674v1.

[6] Nasser Alkhulaifi and Nicholas Bowler. Autoenergy: An automated feature engineering algorithm for energy consumption forecasting with automl. *Knowledge-Based Systems*, 2025. URL https://www.sciencedirect.com/science/article/pii/S0950705125013413. Early access; uses AutoML including TabPFN among evaluated models.

[7] Rawan AlSaad, Majid Alabdulla, Aliya Tabassum, and Rajat Thomas. From mother to infant: predicting infant temperament using maternal mental health measures and tabular machine learning models. *Frontiers in Public Health*, 13:1659987, 2025. doi: 10.3389/fpubh.2025.1659987. URL https://www.frontiersin.org/articles/10.3389/fpubh.2025.1659987.

[8] Saud A. Alzakari, Abdullah Aldrees, Muhammad Fahad Umer, Luca Cascone, Nader Innab, and Imran Ashraf. Artificial intelligence-driven predictive framework for early detection of still birth. *SLAS Technology*, 29(6):100203, 2024. doi: 10.1016/j.slast.2024.100203. URL https://www.sciencedirect.com/science/article/pii/S2472630324000852.

[9] Cinthia Fonseca Araujo, Felipe Mendes Delpino, Lílian Munhoz Figueiredo, Alexandre Dias Porto Chiavegatto Filho, Bruno Pereira Nunes, Helena Silveira Schuch, and Flavio Fernando Demarco. Predicting negative self-rated oral health in adults using machine learning: A longitudinal study in southern brazil. *Journal of Dentistry*, 163:106164, 2025. doi: 10.1016/j.jdent.2025.106164. URL https://pubmed.ncbi.nlm.nih.gov/41075925/.

[10] Davit Aslanyan. Automated supervised identification of thunderstorm ground enhancements (tges). *arXiv preprint arXiv:2510.25125*, 2025. doi: 10.48550/arXiv.2510.25125. URL https://arxiv.org/abs/2510.25125.

[11] Vahid Balazadeh, Hamidreza Kamkari, Valentin Thomas, Benson Li, Junwei Ma, Jesse C. Cresswell, and Rahul G. Krishnan. Causalpfn: Amortized causal effect estimation via in-context learning, 2025. URL https://arxiv.org/abs/2506.07918.

[12] Viacheslav Barkov, Jonas Schmidinger, Robin Gebbers, and Martin Atzmüller. Modern neural networks for small tabular datasets: The new default for field-scale digital soil mapping? *arXiv preprint arXiv:2508.09888*, 2025. URL https://arxiv.org/abs/2508.09888.

[13] Maicon Herverton Lino Ferreira da Silva Barros et al. Machine learning classification of favorable vs unfavorable tuberculosis treatment outcomes using clinical and sociodemographic data from brazil's sinan-tb (2001–2023). Research Square preprint, 2025. URL https://www.researchsquare.com/article/rs-7502054/v1.

[14] Nicolò Bellarmino, Riccardo Cantoro, Martin Huch, and Tobias Kilian. Minimal supervision, maximum accuracy: Tabpfn for microcontroller performance prediction. In *Proceedings of the International Test Conference (ITC)*, 2025. doi: 10.1109/ITC58126.2025.00067. URL https://iris.polito.it/handle/11583/3002056. Applies TabPFN for MCU performance screening with minimal supervision.

[15] Vincent Michel Borderie, Cristina Georgeon, Nassim Louissi, Benjamin Memmi, Malika Hamrani, Nacim Bouheraoua, and Anatole Chessel. CorvisST biomechanical indices in the diagnosis of corneal stromal and endothelial disorders: an artificial intelligence-based comparative study. *British Journal of Ophthalmology*, 2025. doi: 10.1136/bjo-2025-327855. URL https://pubmed.ncbi.nlm.nih.gov/41130662/. Online ahead of print.

[16] Alexej Brauer. Enhancing actuarial non-life pricing models via transformers. *European Actuarial Journal*, 14:991–1012, 2024. doi: 10.1007/s13385-024-00388-2. URL https://link.springer.com/article/10.1007/s13385-024-00388-2.

[17] Bruno-LSo. Ml-health-tabpfn. https://github.com/Bruno-LSo/ML-Health-TABPFN. GitHub repository for cardiovascular risk stratification using TabPFN. Accessed 7 Nov 2025.

[18] Sirin Cetin, Ayse Ulgen, Ozge Pasin, Hakan Sıvgın, and Meryem Cetin. Determination of malignancy risk factors using gallstone data and comparing machine learning methods to predict malignancy. *Journal of Clinical Medicine*, 14(17):6091, 2025. doi: 10.3390/jcm14176091. URL https://www.mdpi.com/2077-0383/14/17/6091.

[19] S. Chang and co authors. Cryogenic assisted abrasive waterjet machining of ti-6al-4v alloy: Thermo-mechanical optimization and ai-based surface integrity prediction. Article available via ScienceDirect, 2025. URL https://www.sciencedirect.com/science/article/abs/pii/S2214993725004531. Includes TabPFN-based modeling for surface integrity.

[20] Bowen Chen, Zhuo Xiong, Yongchun Zhao, and Junying Zhang. Multi-view machine learning model of ash chemical composition–minerals: Improving ash fusibility prediction and interpretability of high-alkali coal. SSRN preprint 5406504, 2025. URL https://ssrn.com/abstract=5406504.

[21] Gahao Chen and Ziwei Yang. Clinical prediction of intravenous immunoglobulin-resistant kawasaki disease based on interpretable transformer model. *PLOS ONE*, 20(7):e0327564, 2025. doi: 10.1371/journal.pone.0327564. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0327564.

[22] Gahao Chen and Ziwei Yang. Risk prediction for gastrointestinal bleeding in pediatric henoch–schönlein purpura using an interpretable transformer model. *Frontiers in Physiology*, 16:1630807, 2025. doi: 10.3389/fphys.2025.1630807. URL https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2025.1630807.

[23] Gang Chen, Zihan Yang, Peng Sun, Chenglong Wang, Jinliang Li, Guang Yang, and Likun Pan. Data-augmented machine learning for predicting biomass-derived hard carbon anode performance in sodium-ion batteries. *arXiv preprint arXiv:2510.12833*, 2025. URL https://arxiv.org/abs/2510.12833.

[24] Hao Chen et al. Coupling eur prediction with fracturing optimization: An integrated machine learning framework for shale gas development. Preprint / article as indexed via ScienceDirect (S2666519025001128), 2025. URL https://www.researchgate.net/publication/395761327_Coupling_EUR_Prediction_with_Fracturing_Optimization_An_Integrated_Machine_Learning_Framework_for_Shale_Gas_Development. Uses ML, including TabPFN-based models, for EUR prediction and fracturing design; update with final journal info when confirmed.

[25] Shisheng Chen, Shanshan Tong, Nyuk Hien Wong, May Lwin Oo, Joie Lim, Erna Tan, Ruohan Xu, Marcel Ignatius, Yang He, and Zhenjiang Shen. Physics-informed regression modelling for vertical facade surface temperature: A tropical case study on solar-reflective material. *arXiv preprint arXiv:2507.16174*, 2025. doi: 10.48550/arXiv.2507.16174. URL https://arxiv.org/abs/2507.16174.

[26] 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*, pages 785–794, 2016.

[27] Woruo Chen, Yao Tian, Youchao Deng, Dejun Jiang, and Dongsheng Cao. Tabpfn opens new avenues for small-data tabular learning in drug discovery. ChemRxiv preprint, 2025. URL https://chemrxiv.org/engage/chemrxiv/article-details/68d29b1cf2aff1677025b18f.

[28] Seungeon Choi, Joshep Shin, Yunu Kim, Jaemyung Shin, and Minsam Ko. Estimating sleep-stage distribution from respiratory sounds via deep audio segmentation. *Sensors*, 25(20):6282, 2025. doi: 10.3390/s25206282. URL https://www.mdpi.com/1424-8220/25/20/6282.

[29] Jasmin Ze Kee Chu, Joel Chia Ming Than, and Hudyjaya Siswoyo Jo. Deep learning for cross-selling health insurance classification. In *Proceedings of the 2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, Miri, Sarawak, Malaysia, 2024. IEEE. URL https://ieeexplore.ieee.org/abstract/document/10475046.

[30] Alicia Curth, David Svensson, Jim Weatherall, and Mihaela Van Der Schaar. Really doing great at estimating cate? a critical look at ml benchmarking practices in treatment effect estimation. In *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 2)*, 2021.

[31] Vinh Nguyen Dao et al. Early prediction of gestational diabetes using integrated cell-free dna features and omics-derived genetic scores. medRxiv preprint, 2025. URL https://www.medrxiv.org/content/10.1101/2025.09.03.25334985v1.

[32] Olanrewaju Daramola, Emmanuel Olanrewaju, Israel Trejo, and Elvis Enebeli. A target-specific machine learning framework for predicting fuel blend properties. ChemRxiv preprint, 2025. URL https://chemrxiv.org/engage/chemrxiv/article-details/68dc888d3e708a7649ff0ec9.

[33] Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *International Conference on Learning Representations (ICLR) 2024*, 2024. URL https://arxiv.org/abs/2309.16588. arXiv preprint arXiv:2309.16588v2, submitted 28 Sep 2023, revised 12 Apr 2024.

[34] Anish Dhir, Cristiana Diaconu, Valentinian Mihai Lungu, James Requeima, Richard E. Turner, and Mark van der Wilk. Estimating interventional distributions with uncertain causal graphs through meta-learning, 2025. URL https://arxiv.org/abs/2507.05526.

[35] Yilang Ding, Jiawen Ren, Jiaying Lu, Gloria Hyunjung Kwak, Armin Iraji, and Alex Fedorov. Longitudinal progression prediction of alzheimer's disease with tabular foundation model. *arXiv preprint arXiv:2508.17649*, 2025. URL https://arxiv.org/abs/2508.17649.

[36] Okan Düzyel, Mehmet Kuntalp, Fevzi Yasin Karabulut, and Damla Kuntalp. Tabpfn achieves superior performance in respiratory disease classification based on respiratory sound data. SSRN preprint, 2025. URL https://ssrn.com/abstract=5529540.

[37] Daniiar Dyikanov, Aleksandr Zaitsev, Tatiana Vasileva, Iris Wang, Arseniy A. Sokolov, Evgenii S. Bolshakov, and et al. Comprehensive peripheral blood immunoprofiling reveals five immunotypes with immunotherapy response characteristics in patients with cancer. *Cancer Cell*, 42(5):759–779.e12, 2024. doi: 10.1016/j.ccell.2024.04.008. URL https://www.cell.com/cancer-cell/fulltext/S1535-6108(24)00132-6.

[38] Moumen El-Melegy, Ahmed Mamdouh, Samia Ali, Mohamed Badawy, Mohamed A. El-Ghar, Norah S. Alghamdi, and Ayman El-Baz. Prostate cancer diagnosis via visual representation of tabular data and deep transfer learning. *Bioengineering*, 11(7):635, 2024. doi: 10.3390/bioengineering11070635. URL https://www.mdpi.com/2306-5354/11/7/635.

[39] Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, 2020.

[40] Nick Erickson, Lennart Purucker, Andrej Tschalzev, David Holzmüller, Prateek Mutalik Desai, Frank Hutter, et al. Tabarena: A living benchmark for machine learning on tabular data. *arXiv preprint arXiv:2506.16791*, 2025.

[41] Jacob Feitelberg, Dwaipayan Saha, Kyuseong Choi, Zaid Ahmad, Anish Agarwal, and Raaz Dwivedi. Tabimpute: Accurate and fast zero-shot missing-data imputation with a pre-trained transformer. *arXiv preprint arXiv:2510.02625*, 2025. doi: 10.48550/arXiv.2510.02625. URL https://arxiv.org/abs/2510.02625.

[42] Open Climate Fix. Adjuster this! tabpfn for solar forecast error adjustment. https://gist.github.com/anshulg954/5f4423ee6b3d3151fa8d0d7fcd98d3eb, 2025. Prototype from Open Climate Fix Summer of Code project for TabPFN-based solar forecast error adjustment.

[43] Rodrigo J. Galindo et al. Development of machine learning models to predict hypoglycemia and hyperglycemia on days of hemodialysis in patients with diabetes based on continuous glucose monitoring. *medRxiv*, 2025. doi: 10.1101/2025.10.24.25338707. URL https://www.medrxiv.org/content/10.1101/2025.10.24.25338707v1.

[44] Pablo García, Jordi de Curtò, Ignacio de Zarzà, Juan Carlos Cano, and Carlos T. Calafate. Foundation models for cybersecurity: A comprehensive multi-modal evaluation of tabpfn and tabicl for tabular intrusion detection. *Electronics*, 14(19):3792, 2025. doi: 10.3390/electronics14193792. URL https://www.mdpi.com/2079-9292/14/19/3792.

[45] Anurag Garg, Muhammad Ali, Noah Hollmann, Lennart Purucker, Samuel Müller, and Frank Hutter. Real-tabpfn: Improving tabular foundation models via continued pre-training with real-world data. *arXiv preprint arXiv:2507.03971*, 2025.

[46] Sachin Goyal, Ziwei Ji, Ankit Singh Rawat, Aditya Krishna Menon, Sanjiv Kumar, and Vaishnavh Nagarajan. Think before you speak: Training language models with pause tokens. In *International Conference on Learning Representations (ICLR) 2024 Poster*, 2024. URL https://openreview.net/forum?id=ph04CRkPdC. Poster paper; published 16 Jan 2024, last modified 17 Mar 2024.

[47] Ping He, Zhanlin Cao, Honggui Di, Guangxin Shen, and Shunhua Zhou. Application of machine learning in caisson inclination prediction: Model performance comparison and interpretability analysis. *Underground Space*, 2025. URL https://www.sciencedirect.com/science/article/abs/pii/S2214391225001734. Includes TabPFN-based models among compared approaches.

[48] Carola Sophia Heinzel, Lennart Purucker, Frank Hutter, and Peter Pfaffelhuber. Advancing biogeographical ancestry predictions through machine learning. In *Forensic Science International: Genetics*. Elsevier, 2025. doi: 10.1016/j.fsigen.2025.103290.

[49] Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. Tabpfn: A transformer that solves small tabular classification problems in a second. *arXiv preprint arXiv:2207.01848*, 2022.

[50] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):319–326, 2025. ISSN 1476-4687. doi: 10.1038/s41586-024-08328-6. URL https://doi.org/10.1038/s41586-024-08328-6.

[51] David Holzmüller, Léo Grinsztajn, and Ingo Steinwart. Better by default: Strong pre-tuned mlps and boosted trees on tabular data. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/2ee1c87245956e3eaa71aaba5f5753eb-Abstract-Conference.html.

[52] Sunil Kumar Jha, James Brinkhoff, Andrew J. Robson, and Brian W. Dunn. Integrating remote sensing and weather time series for australian irrigated rice phenology prediction. *Remote Sensing*, 17(17):3050, 2025. doi: 10.3390/rs17173050. URL https://www.mdpi.com/2072-4292/17/17/3050.

[53] Yongyong Jia, Xiaohui Gao, Zhihui Cai, Yafeng Ji, and Qiwei He. The multimodal fusion framework reveals the mapping relationship between microstructure and friction behavior. SSRN preprint 5616984, 2025. URL https://ssrn.com/abstract=5616984. Integrates image features with a TabPFN-based module for wear prediction.

[54] Mert Karabacak, Alexander Schupper, Matthew Carr, and Konstantinos Margetis. A machine learning-based approach for individualized prediction of short-term outcomes after anterior cervical corpectomy. *Asian Spine Journal*, 18(4):541–549, 2024. doi: 10.31616/asj.2024.0048. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11366553/.

[55] Mert Karabacak, Burak Berksu Ozkara, Tobias D. Faizy, Trevor Hardigan, Jeremy J. Heit, Dheeraj A. Lakhani, Konstantinos Margetis, Kambiz Nael, Max Wintermark, and V. Sreenivasan Yedavalli. Data-driven prognostication in distal medium vessel occlusions using explainable machine learning. *American Journal of Neuroradiology*, 46(4):725–732, 2025. doi: 10.3174/ajnr.A8547. URL https://www.ajnr.org/content/46/4/725.

[56] Rickard Karlsson and Jesse Krijthe. Detecting hidden confounding in observational data using multiple environments. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 44280–44309. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/89e541b817ea043a971840a926e12b37-Paper-Conference.pdf.

[57] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3146–3154. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

[58] Sadegh Khanmohammadi, Miguel G. Cruz, Daniel D. B. Perrakis, Martin E. Alexander, and Mehrdad Arashpour. Using automl and generative ai to predict the type of wildfire propagation in canadian conifer forests. *Ecological Informatics*, 82:102711, 2024. doi: 10.1016/j.ecoinf.2024.102711. URL https://www.sciencedirect.com/science/article/pii/S157495412400253X.

[59] Thomas Derya Kocar, Simone Brefka, Christoph Leinert, Utz Lovis Rieger, Hans Kestler, Dhayana Dallmeier, Jochen Klenk, and Michael Denkinger. Deep learning predicts postoperative mobility, activities of daily living, and discharge destination in older adults from sensor data. *Sensors*, 25(16):5021, 2025. doi: 10.3390/s25165021. URL https://www.mdpi.com/1424-8220/25/16/5021.

[60] Christopher Kolberg, Katharina Eggensperger, and Nico Pfeifer. Tabpfn-wide: Continued pre-training for extreme feature counts. *arXiv preprint arXiv:2510.06162*, 2025. doi: 10.48550/arXiv.2510.06162. URL https://arxiv.org/abs/2510.06162.

[61] Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

[62] Ellen L. Larson et al. Machine learning models of rna expression landscapes help predict overall tumor response to chemotherapy in cholangiocarcinoma. In *AACR Special Conference in Cancer Research: Artificial Intelligence and Machine Learning*, volume 31, page A020, 2025. URL https://aacrjournals.org/clincancerres/article/31/13_Supplement/A020/763312. Abstract A020.

[63] Jie Li, Andrew McCarthy, Zhizhuo Zhang, and Stephen Young. Uncertainty-guided model selection for tabular foundation models in biomolecule efficacy prediction. *arXiv preprint arXiv:2510.02476*, 2025. URL https://arxiv.org/abs/2510.02476.

[64] Yunhua Li et al. Mri delta-radiomics and morphological feature-driven tabpfn model for preoperative prediction of lymphovascular invasion in invasive breast cancer. *Technology in Cancer Research & Treatment*, 24:15330338251362050, 2025. doi: 10.1177/15330338251362050. URL https://journals.sagepub.com/doi/10.1177/15330338251362050.

[65] Zongzheng Li, Chunru Xiong, Kai Zheng, and Qiang Li. An rf-tabpfn-based framework for few-shot iot network attack recognition using lasso-rfe feature selection. *IEEE Access*, 13:151452–151465, 2025. URL https://ieeexplore.ieee.org/document/11142329. Combines Random Forest and TabPFN; DOI to be taken from the IEEE record.

[66] Xiaohui Lin, Yujia Wang, Lingling Zhang, and et al. Construction of machine learning classification prediction model for vancomycin blood concentrations based on mimic-iv database. *China Pharmacy (ZHONGGUO YAOFANG)*, 36(19):2448–2453, 2025. doi: 10.6039/j.issn.1001-0408.2025.19.16. URL https://journal.china-pharmacy.com/en/article/doi/10.6039/j.issn.1001-0408.2025.19.16/.

[67] Zheyuan Lin, Xinhang Lin, Wanxin Li, Zhongxing Tian, Xudong Chai, Dongdong Zou, Weilin Xie, Yi Dong, and Yi Cai. Rapid few-shot tabular machine learning for ϕ-otdr event classification. *Optics Express*, 33(17):36646–36662, 2025. doi: 10.1364/OE.571235. URL https://opg.optica.org/oe/fulltext.cfm?uri=oe-33-17-36646&id=575783.

[68] Hao Liu et al. Characterizing clinical risk profiles of major complications in type 2 diabetes mellitus using deep learning algorithms. *Frontiers in Endocrinology*, 16:1657366, 2025. doi: 10.3389/fendo.2025.1657366. URL https://www.frontiersin.org/articles/10.3389/fendo.2025.1657366.

[69] Hejia Liu, Mochen Yang, and Gediminas Adomavicius. Just because you can, doesn't mean you should: Llms for data fitting. *arXiv preprint arXiv:2508.19563*, 2025. URL https://arxiv.org/abs/2508.19563.

[70] Bihui Lu, Kun Yu, Lin Qiu, Huayong Li, Hongxing Wang, Xiaohong Liu, Jie Shan, and Nan Li. Predicting county-level winter wheat yield in eastern china using multi-source spatiotemporal data: An explainable machine learning approach. SSRN preprint, 2025. URL https://ssrn.com/abstract=5380177.

[71] Junwei Ma, Apoorv Dankar, George Stein, Guangwei Yu, and Anthony L. Caterini. Tabpfgen – tabular data generation with tabpfn. *arXiv preprint arXiv:2406.05216*, 2024. doi: 10.48550/arXiv.2406.05216. URL https://arxiv.org/abs/2406.05216.

[72] Junwei Ma, Valentin Thomas, Rasa Hosseinzadeh, Hamidreza Kamkari, Alex Labach, Jesse C. Cresswell, Keyvan Golestan, Guangwei Yu, Anthony L. Caterini, and Maksims Volkovs. Tabdpt: Scaling tabular foundation models on real data, 2025. URL https://arxiv.org/abs/2410.18164.

[73] Yuchen Ma, Dennis Frauen, Emil Javurek, and Stefan Feuerriegel. Foundation models for causal inference via prior-data fitted networks, 2025. URL https://arxiv.org/abs/2506.10914.

[74] Luis Magadán, José Roldán-Gómez, Juan Carlos Granda, and Francisco José Suárez. Early fault classification in rotating machinery with limited data using TabPFN. *IEEE Sensors Journal*, 23(24):30960–30970, 2023. doi: 10.1109/JSEN.2023.3331100. URL https://ieeexplore.ieee.org/document/10318062.

[75] Asmaa A. Mahdi. Diagnosing patient stroke status using modern ai after dataset balancing: A comprehensive comparative study. *Journal of Scientific Reports*, 9(1):219–228, 2025. doi: 10.58970/JSR.1105. URL https://www.ijsab.com/jsr-volume-9-issue-1/8205.

[76] William Merrill and Ashish Sabharwal. Exact expressive power of transformers with padding. *CoRR*, abs/2505.18948, 2025. URL https://arxiv.org/abs/2505.18948. arXiv preprint.

[77] Brady Neal, Chin-Wei Huang, and Sunand Raghupathi. Realcause: Realistic causal inference benchmarking. *CoRR*, abs/2011.15007, 2020. URL https://arxiv.org/abs/2011.15007.

[78] Hoang Nguyen. Recovery rate prediction for corporate bonds — experiments. https://github.com/hoanguyen94/Recovery-rate-prediction. GitHub repository; accessed 7 Nov 2025.

[79] Ryunosuke Noda, Daisuke Ichikawa, and Yugo Shibagaki. Machine learning-based diagnostic prediction of minimal change disease: Model development study. *Scientific Reports*, 14:23460, 2024. doi: 10.1038/s41598-024-73898-4. URL https://www.nature.com/articles/s41598-024-73898-4.

[80] Authors not clearly specified. Enhancing reservoir parameter prediction workflows via advanced core data augmentation. ResearchGate preprint 395434405, 2025. URL https://www.researchgate.net/publication/395434405_Enhancing_Reservoir_Parameter_Prediction_Workflows_via_Advanced_Core_Data_Augmentation. Machine learning workflow including TabPFN for improved reservoir parameter prediction; please update with definitive metadata if published.

[81] Authors not clearly specified in the available metadata. The first 0.2 degrees resolution global continental heat flow map: Advancing fine-scale geothermal modeling. Preprint / technical report as indexed via ResearchGate, 2025. URL https://www.researchgate.net/publication/396728153_The_First_02_Resolution_Global_Continental_Heat_Flow_Map_Advancing_Fine-Scale_Geothermal_Modeling. Combines GeoClimaProx and TabPFN-style models for global heat flow estimation; please update with full author list and venue from the official publication if available.

[82] Fabian Offensperger, Ario de la Tin, Kevin Ogilvie, and et al. Large-scale chemoproteomics expedites ligand discovery and predicts ligand behavior in cells. *Science*, 384(6694):eadk5864, 2024. doi: 10.1126/science.adk5864. URL https://www.science.org/doi/10.1126/science.adk5864.

[83] Miruna Oprescu, Vasilis Syrgkanis, Keith Battocchi, Maggie Hei, and Greg Lewis. Econml: A machine learning library for estimating heterogeneous treatment effects. In *33rd Conference on Neural Information Processing Systems*, page 6, 2019.

[84] Mayra Pacheco-Cardín, Juan Luis Hernández-Arellano, José-Manuel Mejía-Muñoz, and Aide Aracely Maldonado-Macías. Comparison of machine learning and deep learning models in manual strength prediction using anthropometric variables. *International Journal of Occupational Safety and Ergonomics*, pages 1–10, 2025. doi: 10.1080/10803548.2025.2554461. Online ahead of print.

[85] Chaochao Pan et al. Sense-of-agency as clinically accessible features for schizophrenia prediction: Interpretable ensemble machine learning research and webserver development. *Asian Journal of Psychiatry*, 111:104674, 2025. doi: 10.1016/j.ajp.2025.104674. URL https://www.sciencedirect.com/science/article/pii/S187620182500317X.

[86] Nikunj Panchal, Abdul Qayum, Abdul Shahid, et al. Metrics-first, language-aware clone type recognition: Auditable signals across c, c#, java, and python. *Authorea Preprints*, 2025. doi: 10.22541/au.176059643.37779565/v1. URL https://wiley.authorea.com/users/980519/articles/1346750-metrics-first-language-aware-clone-type-recognition-auditable-signals-across-c-c-j

[87] Eloy Peña-Asensio, Josep M. Trigo-Rodríguez, Jordi Sort, Jordi Ibáñez-Insa, and Albert Rimola. Machine learning applications on lunar meteorite minerals: From classification to mechanical properties prediction. *International Journal of Mining Science and Technology*, 34(9):1283–1292, 2024. doi: 10.1016/j.ijmst.2024.08.001. URL https://www.sciencedirect.com/science/article/pii/S2095268624001010.

[88] Giulia Perciballi, Federica Granese, Ahmad Fall, Farida Zehraoui, Edi Prifti, and Jean-Daniel Zucker. Adapting tabpfn for zero-inflated metagenomic data. In *Table Representation Learning Workshop at NeurIPS 2024*, 2024. URL https://openreview.net/forum?id=3I0bVvUj25.

[89] Sindy Licette Piñero, Xiaomei Li, Lin Liu, Jiuyong Li, Sang Hong Lee, Marnie Winter, Thin Nguyen, Junpeng Zhang, and Thuc Duy Le. Taco: Tabpfn augmented causal outcomes for early detection of long covid. *medRxiv*, 2025. doi: 10.1101/2025.10.02.25337138. URL https://www.medrxiv.org/content/10.1101/2025.10.02.25337138v1.

[90] Prior Labs. How bostongene utilized tabpfn to identify immune system profiles associated with immunotherapy response in cancer patients. https://www.linkedin.com/pulse/how-bostongene-utilized-tabpfn-identify-immune-system-profiles-vexle/, 2025. Online case study on TabPFN in immune profiling. Accessed 7 Nov 2025.

[91] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

[92] Jingang Qu, David Holzmüller, Gaël Varoquaux, and Marine Le Morvan. Tabicl: A tabular foundation model for in-context learning on large data. *arXiv preprint arXiv:2502.05564*, 2025.

[93] Muhammad Moshiur Rahman, Andrew Robson, and Theo Bekker. Machine learning approaches for assessing avocado alternate bearing using sentinel-2 and climate variables—a case study in limpopo, south africa. *Preprints*, 2025(202510.2413), 2025. doi: 10.20944/preprints202510.2413.v1. URL https://www.preprints.org/manuscript/202510.2413. Preprint, version 1.

[94] Madhushan Ramalingam. Uncertainty-aware tabular prediction: Evaluating vbll-enhanced tabpfn in safety-critical medical data. *arXiv preprint arXiv:2509.10048*, 2025. URL https://arxiv.org/abs/2509.10048.

[95] Jake Robertson, Arik Reuter, Siyuan Guo, Noah Hollmann, Frank Hutter, and Bernhard Schölkopf. Do-pfn: In-context learning for causal effect estimation. *arXiv preprint arXiv:2506.06039*, 2025.

[96] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[97] Sergio Ruiz-Villafranca et al. Wfe-tab: Overcoming limitations of tabpfn in iiot-mec intrusion detection. *Future Generation Computer Systems*, 2025. URL https://www.sciencedirect.com/science/article/pii/S0167739X25000020. Weighted fusion-ensemble TabPFN for Industrial IoT intrusion detection.

[98] Taiga Saito, Yu Otake, and Stephen Wu. Tabular foundation model for geoai benchmark problems bm/airportsoilproperties/2/2025. *arXiv preprint arXiv:2509.03191*, 2025. URL https://arxiv.org/abs/2509.03191.

[99] Baha'a Zaher Saleh et al. Machine learning framework for energy consumption optimization using the tabpfnregressor algorithm. Preprint / technical report on wastewater treatment plant energy optimization, 2025. URL https://www.researchgate.net/publication/390516459_Machine_learning_framework_for_energy_consumption_optimization_using_the_TabPFNRegressor_algorithm. Details via ResearchGate preprint 390516459; please update with final publication metadata if available.

[100] Andreas Sauter, Saber Salehkaleybar, Aske Plaat, and Erman Acar. Activa: Amortized causal effect estimation via transformer-based variational autoencoder, 2025. URL https://arxiv.org/abs/2503.01290.

[101] Jonas Schmidinger, Viacheslav Barkov, Sebastian Vogel, Martin Atzmüller, and Gerard B. M. Heuvelink. Kriging prior regression: A case for kriging-based spatial features with tabpfn in soil mapping. *arXiv preprint arXiv:2509.09408*, 2025. URL https://arxiv.org/abs/2509.09408.

[102] Jay Shah, Ganesh Bikshandi, Ying Zhang, Vijay Thakkar, Pradeep Ramani, and Tri Dao. Flashattention-3: Fast and accurate attention with asynchrony and low-precision. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/7ede97c3e082c6df10a8d6103a2eebd2-Abstract-Conference.html.

[103] Bichen Shang, Guanzhe Li, Wei Sun, Liang Zhang, et al. In-context learning for nano-pcm thermal behavior prediction in battery thermal management via lattice boltzmann simulation. *Energy*, 2025. URL https://www.sciencedirect.com/science/article/pii/S036054422504335X. Evaluates TabPFN-style in-context learning for nano-PCM thermal behavior.

[104] Sandeep Sharma. Data and models for shape-selective adsorption in zeolites for long-chain alkane hydroisomerization. https://doi.org/10.4233/uuid:f36da034-5cb3-42ca-a53d-d351f68a9ffa, 2025. Repository associated with shape-selectivity modeling in zeolites; includes TabPFN-based components.

[105] Sandeep Sharma et al. Machine learning-based predictions of henry coefficients for long-chain alkanes in one-dimensional zeolites: Application to hydroisomerization. *The Journal of Physical Chemistry C*, 2025. doi: 10.1021/acs.jpcc.5c03868. URL https://pubs.acs.org/doi/10.1021/acs.jpcc.5c03868. In press / early access; uses ML including TabPFN-style approaches for Henry coefficient prediction.

[106] Shriyank Somvanshi, Pavan Hebli, Gaurab Chhetri, and Subasish Das. Tabular data with class imbalance: Predicting electric vehicle crash severity with pretrained transformers (tabpfn) and mamba-based models. *arXiv preprint arXiv:2509.11449*, 2025. URL https://arxiv.org/abs/2509.11449.

[107] Nurdaulet Tasmurzayev, Baglan Imanbek, Assiya Boltaboyeva, Gulmira Dikhanbayeva, Sarsenbek Zhussupbekov, Qarlygash Saparbayeva, and Gulshat Amirkhanova. Explainable ai for coronary artery disease stratification using routine clinical data. *Algorithms*, 18(11):693, 2025. doi: 10.3390/a18110693. URL https://www.mdpi.com/1999-4893/18/11/693.

[108] Melina Thegarza and co authors. Ml_climate final project: Flood impact on housing prices. Course project report, GitHub repository, 2023. URL https://github.com/melina-thegarza/ml-climate/blob/main/doc/ML_Climate___Final.pdf. Student project using machine learning (incl. TabPFN) for flood impact assessment.

[109] Vinh Quang Tran and Haewon Byeon. Predicting dementia in parkinson's disease on a small tabular dataset using hybrid lightgbm–tabpfn and shap. *Digital Health*, 10:20552076241272585, 2024. doi: 10.1177/20552076241272585. URL https://journals.sagepub.com/doi/10.1177/20552076241272585.

[110] Author(s) unavailable. Multimodal clinical prediction framework with tabular and phenotypic data from large-scale projects. MBZUAI thesis, institutional repository item 3e3d4c0d-dbcb-4d5b-a23e-e28aea840660, 2025. URL https://irep.mbzuai.ac.ae/items/3e3d4c0d-dbcb-4d5b-a23e-e28aea840660. Metadata limited; please update author and exact title from the repository.

[111] Toon Vanderschueren, Tim Verdonck, Mihaela van der Schaar, and Wouter Verbeke. AutoCATE: End-to-end, automated treatment effect estimation. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=QbOoz74GNO.

[112] Justus Viga, Penelope Mueck, Alexander Löser, and Torben Weis. Fuelcast: Benchmarking tabular and temporal models for ship fuel consumption. *arXiv preprint arXiv:2510.08217*, 2025. doi: 10.48550/arXiv.2510.08217. URL https://arxiv.org/abs/2510.08217.

[113] Konstantinos Vrettos, Konstantina Kasioumi, Nikolaos Galanakis, Elias Kehagias, Nikolaos Kontopodis, Nikolas Matthaiou, and Michail E. Klontzas. Radiomics enhance the prediction of endovascular treatment success for femoropopliteal chronic total occlusions: a proof-of-concept study. *European Journal of Radiology*, 194:112496, 2025. doi: 10.1016/j.ejrad.2025.112496. URL https://pubmed.ncbi.nlm.nih.gov/41166916/.

[114] Tuyen Vu, Ha Xuan Tran, Lin Liu, Jiuyong Li, Jia Tina Du, and Thuc Duy Le. Foundation model-based recommendation of optimal neoadjuvant therapy in breast cancer. *medRxiv*, 2025. doi: 10.1101/2025.10.03.25337255. URL https://www.medrxiv.org/content/10.1101/2025.10.03.25337255v1.

[115] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests, 2017. URL https://arxiv.org/abs/1510.04342.

[116] Chi Wang and Qingyun Wu. FLO: fast and lightweight hyperparameter optimization for automl. *CoRR*, abs/1911.04706, 2019. URL http://arxiv.org/abs/1911.04706.

[117] Hongyu Wang et al. Application of tabpfn model on the energy performance improvement of high-power multistage centrifugal pump. *Energy*, 2025. URL https://www.sciencedirect.com/science/article/abs/pii/S0360544225040411. Uses TabPFN-based modelling for entropy generation and efficiency optimization; see article S0360544225040411.

[118] Jie Wang, Junqi Deng, Siyi Li, Weijie Du, Zengqi Zhang, and Xiaoming Liu. Explainable machine learning for multicomponent concrete: Predictive modeling and feature interaction insights. *Materials*, 18(19):4456, 2025. doi: 10.3390/ma18194456. URL https://www.mdpi.com/1996-1944/18/19/4456.

[119] Peng Wang, Hongjun Liu, Yiming Shi, Ao Liu, Qingyu Zhu, Irina Albu, Maja Pacholec, Lulu Cheng, Xu Sun, and Xinli Chi. Harnessing small-data machine learning for transformative mental health forecasting: Towards precision psychiatry with personalised digital phenotyping. *Med Research*, 2025. doi: 10.1002/mdr2.70017. URL https://onlinelibrary.wiley.com/doi/10.1002/mdr2.70017.

[120] Yuandou Wang, Filip Gunnarsson, and Rihan Hai. Imlp: An energy-efficient continual learning method for tabular data streams. *arXiv preprint arXiv:2510.04660*, 2025. URL https://arxiv.org/abs/2510.04660.

[121] Wen Wen, Tingrui Zhang, Haina Zhao, Jingyan Liu, Heng Jiang, Yushuang He, and Zekun Jiang. Multimodal model enhances qualitative diagnosis of hypervascular thyroid nodules: integrating radiomics and deep learning features based on b-mode and pdi images. *Gland Surgery*, 14(8):1558–1571, 2025. doi: 10.21037/gs-2025-183. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC12432950/.

[122] H. Xu et al. Vision-language ai model for detecting pet/ct-occult nodal disease in patients with non-small-cell lung cancer treated with stereotactic ablative radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, 2025. Details from Red Journal abstract S0360-3016(25)05890-0.

[123] J. Xu and co authors. Multiscale prediction from ion concentrations to soil salinity in salinized farmland using machine learning. SSRN preprint 5591702, 2025. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5591702. Compares multiple models; TabPFN achieves strong performance for soil salinity prediction.

[124] Yan Xu, Zheng Xu, Chenyu Li, Lingyu Xu, Xinyuan Wang, Chen Guan, Siqi Jiang, Ningxin Zhang, Minghao Gu, and Yanlu Xin. Tabular prior data fitted network predicts acute kidney injury with routine clinical data. SSRN preprint, 2025. URL https://ssrn.com/abstract=5397006.

[125] Hyunseok Yang and Jungsu Park. Comparing the performance of a deep learning model (tabpfn) for predicting river algal blooms with varying data composition. *Journal of the Korean Wetlands Society*, 26(3):197–203, 2024. URL https://www.earticle.net/Article/A456244.

[126] Hang Yu, Sina Saffaran, Israel S. Maia, Enrico Clini, Declan G. Bates, and NIVPredict study group. Early prediction of non-invasive ventilation outcome using the tabpfn machine learning model: A multi-centre validation study. *Intensive Care Medicine*, 51(8):1542–1544, 2025. doi: 10.1007/s00134-025-08025-6. URL https://link.springer.com/article/10.1007/s00134-025-08025-6.

[127] Qiong Zhang, Yan Shuo Tan, Qinglong Tian, and Pengfei Li. Tabpfn: One model to rule them all? *arXiv preprint arXiv:2505.20003*, 2025.

[128] Xingxuan Zhang, Gang Ren, Han Yu, Hao Yuan, Hui Wang, Jiansheng Li, Jiayun Wu, Lang Mo, Li Mao, Mingchao Hao, Ningbo Dai, Renzhe Xu, Shuyang Li, Tianyang Zhang, Yue He, Yuanrui Wang, Yunjia Zhang, Zijing Xu, Dongzhe Li, Fang Gao, Hao Zou, Jiandong Liu, Jiashuo Liu, Jiawei Xu, Kaijie Cheng, Kehan Li, Linjun Zhou, Qing Li, Shaohua Fan, Xiaoyu Lin, Xinyan Han, Xuanyue Li, Yan Lu, Yuan Xue, Yuanyuan Jiang, Zimu Wang, Zhenlei Wang, and Peng Cui. Limix: Unleashing structured-data modeling capability for generalist intelligence. *arXiv preprint arXiv:2509.03505*, 2025.

[129] R. Zheng. A multitask deep learning framework for clinical decision-making in assisted reproductive technology. Master's thesis, Massachusetts Institute of Technology, 2025. URL https://dspace.mit.edu/handle/1721.1/162969. M.Eng. thesis.

[130] Jinying Zhu, Ping Xiong, Wei Wang, Tianshu Lu, and Defang Ouyang. Integrating artificial intelligence and physiologically based pharmacokinetic modeling to predict in vitro and in vivo fate of amorphous solid dispersions. *Journal of Controlled Release*, 386:114123, 2025. doi: 10.1016/j.jconrel.2025.114123. URL https://doi.org/10.1016/j.jconrel.2025.114123.

[131] Shidian Zhu, Hui Zhang, Yanlin Liu, Wenyu Bu, Qiang Wu, Jin Wang, Wandi Chen, Qiannong Wu, Zhirong Geng, and Fuming Liu. Development of an optimized risk evaluation system for cardiovascular-kidney-metabolic syndrome-associated coronary heart disease based on tabular prior-data fitted network. *Digital Health*, 11:20552076251379379, 2025. doi: 10.1177/20552076251379379. URL https://doi.org/10.1177/20552076251379379.

[132] Xiangang Zhu, Peidong Su, Jiang Yu, Jiaheng Pei, Zhaoyong Teng, Yougui Li, and Yuxuan Liu. A prediction model for hazard levels of shallow natural gas in tunnel based on k-means clustering and tabular prior-data fitted network. *Results in Engineering*, 21:106873, 2025. doi: 10.1016/j.rineng.2025.106873. URL https://www.sciencedirect.com/science/article/pii/S2590123025029366.

A Contributors

Model dev & Deployment

Léo Grinsztajn, Klemens Flöge, Oscar Key, Felix Birkel, Brendan Roof, Phil Jund, Benjamin Jäger, Adrian Hayler, Dominik Safaric, Simone Alessi, Felix Jablonski, Mihir Manium, Rosen Yu, Anurag Garg, Jake Robertson, Shi Bin (Liam) Hoo, Vladyslav Moroshan, Magnus Bühler, Lennart Purucker, Bernhard Schölkopf, Noah Hollmann, Frank Hutter

Distribution & Product

Clara Cornu, Lilly Charlotte Wehrhahn, Alessandro Bonetto, Sauraj Gambhir

B TabPFN Use Case Overview

TabPFNv2 has been applied to a broad set of use cases. We now list close to 100 published use cases across different industries.

Healthcare and Life Sciences

We collected 51 published TabPFN use cases in this area, by far more than in any other area; we attribute this partly to the scarcity of data in healthcare and life sciences, and partly to the open publishing culture in this area. Use cases span oncology, neurology, cardiology, psychiatry, nephrology, and pharmacology. Applications include diagnosis, prognosis, and treatment response prediction from multimodal clinical, imaging, and omics data, often under severe data scarcity.

1. TabPFN was applied to distinguish cancer patients from healthy individuals using immune system profiles from peripheral blood, facilitating predictions of immunotherapy responses [90]. Link
2. A machine learning model employing TabPFN was developed for non-invasive diagnostic prediction of minimal change disease in patients with nephrotic syndrome, utilizing clinical biomarkers [79]. Link
3. TabPFN was integrated into a system for analyzing T-cell receptor repertoires combined with clinical biomarkers to forecast immunotherapy outcomes in cancer patients, as explored by researchers at BostonGene [37]. Link
4. TabPFN enabled early detection of stillbirth risks through analysis of cardiotocography data, supporting improved prenatal care [8]. Link
5. Predictive modeling for postoperative outcomes following anterior cervical corpectomy utilized TabPFN to assess patient demographics and surgical parameters [54]. Link
6. A hybrid model incorporating TabPFN was introduced to predict dementia progression in Parkinson's disease patients, handling small datasets and missing values effectively [109]. Link
7. A machine learning model based on TabPFN was developed to predict 90-day unfavorable outcomes in stroke patients with distal vessel occlusions using CT perfusion imaging [55]. Link
8. TabPFN was utilized in chemoproteomics for identifying small-molecule fragment-protein interactions, aiding ligand discovery in drug development [82]. Link
9. TabPFN facilitated the prediction of non-invasive ventilation outcomes in patients with acute hypoxemic respiratory failure, supporting early identification of treatment failures [126]. Link
10. An interpretable Transformer-based model leveraging TabPFN was created to predict intravenous immunoglobulin resistance in pediatric patients with Kawasaki disease [21]. Link
11. TabPFN was employed in visual representation techniques for prostate cancer diagnosis, converting clinical biomarkers and symptom data into formats suitable for analysis [38]. Link
12. TabPFN was used to combine clinical, MR morphological, and delta-radiomics features to predict lymphovascular invasion in invasive breast cancer patients [64]. Link
13. TabPFN is proposed to predict mental health trajectories through digital phenotyping, enabling proactive and personalized interventions in precision psychiatry [119]. Link
14. TabPFN contributed to cardiovascular disease risk stratification using clinical features from a large patient cohort, incorporating interpretability techniques [17]. Link

15. TabPFN outperformed traditional machine learning models for early prediction of acute kidney injury in hospitalized patients, demonstrating generalizability across datasets [124]. Link
16. TabPFN was integrated into a framework for predicting postoperative mobility and discharge destinations in older adults using sensor data [59]. Link
17. TabPFN supported the prediction of infant temperament from maternal mental health data, aiding early identification of at-risk infants [7]. Link
18. TabPFN was employed to characterize clinical risk profiles for complications in type 2 diabetes mellitus patients, focusing on neuropathy and retinopathy [68]. Link
19. TabPFN was extended with a longitudinal-to-cross-sectional transformation to forecast Alzheimer's disease progression on neuroimaging datasets [35]. Link
20. TabPFN supported uncertainty calibration evaluation in medical data using variational techniques [94]. Link
21. TabPFN was applied to predict tumor response to chemotherapy in cholangiocarcinoma patients using RNA expression landscapes [62]. Link
22. TabPFN was incorporated into a generative model framework for tasks like data augmentation and imputation in biomedicine [71]. Link
23. TabPFN facilitated the prediction of gallstone malignancy risks through analysis of associated disease factors [18]. Link
24. TabPFN was used in classifying tuberculosis treatment outcomes based on clinical and sociodemographic data from national registries [13]. Link
25. TabPFN contributed to early prediction of gestational diabetes using cell-free DNA and genetic scores from early pregnancy blood samples [31]. Link
26. TabPFN was used for predicting schizophrenia based on sense of agency features, emphasizing interpretability [85]. Link
27. TabPFN was integrated into a physiologically based pharmacokinetic model for predicting dissolution and absorption of amorphous solid dispersions in drug development [130]. Link
28. TabPFN enabled classification of respiratory diseases from sound data, addressing clinical spectrum diversity [36]. Link
29. TabPFN was applied to small-data tabular learning in drug discovery, handling data scarcity and distribution shifts [27]. Link
30. TabPFN facilitated prediction of coronary heart disease risk in patients with cardiovascular-kidney-metabolic syndrome, optimizing evaluation in small samples [131]. Link
31. TabPFN was used to predict success of allogeneic stem cell mobilization in donors, aiding transplant therapies [5]. Link
32. TabPFN contributed to predicting manual strength using anthropometric data, focusing on accuracy and interpretability [84]. Link
33. TabPFN supported uncertainty-guided model selection for biomolecule efficacy prediction, enhancing ensemble optimization in drug discovery, as studied at GSK [63]. Link
34. TabPFN was utilized in a multitask deep learning framework for optimizing in vitro fertilization decisions, including embryo transfer and pregnancy prediction [129]. Link
35. TabPFN enabled a framework for early Long COVID detection through causal gene identification and interpretability [89]. Link
36. TabPFN was used in a foundation model approach for neoadjuvant therapy recommendations in breast cancer, integrating multi-omics data [114]. Link
37. Recent work has demonstrated explainable machine learning pipelines for coronary artery disease stratification from routine clinical data [107]. Link
38. TabPFN facilitated prediction of recurrence and progression in oral potentially malignant disorder patients post-surgery [4]. Link
39. TabPFN supported prediction of occult lymph node metastasis in non-small cell lung cancer patients treated with stereotactic ablative radiotherapy [122]. Link

40. TabPFN was used in stroke diagnosis, addressing dataset imbalance and model interpretability for clinical decisions [75]. Link
41. TabPFN was integrated into a multimodal thesis framework for clinical predictions using tabular and phenotypic data from large-scale projects [110]. Link
42. TabPFN was used to predict diabetes-related hypo- and hyperglycemia during hemodialysis using continuous glucose monitoring data, facilitating improved patient management [43]. Link
43. TabPFN was applied to enhance diagnosis of hypervascular thyroid nodules using multimodal ultrasound features [121]. Link
44. TabPFN was integrated with radiomics and clinical features to predict endovascular treatment success in femoropopliteal chronic total occlusions, supporting interventional planning [113]. Link
45. TabPFN was applied to CorvisST biomechanical indices to classify corneal disorders, improving diagnostic accuracy in ophthalmology [15]. Link
46. TabPFN was incorporated into a non-invasive sleep staging framework using respiratory sound features, advancing passive sleep monitoring [28]. Link
47. TabPFN supported prediction of vancomycin blood concentrations to optimize antimicrobial dosing strategies in clinical practice [66]. Link
48. TabPFN was used to predict negative self-rated oral health in adults, identifying risk factors for targeted public-health interventions [9]. Link
49. TabPFN was extended to very high-dimensional feature spaces to enable robust analysis of biomedical data, improving stability and interpretability in clinical applications [60]. Link
50. TabPFN predicted gastrointestinal bleeding risk in pediatric Henoch–Schönlein purpura patients, supporting early clinical intervention [22]. Link
51. TabPFN was used as the pre-trained backbone (embeddings + in-context learning) for silica nanoparticle cellular toxicity prediction [2]. Link

Financial Services, Banking, and Insurance

While we have seen strong customer interest in this area, this is not reflected by the relatively few published use cases (only 3) we managed to collect; we attribute this to the domain's competitive nature and disinclination to publish.

1. TabPFN was applied to usage-based premium calculations in actuarial science, leveraging driving behavior data from IoT devices [16]. Link
2. TabPFN facilitated cross-selling of health insurance products through deep learning analysis of customer data [29]. Link
3. TabPFN was used in corporate bond recovery rate prediction for credit risk management [78]. Link

Energy and Utilities

We collected 15 use cases focused on environmental forecasting (algal blooms, wildfire, rainfall), renewable-energy nowcasting, process/asset optimization across water, oil & gas, and materials.

1. TabPFN was employed to predict river algal blooms through multi-classification of chlorophyll-a concentrations, aiding water management [125]. Link
2. TabPFN facilitated wildfire propagation prediction in Canadian conifer forests, classifying fire types for environmental risk assessment [58]. Link
3. TabPFN was integrated into a machine learning framework for optimizing energy consumption at wastewater treatment plants [99]. Link
4. TabPFN supported rainfall forecast post-processing using historical error patterns from environmental data [3]. Link

5. TabPFN enabled solar forecast error adjustment, particularly during rapid weather changes, as developed by Open Climate Fix [42]. Link
6. TabPFN was applied to predict ash fusibility in high-alkali coal for improved energy production [20]. Link
7. TabPFN contributed to predicting Henry coefficients for alkanes in zeolites, aiding hydroisomerization in sustainable fuel production [105]. Link
8. TabPFN facilitated shape-selectivity modeling in zeolites for long-chain alkane hydroisomerization, optimizing catalyst design [104]. Link
9. TabPFN was used in an integrated framework for estimated ultimate recovery prediction and fracturing optimization in shale gas reservoirs [24]. Link
10. TabPFN supported core data augmentation for enhanced reservoir parameter prediction in oil and gas exploration [80]. Link
11. TabPFN was employed to optimize energy performance in multistage centrifugal pumps through entropy generation analysis [117]. Link
12. TabPFN contributed to physics-informed regression for evaluating solar-reflective materials in facade temperature modeling [25]. Link
13. TabPFN was applied to generate advanced global heat flow maps at 0.2° resolution, integrating high-resolution geophysical data to improve geothermal resource modeling [81]. Link
14. TabPFN contributed to FuelCast, standardizing benchmarks for ship fuel consumption prediction and improving efficiency in maritime operations [112]. Link
15. TabPFN was used as the main supervised classifier to automatically identify thunderstorm ground enhancements from particle detector and environmental measurements [10]. Link

Manufacturing and Industrial

We collected 12 diverse use cases including anomaly detection, predictive maintenance, physics-aware optimization—spanning IIoT security, rotating machinery, semiconductor testing, geotechnical/optical sensing, machining, battery thermal modeling, and concrete mix design.

1. TabPFN enabled early fault classification in rotating machinery, addressing data scarcity in industrial scenarios [74]. Link
2. TabPFN facilitated microcontroller performance prediction, aiding semiconductor screening with minimal supervision, as studied at Infineon Technologies [14]. Link
3. TabPFN was applied to caisson inclination prediction in ultra-deep construction, combining data denoising techniques [47]. Link
4. TabPFN supported event classification in phase-sensitive optical time-domain reflectometry systems for distributed fiber sensing [67]. Link
5. TabPFN was integrated into an adaptive ensemble for intrusion detection in Industrial Internet of Things networks [97]. Link
6. TabPFN enabled a random forest-based framework for attack recognition in Internet of Things networks, improving interpretability [65]. Link
7. TabPFN facilitated geotechnical site characterization for predicting soil strength and imputing mechanical parameters [98]. Link
8. TabPFN was used in cryogenic-assisted abrasive waterjet machining for improving surface integrity in titanium alloys [19]. Link
9. TabPFN supported in-context learning for thermal behavior prediction in nano-phase change materials for battery systems [103]. Link
10. TabPFN was applied to explainable strength evaluation in multicomponent concrete mixtures [118]. Link

11. TabPFN was integrated into a multimodal fusion framework linking microstructure to friction behavior in martensitic stainless steel, improving wear resistance in materials engineering applications [53]. Link
12. TabPFN supported multiscale modeling to predict soil salinity in arid farmland, advancing sustainable agricultural management in regions such as Xinjiang [123]. Link

Other Industries

We collected 19 further heterogeneous TabPFN applications spanning geoscience, forensic science, agriculture, materials, and engineering domains—ranging from microbiome classification and lunar regolith analysis to soil property modeling, crop yield and phenology forecasting, fuel-blend optimization, and spatial regression.

1. TabPFN was modified for microbiome data classification in metagenomics, matching species abundance patterns with synthetic priors [88]. Link
2. TabPFN enabled lunar regolith analysis for classifying meteorite compositions from spectral data [87]. Link
3. TabPFN facilitated winter wheat yield forecasting in agricultural regions by integrating climate and remote sensing data [70]. Link
4. TabPFN was applied to flood impact assessment on housing prices by geographic areas [108]. Link
5. TabPFN showed the strongest performance on 31 predictive soil modeling datasets containing 30 to 460 samples [12]. Link
6. TabPFN was applied to shallow natural gas hazard prediction in tunnel construction [132]. Link
7. TabPFN supported automated feature engineering for energy consumption forecasting in domain-specific applications [6]. Link
8. TabPFN enabled Australian rice phenology prediction using remote sensing and weather data for crop management [52]. Link
9. TabPFN was applied to a multi-stage framework for predicting fuel blend properties through automated feature engineering [32]. Link
10. TabPFN enabled kriging prior regression for incorporating spatial context in soil mapping predictions [101]. Link
11. TabPFN was applied to predicting electric vehicle crash severity using deep learning models [106]. Link
12. TabPFN enhanced clone-type recognition across programming languages through metrics-driven analysis, improving stability and interpretability in software engineering [86]. Link
13. TabPFN was used to predict biomass-derived hard carbon performance in sodium-ion batteries, facilitating material selection for energy storage systems [23]. Link
14. TabPFN informed the development of TabImpute, enabling efficient zero-shot imputation for missing tabular data and improving preprocessing pipelines [41]. Link
15. TabPFN, alongside TabICL and related foundation models, was evaluated for intrusion detection, improving cybersecurity performance in IoT networks [44]. Link
16. TabPFN supported continual learning for tabular data streams in resource-constrained environments [120]. Link
17. TabPFN contributed to assessing robustness of language models for data fitting under irrelevant variations [69]. Link
18. TabPFN was used in forensic science to advance biogeographical ancestry predictions [48]. Link
19. TabPFN was used as a benchmark model for predicting avocado alternate bearing from Sentinel-2 and climate features [93]. Link

Healthcare and Life Sciences

We collected 51 published TabPFN use cases in this area, by far more than in any other area; we attribute this partly to the scarcity of data in healthcare and life sciences, and partly to the open publishing culture in this area. Use cases span oncology, neurology, cardiology, psychiatry, nephrology, and pharmacology. Applications include diagnosis, prognosis, and treatment response prediction from multimodal clinical, imaging, and omics data, often under severe data scarcity.

1. TabPFN was applied to distinguish cancer patients from healthy individuals using immune system profiles from peripheral blood, facilitating predictions of immunotherapy responses [90]. Link
2. A machine learning model employing TabPFN was developed for non-invasive diagnostic prediction of minimal change disease in patients with nephrotic syndrome, utilizing clinical biomarkers [79]. Link
3. TabPFN was integrated into a system for analyzing T-cell receptor repertoires combined with clinical biomarkers to forecast immunotherapy outcomes in cancer patients, as explored by researchers at BostonGene [37]. Link
4. TabPFN enabled early detection of stillbirth risks through analysis of cardiotocography data, supporting improved prenatal care [8]. Link
5. Predictive modeling for postoperative outcomes following anterior cervical corpectomy utilized TabPFN to assess patient demographics and surgical parameters [54]. Link
6. A hybrid model incorporating TabPFN was introduced to predict dementia progression in Parkinson's disease patients, handling small datasets and missing values effectively [109]. Link
7. A machine learning model based on TabPFN was developed to predict 90-day unfavorable outcomes in stroke patients with distal vessel occlusions using CT perfusion imaging [55]. Link
8. TabPFN was utilized in chemoproteomics for identifying small-molecule fragment-protein interactions, aiding ligand discovery in drug development [82]. Link
9. TabPFN facilitated the prediction of non-invasive ventilation outcomes in patients with acute hypoxemic respiratory failure, supporting early identification of treatment failures [126]. Link
10. An interpretable Transformer-based model leveraging TabPFN was created to predict intravenous immunoglobulin resistance in pediatric patients with Kawasaki disease [21]. Link
11. TabPFN was employed in visual representation techniques for prostate cancer diagnosis, converting clinical biomarkers and symptom data into formats suitable for analysis [38]. Link
12. TabPFN was used to combine clinical, MR morphological, and delta-radiomics features to predict lymphovascular invasion in invasive breast cancer patients [64]. Link
13. TabPFN is proposed to predict mental health trajectories through digital phenotyping, enabling proactive and personalized interventions in precision psychiatry [119]. Link
14. TabPFN contributed to cardiovascular disease risk stratification using clinical features from a large patient cohort, incorporating interpretability techniques [17]. Link
15. TabPFN outperformed traditional machine learning models for early prediction of acute kidney injury in hospitalized patients, demonstrating generalizability across datasets [124]. Link
16. TabPFN was integrated into a framework for predicting postoperative mobility and discharge destinations in older adults using sensor data [59]. Link
17. TabPFN supported the prediction of infant temperament from maternal mental health data, aiding early identification of at-risk infants [7]. Link
18. TabPFN was employed to characterize clinical risk profiles for complications in type 2 diabetes mellitus patients, focusing on neuropathy and retinopathy [68]. Link

19. TabPFN was extended with a longitudinal-to-cross-sectional transformation to forecast Alzheimer's disease progression on neuroimaging datasets [35]. Link
20. TabPFN supported uncertainty calibration evaluation in medical data using variational techniques [94]. Link
21. TabPFN was applied to predict tumor response to chemotherapy in cholangiocarcinoma patients using RNA expression landscapes [62]. Link
22. TabPFN was incorporated into a generative model framework for tasks like data augmentation and imputation in biomedicine [71]. Link
23. TabPFN facilitated the prediction of gallstone malignancy risks through analysis of associated disease factors [18]. Link
24. TabPFN was used in classifying tuberculosis treatment outcomes based on clinical and sociodemographic data from national registries [13]. Link
25. TabPFN contributed to early prediction of gestational diabetes using cell-free DNA and genetic scores from early pregnancy blood samples [31]. Link
26. TabPFN was used for predicting schizophrenia based on sense of agency features, emphasizing interpretability [85]. Link
27. TabPFN was integrated into a physiologically based pharmacokinetic model for predicting dissolution and absorption of amorphous solid dispersions in drug development [130]. Link
28. TabPFN enabled classification of respiratory diseases from sound data, addressing clinical spectrum diversity [36]. Link
29. TabPFN was applied to small-data tabular learning in drug discovery, handling data scarcity and distribution shifts [27]. Link
30. TabPFN facilitated prediction of coronary heart disease risk in patients with cardiovascular-kidney-metabolic syndrome, optimizing evaluation in small samples [131]. Link
31. TabPFN was used to predict success of allogeneic stem cell mobilization in donors, aiding transplant therapies [5]. Link
32. TabPFN contributed to predicting manual strength using anthropometric data, focusing on accuracy and interpretability [84]. Link
33. TabPFN supported uncertainty-guided model selection for biomolecule efficacy prediction, enhancing ensemble optimization in drug discovery, as studied at GSK [63]. Link
34. TabPFN was utilized in a multitask deep learning framework for optimizing in vitro fertilization decisions, including embryo transfer and pregnancy prediction [129]. Link
35. TabPFN enabled a framework for early Long COVID detection through causal gene identification and interpretability [89]. Link
36. TabPFN was used in a foundation model approach for neoadjuvant therapy recommendations in breast cancer, integrating multi-omics data [114]. Link
37. Recent work has demonstrated explainable machine learning pipelines for coronary artery disease stratification from routine clinical data [107]. Link
38. TabPFN facilitated prediction of recurrence and progression in oral potentially malignant disorder patients post-surgery [4]. Link
39. TabPFN supported prediction of occult lymph node metastasis in non-small cell lung cancer patients treated with stereotactic ablative radiotherapy [122]. Link
40. TabPFN was used in stroke diagnosis, addressing dataset imbalance and model interpretability for clinical decisions [75]. Link
41. TabPFN was integrated into a multimodal thesis framework for clinical predictions using tabular and phenotypic data from large-scale projects [110]. Link
42. TabPFN was used to predict diabetes-related hypo- and hyperglycemia during hemodialysis using continuous glucose monitoring data, facilitating improved patient management [43]. Link
43. TabPFN was applied to enhance diagnosis of hypervascular thyroid nodules using multimodal ultrasound features [121]. Link

44. TabPFN was integrated with radiomics and clinical features to predict endovascular treatment success in femoropopliteal chronic total occlusions, supporting interventional planning [113]. Link
45. TabPFN was applied to CorvisST biomechanical indices to classify corneal disorders, improving diagnostic accuracy in ophthalmology [15]. Link
46. TabPFN was incorporated into a non-invasive sleep staging framework using respiratory sound features, advancing passive sleep monitoring [28]. Link
47. TabPFN supported prediction of vancomycin blood concentrations to optimize antimicrobial dosing strategies in clinical practice [66]. Link
48. TabPFN was used to predict negative self-rated oral health in adults, identifying risk factors for targeted public-health interventions [9]. Link
49. TabPFN was extended to very high-dimensional feature spaces to enable robust analysis of biomedical data, improving stability and interpretability in clinical applications [60]. Link
50. TabPFN predicted gastrointestinal bleeding risk in pediatric Henoch–Schönlein purpura patients, supporting early clinical intervention [22]. Link
51. TabPFN was used as the pre-trained backbone (embeddings + in-context learning) for silica nanoparticle cellular toxicity prediction [2]. Link

Financial Services, Banking, and Insurance

While we have seen strong customer interest in this area, this is not reflected by the relatively few published use cases (only 3) we managed to collect; we attribute this to the domain's competitive nature and disinclination to publish.

1. TabPFN was applied to usage-based premium calculations in actuarial science, leveraging driving behavior data from IoT devices [16]. Link
2. TabPFN facilitated cross-selling of health insurance products through deep learning analysis of customer data [29]. Link
3. TabPFN was used in corporate bond recovery rate prediction for credit risk management [78]. Link

Energy and Utilities

We collected 15 use cases focused on environmental forecasting (algal blooms, wildfire, rainfall), renewable-energy nowcasting, process/asset optimization across water, oil & gas, and materials.

1. TabPFN was employed to predict river algal blooms through multi-classification of chlorophyll-a concentrations, aiding water management [125]. Link
2. TabPFN facilitated wildfire propagation prediction in Canadian conifer forests, classifying fire types for environmental risk assessment [58]. Link
3. TabPFN was integrated into a machine learning framework for optimizing energy consumption at wastewater treatment plants [99]. Link
4. TabPFN supported rainfall forecast post-processing using historical error patterns from environmental data [3]. Link
5. TabPFN enabled solar forecast error adjustment, particularly during rapid weather changes, as developed by Open Climate Fix [42]. Link
6. TabPFN was applied to predict ash fusibility in high-alkali coal for improved energy production [20]. Link
7. TabPFN contributed to predicting Henry coefficients for alkanes in zeolites, aiding hydroisomerization in sustainable fuel production [105]. Link
8. TabPFN facilitated shape-selectivity modeling in zeolites for long-chain alkane hydroisomerization, optimizing catalyst design [104]. Link

9. TabPFN was used in an integrated framework for estimated ultimate recovery prediction and fracturing optimization in shale gas reservoirs [24]. Link
10. TabPFN supported core data augmentation for enhanced reservoir parameter prediction in oil and gas exploration [80]. Link
11. TabPFN was employed to optimize energy performance in multistage centrifugal pumps through entropy generation analysis [117]. Link
12. TabPFN contributed to physics-informed regression for evaluating solar-reflective materials in facade temperature modeling [25]. Link
13. TabPFN was applied to generate advanced global heat flow maps at 0.2° resolution, integrating high-resolution geophysical data to improve geothermal resource modeling [81]. Link
14. TabPFN contributed to FuelCast, standardizing benchmarks for ship fuel consumption prediction and improving efficiency in maritime operations [112]. Link
15. TabPFN was used as the main supervised classifier to automatically identify thunderstorm ground enhancements from particle detector and environmental measurements [10]. Link

Manufacturing and Industrial

We collected 12 diverse use cases including anomaly detection, predictive maintenance, physics-aware optimization—spanning IIoT security, rotating machinery, semiconductor testing, geotechnical/optical sensing, machining, battery thermal modeling, and concrete mix design.

1. TabPFN enabled early fault classification in rotating machinery, addressing data scarcity in industrial scenarios [74]. Link
2. TabPFN facilitated microcontroller performance prediction, aiding semiconductor screening with minimal supervision, as studied at Infineon Technologies [14]. Link
3. TabPFN was applied to caisson inclination prediction in ultra-deep construction, combining data denoising techniques [47]. Link
4. TabPFN supported event classification in phase-sensitive optical time-domain reflectometry systems for distributed fiber sensing [67]. Link
5. TabPFN was integrated into an adaptive ensemble for intrusion detection in Industrial Internet of Things networks [97]. Link
6. TabPFN enabled a random forest-based framework for attack recognition in Internet of Things networks, improving interpretability [65]. Link
7. TabPFN facilitated geotechnical site characterization for predicting soil strength and imputing mechanical parameters [98]. Link
8. TabPFN was used in cryogenic-assisted abrasive waterjet machining for improving surface integrity in titanium alloys [19]. Link
9. TabPFN supported in-context learning for thermal behavior prediction in nano-phase change materials for battery systems [103]. Link
10. TabPFN was applied to explainable strength evaluation in multicomponent concrete mixtures [118]. Link
11. TabPFN was integrated into a multimodal fusion framework linking microstructure to friction behavior in martensitic stainless steel, improving wear resistance in materials engineering applications [53]. Link
12. TabPFN supported multiscale modeling to predict soil salinity in arid farmland, advancing sustainable agricultural management in regions such as Xinjiang [123]. Link

Other Industries

We collected 19 further heterogeneous TabPFN applications spanning geoscience, forensic science, agriculture, materials, and engineering domains—ranging from microbiome classification and lunar regolith analysis to soil property modeling, crop yield and phenology forecasting, fuel-blend optimization, and spatial regression.

1. TabPFN was modified for microbiome data classification in metagenomics, matching species abundance patterns with synthetic priors [88]. Link
2. TabPFN enabled lunar regolith analysis for classifying meteorite compositions from spectral data [87]. Link
3. TabPFN facilitated winter wheat yield forecasting in agricultural regions by integrating climate and remote sensing data [70]. Link
4. TabPFN was applied to flood impact assessment on housing prices by geographic areas [108]. Link
5. TabPFN showed the strongest performance on 31 predictive soil modeling datasets containing 30 to 460 samples [12]. Link
6. TabPFN was applied to shallow natural gas hazard prediction in tunnel construction [132]. Link
7. TabPFN supported automated feature engineering for energy consumption forecasting in domain-specific applications [6]. Link
8. TabPFN enabled Australian rice phenology prediction using remote sensing and weather data for crop management [52]. Link
9. TabPFN was applied to a multi-stage framework for predicting fuel blend properties through automated feature engineering [32]. Link
10. TabPFN enabled kriging prior regression for incorporating spatial context in soil mapping predictions [101]. Link
11. TabPFN was applied to predicting electric vehicle crash severity using deep learning models [106]. Link
12. TabPFN enhanced clone-type recognition across programming languages through metrics-driven analysis, improving stability and interpretability in software engineering [86]. Link
13. TabPFN was used to predict biomass-derived hard carbon performance in sodium-ion batteries, facilitating material selection for energy storage systems [23]. Link
14. TabPFN informed the development of TabImpute, enabling efficient zero-shot imputation for missing tabular data and improving preprocessing pipelines [41]. Link
15. TabPFN, alongside TabICL and related foundation models, was evaluated for intrusion detection, improving cybersecurity performance in IoT networks [44]. Link
16. TabPFN supported continual learning for tabular data streams in resource-constrained environments [120]. Link
17. TabPFN contributed to assessing robustness of language models for data fitting under irrelevant variations [69]. Link
18. TabPFN was used in forensic science to advance biogeographical ancestry predictions [48]. Link
19. TabPFN was used as a benchmark model for predicting avocado alternate bearing from Sentinel-2 and climate features [93]. Link

C License and Availability

We release TabPFN-2.5 under our `TABPFN-2.5 License v1.0` designed to be permissive for research and internal evaluation. It *explicitly allows* testing, evaluation, and internal benchmarking, so an organization can download the model and run preliminary assessments on its own datasets.

The key restriction is that the model, its derivatives, and its outputs cannot be used for any commercial or production purpose. This includes, but is not limited to, revenue-generating products, competitive benchmarking for procurement, client deliverables, or using the model's results for internal commercial decision-making.

For production use cases, we offer a *Commercial Enterprise License*. This provides access to our proprietary high-speed inference engine, dedicated support, integration tooling, and other internal models.

Please contact us at sales@priorlabs.ai for commercial licensing inquiries. The full non-commercial mode license text can be found at `https://huggingface.co/Prior-Labs/tabpfn_2_5/blob/main/LICENSE`.

C.1 Cloud API

We provide a managed TabPFN-2.5 cloud endpoint, which runs on our optimized GPU infrastructure. This is the recommended option for users who do not have a dedicated local GPU or for those who wish to use TabPFN commercially without purchasing a full on-premise license.

The API is accessible via a simple Python SDK[2] (`pip install tabpfn-client`) or a standard REST API, allowing for integration into both non-commercial and commercial applications.

D The TabPFN Ecosystem

Figure 3 provides a minimal user workflow through components in the TabPFN–Extensions ecosystem.

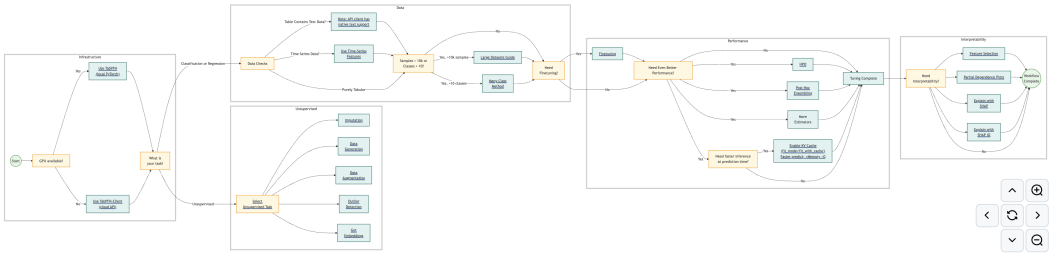


Figure 3: A minimal user workflow through components in the TabPFN–Extensions ecosystem.

E How to Get Optimal Fit + Predict Speed from TabPFN-2.5

To achieve good performance, we recommend the following:

- **Use a dedicated GPU or GPUs**: We recommend NVIDIA H100 or A100 GPUs. Any dedicated GPU supported by PyTorch is compatible, but some models may not have enough memory for larger datasets or perform slowly. Integrated GPUs, MPS (Apple Silicon), and CPUs are also supported, but are only suitable for small datasets.
- **Use multiple GPUs**: For larger datasets, fit + predict time can be dramatically reduced by parallelizing inference over several GPUs. To enable this, set the `device` parameter of `TabPFNClassifier` and `TabPFNRegressor`.
- **Use batch inference**: Unless the fitted-model cache is enabled (see below), the model is retrained each time `.predict()` is called. This means that it is much faster to make a prediction for all your test points in a single `.predict()` call. If you run out of memory, split the test points into batches of 1000 to 10000 and call `.predict()` for each batch.
- **Use PyTorch 2.8 or above**: TabPFN-2.5 also supports earlier versions of PyTorch, but these may have lower performance.

[2]The Python client SDK is available on PyPI: `https://github.com/PriorLabs/TabPFN-client`

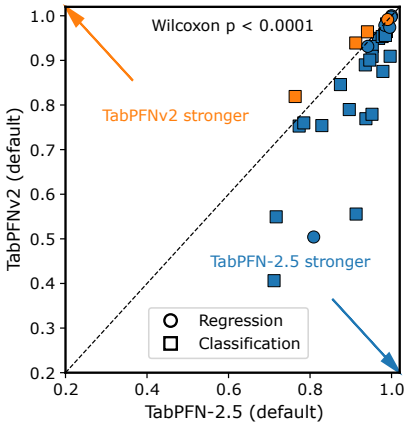


Figure 4: **TabPFN-2.5 clearly outperforms TabPFNv2.** We show here the normalized performance of TabPFN-2.5 and TabPFNv2 for each dataset of the TabPFNv2 subset on TabArena. TabPFN-2.5 often performs much better and is never much worse.

- **For small datasets, enable the fitted-model cache**: This is an experimental feature that trains and stores the model during `.fit()`, making subsequent `.predict()` calls fast by using a KV-Cache. It is enabled by setting the `fit_mode` parameter of `TabPFNClassifier` and `TabPFNRegressor` to `fit_with_cache`. However, with this setting classification models will consume approximately 6.1 KB of GPU memory and 48.8 KB of CPU memory per cell in the training dataset (regression models about 25% less), thus it is currently only suitable for small training datasets. For larger datasets and CPU-based inference, we recommend the TabPFN-as-MLP/Tree output engine.
- If speed is important for your application, you may consider optimizing the `memory_saving_mode` and `n_preprocessing_jobs` parameters of `TabPFNClassifier` and `TabPFNRegressor`. See the code documentation for further information.

Figure 17 in the appendix shows the inference latency you can expect for three common models of GPU, when using one or four GPUs. It also shows the maximum dataset size that fits in memory for each GPU.

F Detailed TabArena Results

In addition to the results shown in Section 3, we also report the pairwise winrates of different models on TabArena in Figure 5 (for TabPFNv2 compatible datasets with less than 10k rows and 500 features) and Figure 6 (all datasets up to 100k training rows and 2k features).

We also compare our TabPFN-2.5 model to other foundation models in more detail below. In Figure 7, we show that TabPFN-2.5 outperforms TabICL when we restrict TabArena to only datasets for which TabICL is designed, and in Figure 8, we show much better performance when compared to LimiX's results on datasets with less than 50,000 samples and 2,000 features, which corresponds to the datasets on which the TabArena maintainers could run LimiX at the time of writing (see this link).

G Additional Internal Benchmark Details

G.1 Performance on Internal Benchmarks

A diverse internal benchmark. In addition to the public TabArena benchmark, we built our own benchmarking framework using proprietary data. It includes over 100 use cases from healthcare, finance, insurance, retail and manufacturing. This benchmark focuses on comparing to gradient-boosted decision tree libraries that are frequently used in industry (XGBoost [26], CatBoost [91], LightGBM [57]), both in their default version and tuned for one hour. In all cases, we show the results of three standard gradient-boosted tree libraries (LightGBM, XGBoost and CatBoost). We

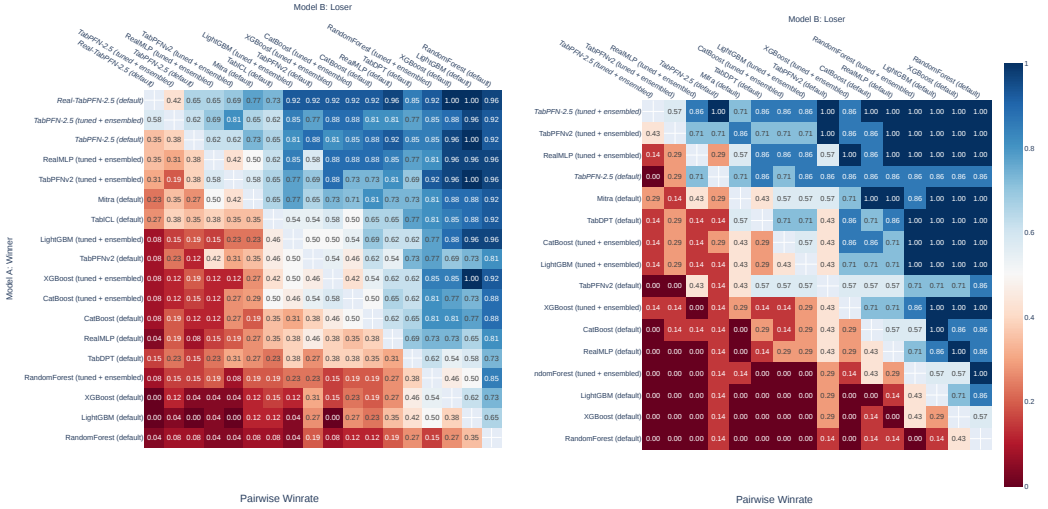


Figure 5: TabArena-Lite pairwise win rates on **classification** (left) and **regression** (right), restricted to TabPFNv2 compatible datasets (less than **10K training samples and 500 features**). Note that tuning for TabPFN-2.5 is only based on 60 random configs compared to 200 for the baselines.

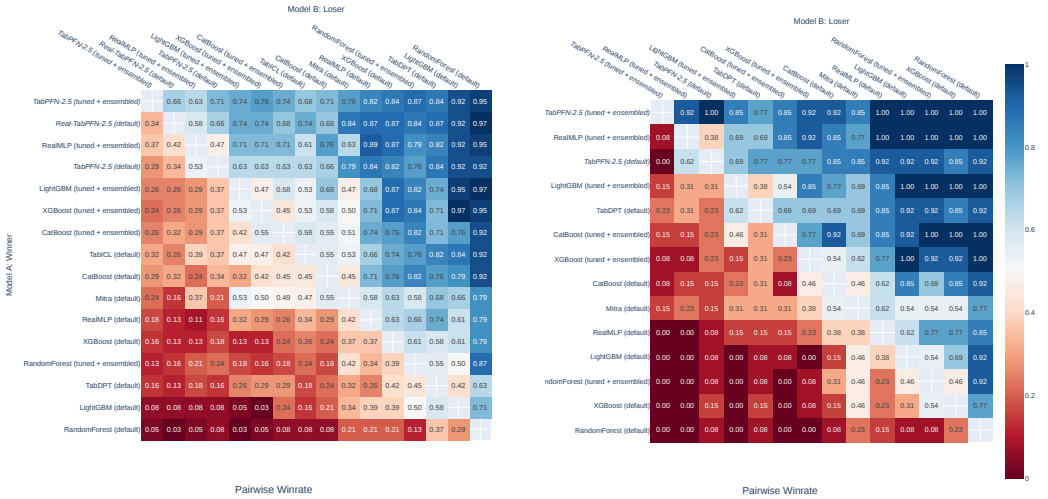


Figure 6: TabArena-Lite pairwise win rates on **classification** (left) and **regression** (right), evaluated on **all datasets** (up to **100k training samples and 2K features**). Note that tuning for TabPFN-2.5 is only based on 60 random configs compared to 200 for the baselines.

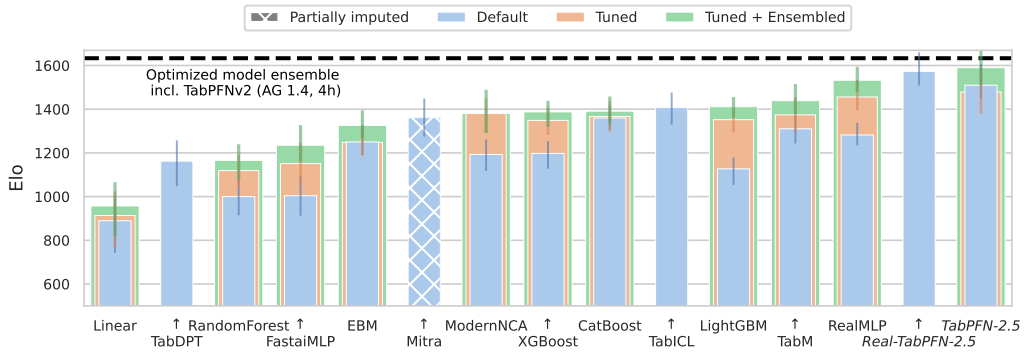


Figure 7: **Comparison with TabICL [92]**. In this plot, we show the performance of TabPFN-2.5 and TabICL on a TabArena-lite subset compatible with TabICL, restricting to **classification datasets with less than 500 features**. On this subset for which TabICL is designed, we see that TabPFN-2.5 significantly outperforms TabICL.

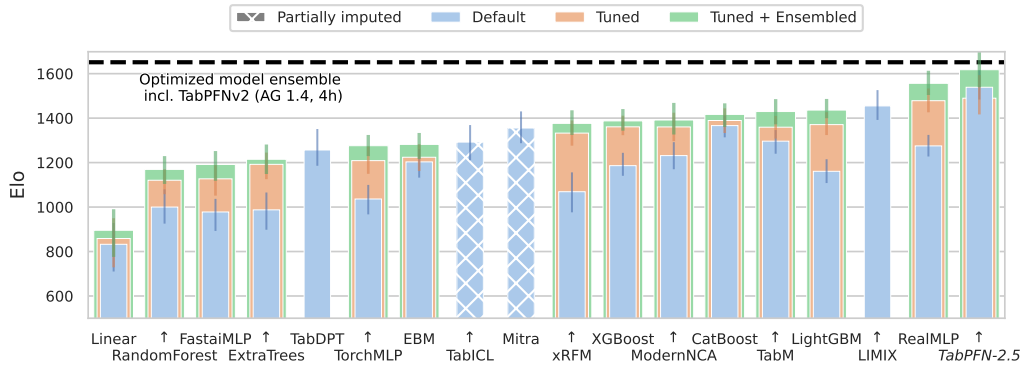


Figure 8: **Comparison with LimiX [128]**. In this plot, we show the performance of TabPFN-2.5 and LimiX on datasets from TabArena-Lite with **less than 50,000 training samples and less than 2,000 features**, which corresponds to the datasets on which the TabArena maintainers could run LimiX at the time of writing (see this link). On this subset, we see that TabPFN-2.5 significantly outperforms LimiX. Note that these results are still unverified by the original authors at the time of writing and thus not included in the main paper results.

tune all of the baselines for 1hr, using random search on the established search spaces from [50]. TabPFN is tuned using our AutoTabPFN system, resulting in a tuned and ensembled model.

TabPFN-2.5 shows strong results up to 50,000 samples and 2,000 features. Figure 9 and Figure 10 show results on our internal benchmark for classification and regression datasets with up to 50,000 data points and 500 features. We can see on these figures that TabPFN outperforms in one forward pass all our tuned baselines. In Section G, we also show strong results on datasets with 500 to 2,000 features, and provide more details on how we normalize the performance of each model across datasets.

G.2 Measuring TabPFN-2.5 Training and Inference Speed

Figure 11 shows how TabPFN-2.5 classification speed scales with training set size, when using one or four GPUs, as we vary the number of rows and columns in the dataset. The time measured includes both the time to process the training rows (equivalent to the combination of "training" a classical ML model) and "prediction" time on test rows. We can observe the expected scaling in $\mathcal{O}(r^2 \min(c, 500) + r \min(c, 500)^2)$, where r is the number of rows and c is the number of

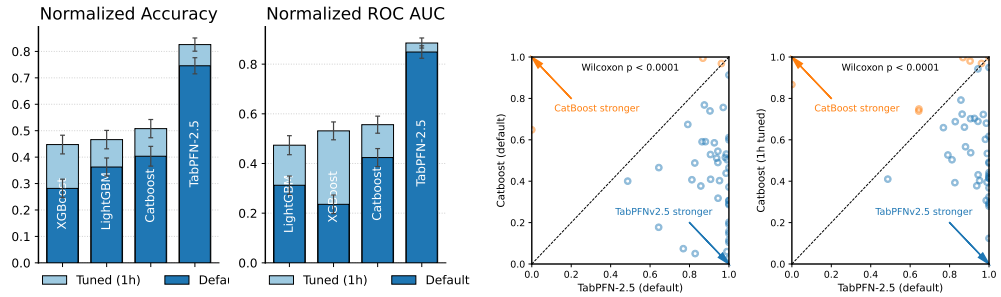


Figure 9: Results from our internal benchmark on **classification datasets with up to 50,000 data points**. More details on the normalization is available in Appendix G. In the scatter plots (right), each point represents a different dataset from our internal benchmark, and the axes measure the normalized performance of TabPFN-2.5 and CatBoost (either default or tuned for 1 hour) on this dataset.

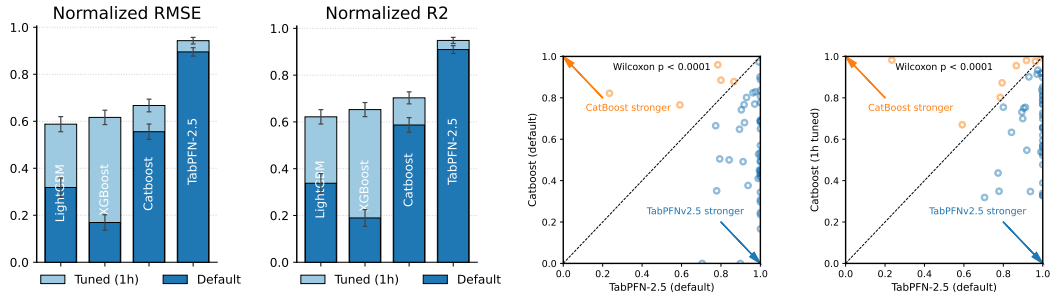
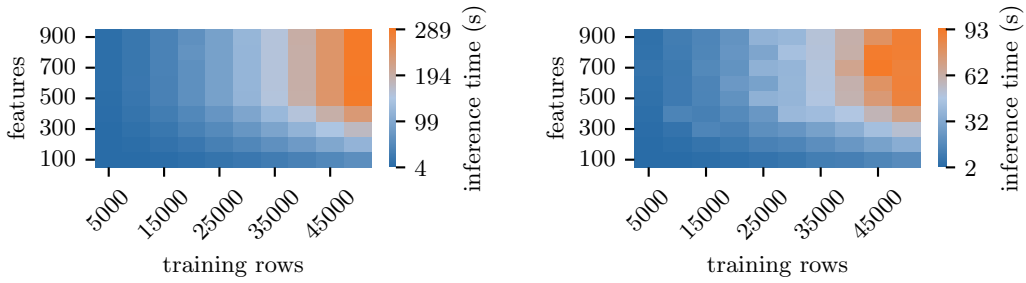


Figure 10: Results from our internal benchmark on **regression datasets with up to 50,000 data points**. More details on the normalization is available in Appendix G. In the scatter plots (right), each point represent a different dataset from our internal benchmark, and the axis measure the normalized performance of TabPFN-2.5 and CatBoost (either default or tuned for 1 hour) on this dataset

columns, due to dual attention over rows and capped per-estimator feature subsampling at 500 features. Section J contains results for regression, performance on common models of GPU, for reference, and a measurement of the speedup from TabPFNv2. The inference speed reported here reflects the latency of the full in-context learning model.

G.3 Fast Inference with TabPFN-2.5-as-MLP

To improve deployment flexibility, we developed a proprietary distillation engine that, given a training data set, outputs a multi-layer perceptron (TabPFN-2.5-as-MLP) or tree ensemble classifier



(a) one H100 GPU

(b) four H100 GPUs

Figure 11: Time taken, in seconds, to fit TabPFN-2.5 classification models on various training set sizes, and then make predictions on 500 test rows. Figure 17 in Section E reports results for regression, alongside performance on A100 and T4 GPUs.

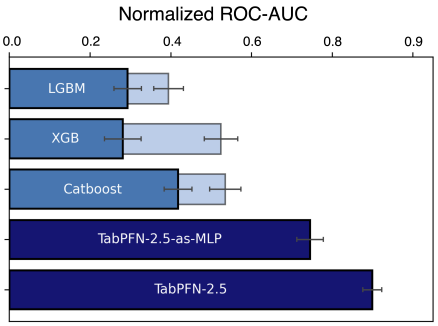


Figure 12: **TabPFN-as-MLP still outperforms tree-based models** while having much faster inference speed than TabPFN. For baseline, light blue represents performance when tuned for 1 hour, and darker blue default performance. For TabPFN, we report default performance.

(TabPFN-2.5-as-TreeEns) whose performance is close to the one of TabPFN on this benchmark (see Figure 12). In contrast to TabPFN, this resulting MLP or tree ensemble classifier is dataset-specific, does not perform in-context learning, takes as input a single data point, and has very low latency and memory footprint for making predictions. It can also be seamlessly integrated into existing production pipelines, including those constrained by latency, interpretability, or regulatory requirements that hinder a change in the class of models being deployed. This increases TabPFN-2.5's practical use in real-world decision systems. Other types of models could easily be supported.

We benchmark TabPFN-2.5-as-MLP against tuned LightGBM, XGBoost, and CatBoost models, as well as the standard TabPFN-2.5 model, on our curated collection of internal open source datasets with less than 10k data points. Figure 12 illustrates representative test-split performance. Empirically, TabPFN-2.5-as-MLP offers competitive accuracy while reducing inference cost, making it attractive for high-throughput or resource-constrained deployment scenarios.

G.4 Details on the normalization

For benchmarking, we normalize scores per dataset to enable averaging and clearer comparison across datasets, ensuring that differences in dataset difficulty do not bias comparisons. For each dataset, we linearly scale scores between 0 (worse model on this dataset) and 1 (best model). For each model, the default and tuned versions are considered as two different models for the normalization. Bar heights show the mean normalized performance, and error bars denote the standard error of the mean (SEM) across datasets, reflecting uncertainty from dataset variability.

G.5 Additional results on many features

In Figure 13, we show results on an internal set of datasets containing from 500 to 2,000 features showing strong default performance.

H Results with Tuned Decision Thresholds

Starting with TabPFN-2.5, our framework supports tuning the decision threshold to optimize for specific metrics. Figure 14 quantifies the performance gains that this procedure can yield, illustrating substantial improvement in F1-score for several imbalanced datasets when tuning the threshold.

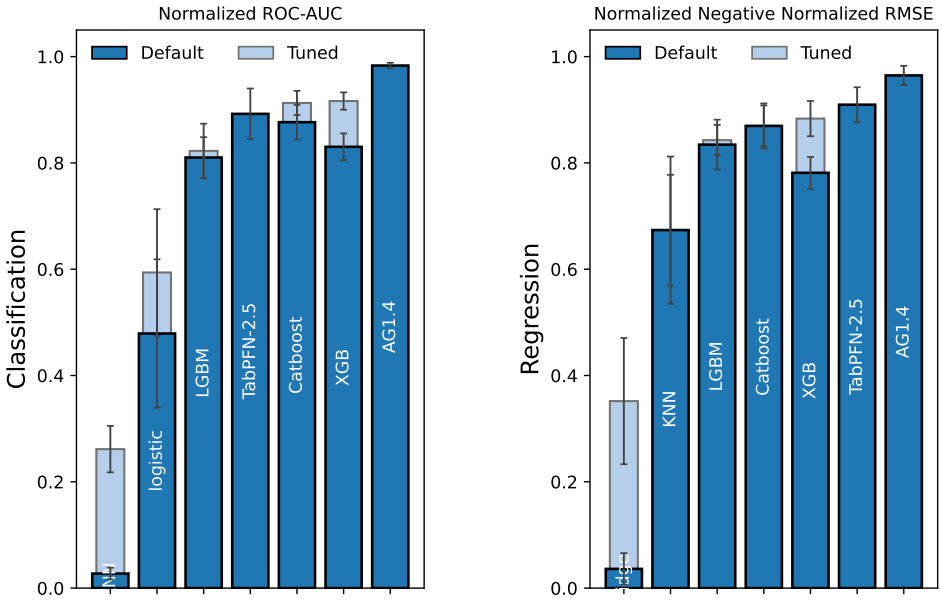


Figure 13: **TabPFN-2.5 default performs well up to 2,000 features**. In our internal benchmark on datasets from 500 features to 2,000 features, we can see that for both classification (left) and regression (right), the default TabPFN-2.5 outperforms any other default model and is better than any tuned single model for regression.

I TabPFN for Causal Inference

RealCause Benchmark. To systematically evaluate TabPFN's potential as a causal estimator, we leverage the RealCause benchmark [77], a semi-synthetic benchmark which begins with real-world randomized control trial (RCT) data and synthetically creates observable confounding effects.[3] We measure the Precision in Estimating Heterogeneous Effects (PEHE), which corresponds to the root-

[3]Descriptions of the ACIC-2016, IHDP, and Lalonde-PSID and Lalonde-CPS datasets are provided in Appendix Table 2.

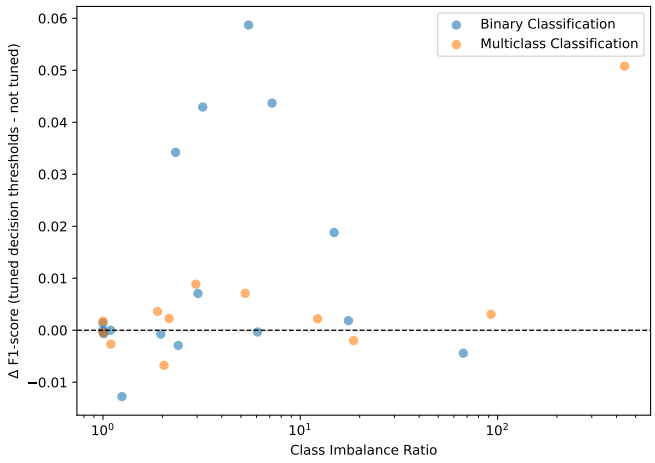


Figure 14: **F1-score sometimes improves substantially by decision threshold tuning.** The plot shows the difference in F1-score (macro) between a model with an optimized decision threshold and the same model using a default (untuned) threshold. This demonstrates the effectiveness of the tuning procedure for metric-specific optimization.

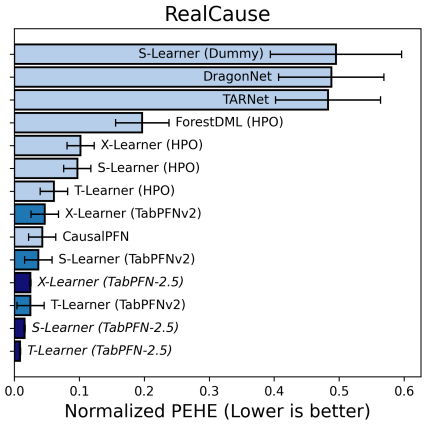


Figure 15: PFN-based CATE estimators dominate RealCause, outperforming specialized tree- and deep-learning-based methods for causal inference. Choice of propensity and outcome model is important for CATE estimation.

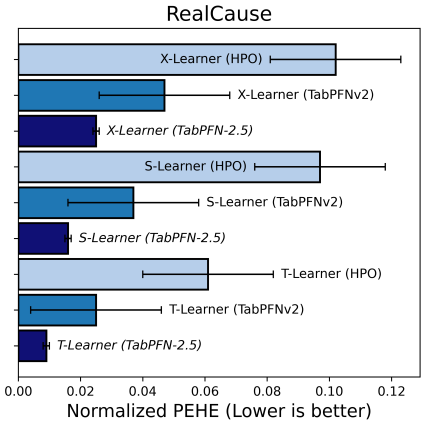


Figure 16: Improvements in base model predictive performance transfer to improved performance in CATE estimation. Our new model, TabPFN-2.5, is the strongest choice of base model for all meta-learners.

mean-squared error between predicted and RealCause's ground-truth CATE values[4]. In Figure 15, we show that PFN-based methods for CATE-estimation dominate the leaderboard, occupying the first seven positions. TabPFN-2.5 applied as a T-Learner, a simple two-model approach that fits a separate model to the treatment and control observations, achieves the strongest overall performance, outperforming specialized tree- and deep-learning-based methods [115]. We also observe in Figure 16 that for each of our three meta-learners, TabPFN-2.5 performs better out-of-the-box than TabPFNv2 and HPO[5]. This result shows that improvements in base model predictive performance transfer to the problem of causal inference.

Foundation Models for Causal Inference. While we show strong results in unconfounded settings, real-world causal inference often involves imperfect data and latent confounders. A growing line of work aims to pre-train PFNs explicitly for causal reasoning—for example, predicting interventional outcomes or learning causal structures directly [11, 34, 73, 95, 100]. We view this as one of the most exciting frontiers for foundation models: extending TabPFN's reasoning from predicting *what is* to inferring *what would happen if*, and ultimately, *understanding why*.

J Supplementary Inference Time Details

Figure 17 shows the inference latency you can expect for three common models of GPUs. Figure 18 shows that the time scales linearly with the number of test rows. Figure 19 compares the fit + training time of TabPFN-2.5 vs TabPFNv2, showing that TabPFN-2.5 is significantly faster, showing between 1x and 2.3x speedup depending on the dataset size.

K Data Contamination and Deduplication for Real-TabPFN-2.5

To ensure fair evaluation and eliminate data contamination, we implemented an enhanced multi-tiered deduplication and filtering pipeline for Real-TabPFN-2.5. While based on the methodology used for Real-TabPFN [45], the process was extended to deduplicate the training datasets against all internal benchmarks, our curated in-house validation suite, and the public TabArena benchmark [40]. Our deduplication procedure combines automated cross-referencing of dataset identifiers, feature schemas, and row- and column-level hashes with manual metadata inspection to ensure that no training dataset

[4]For a description of the CATE estimation task and common estimators, please refer to Appendix L.

[5]Hyperparameter optimization is run for 60 seconds on an H100 per propensity and outcome model using FLAML [116].

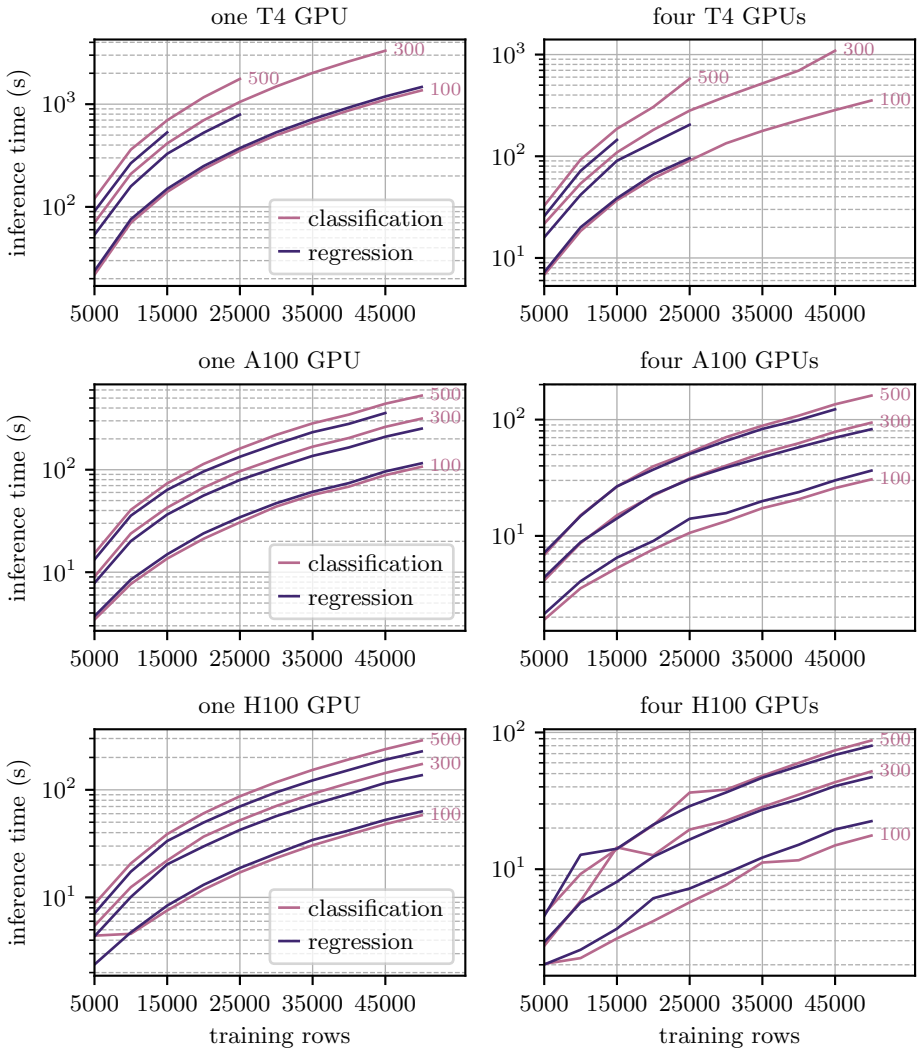


Figure 17: Time taken, in seconds, to train TabPFN-2.5 models on various training set sizes, and then make predictions on 500 test rows, using three common models of NVIDIA GPU: T4 15GB, A100 SXM 40GB, H100 SXM 80GB. Performance is shown for 100, 300, and 500 features. Datasets with more than 500 features have the same performance as datasets with 500, as each estimator will subsample to 500 features. Incomplete lines indicate that the GPU had insufficient memory for that dataset size.

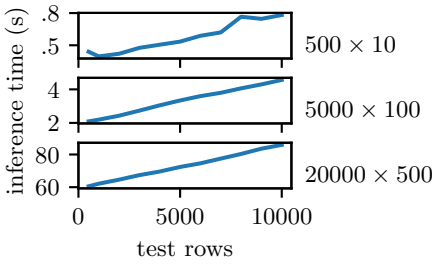


Figure 18: The time taken by TabPFN-2.5 to train and predict scales linearly in the test set size, shown here for a classification model trained on datasets of 500 rows × 10 features, 5,000 rows × 100 features, and 20,000 rows × 500 features. Measured on one H100 GPU.

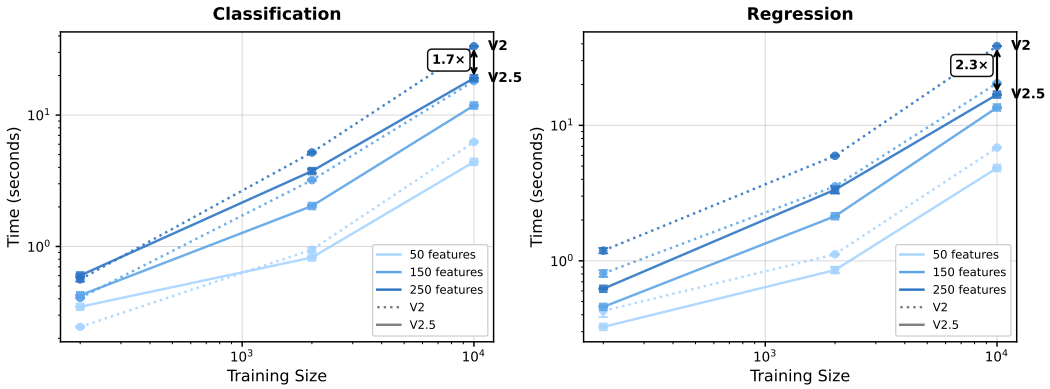


Figure 19: **TabPFN-2.5 is significantly faster than TabPFNv2**. Comparison of the time taken to fit + predict TabPFN-2.5 vs TabPFNv2 on different number of rows and features. Measured for 100 test points on 1 H100, using the same number of estimators (8). Note that this is measured using the v2 and v2.5 versions available on the latest release of the TabPFN package, and thus is on top of the performance improvements since the original release of TabPFNv2.

overlaps with, or is derived from, any evaluation dataset. Datasets failing these criteria were excluded from the final training corpus.

K.1 Training Datasets

The following table lists the datasets curated for fine-tuning, along with their sources and access links.

Name	Source
artificial-characters	OpenML
BNG(breast-w)	OpenML
BNG(tic-tac-toe)	OpenML
connect_4	OpenML
eeg-eye-state	OpenML
Employee-Turnover-at-TECHCO	OpenML
eye_movements	OpenML
FOREX_eurpln-hour-High	OpenML
gas-drift	OpenML
higgs	OpenML
Intersectional-Bias-Assessment-(Training-Data)	OpenML
law-school-admission-binary	OpenML
Medical-Appointment	OpenML
microaggregation2	OpenML
fried	OpenML
mushroom	OpenML
NewspaperChurn	OpenML
nursery	OpenML
WBCAtt	OpenML
Internet Firewall Data	OpenML
aam_avaliacao_dataset	Kaggle
Air Traffic Data	Kaggle
ansible-defects-prediction	Kaggle
AV Healthcare Analytics II	Kaggle
Candidate Selection	Kaggle
Cardio Disease	Kaggle
Classification - Crop Damages in India (2015-2019)	Kaggle
CSGO Round Winner Classification	Kaggle

Name	Source
Flower Type Prediction Machine Hack	Kaggle
Horse Racing - Tipster Bets	Kaggle
How severe the accident could be	Kaggle
hr-comma-sep	Kaggle
ip-network-traffic-flows-labeled-with-87-apps	Kaggle
Janatahack cross-sell prediction	Kaggle
L&T Vehicle Loan Default Prediction	Kaggle
League of Legends Diamond Games (First 15 Minutes)	Kaggle
Richter's Predictor Modeling Earthquake Damage	Kaggle
Server Logs - Suspicious	Kaggle
Sloan Digital Sky Survey DR14	Kaggle
Sloan Digital Sky Survey DR16	Kaggle
Term Deposit Prediction Data Set	Kaggle
trajectory-based-ship-classification	Kaggle
Travel Insurance	Kaggle

L Details on Causal Inference Results

Causal Inference Most real-world decision problems ultimately hinge on causal questions—understanding what would happen if we intervened, rather than merely observing correlations. Estimating Conditional Average Treatment Effects (CATEs) is one of the central ways to answer these "what-if" questions: how would an individual's outcome change if a treatment were applied versus withheld?

Unconfounded Settings. Many causal inference methods require *unconfoundedness*, which broadly states that there are no features not included in the dataset that influence both the *treatment* variable and the outcome [96]. While recent studies have begun to challenge the validity and verifiability of this assumption [56, 95], there are presently a wide variety of causal inference methods designed for the unconfounded setting [30, 83].

Importance of Base Model. Recent empirical findings have shown that when unconfoundedness holds, CATE estimation can be framed as an AutoML problem [111], as many CATE estimators require a choice of classification or regression model to approximate the likelihood (propensity) of a treatment and an outcome given an individual's features. Parallel studies [127, 95] have shown that TabPFN is an especially strong choice for meta-learners such as the X-, T-, and S-Learner [61], hypothesizing that its strong performance in tabular prediction transfers to the problem of causal inference.

Table 2: Description of causal inference datasets in the RealCause benchmark.

Characteristic	ACIC-2016	IHDP	Lalonde-CPS	Lalonde-PSID
Realizations	10	100	100	100
Samples	4,802	747	16,177	2,675
Features	58	25	8	8