SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

In-Context Prior-Data-Fitted Networks for Robust Stroke Outcome Prediction in Tabular Data

Dayeong An, PhD, Northwestern University; Chaochao Zhou, PhD; Syed Faruqui, PhD;
Ramez Abdalla, MD, PhD; Ali Shaibani, MD; Sameer Ansari, MD, PhD; Donald Cantrell, MD, PhD

Introduction/Background

Mechanical Thrombectomy (MT) is the established standard-of-care for the interventional treatment of Acute Ischemic Stroke (AIS), significantly improving patient outcomes. Precise prognostication of clinical outcomes following MT is crucial for informed, personalized postoperative care. However, conventional machine learning (ML) models require dataset specific retraining and hyperparameter tuning, limiting their suitability for real time clinical application. Tabular Prior Data Fitted Network (TabPFN) has emerged as a promising in-context learning model for tabular data, uniquely capable of accurately and instantaneously predicting outcomes without any additional retraining or hyperparameter tuning.

Methods/Intervention

The NeuroVascular Quality Initiative-Quality Outcomes Database (NVQI-QOD) registry is a multi-center initiative that documents neurointerventional procedures performed at 34 US centers between January 2014 and December 2023. The dataset contains 10682 entries and 425 variables. The feature variables were grouped into three categories: Preop (data available prior to MT), Postop (data after MT), and DC (discharge data). TabPFN was employed to predict 90-day mRS severity grouped into binary classes (0-2 vs. 3-6) using classification. Performance was compared against traditional ML models, including hyperparameter-optimized XGBoost and CatBoost. Additionally, TabPFN’s performance was validated using publicly available clinical datasets.

Results/Outcome

TabPFN consistently outperformed traditional ML models across all evaluation stages and metrics. At the Preop stage, TabPFN achieves an AUROC of 0.79, improving to 0.87 at Postop and 0.90 at DC. Additionally, TabPFN leads in prediction, recall, and F1-score, indicating its ability to make accurate and balanced predictions across diverse clinical outcomes. On six publicly available clinical datasets, TabPFN also consistently demonstrated superior performance compared to ML models, further validating to its robustness and generalizability.

Conclusion

Our finding highlight TabPFN as an exceptionally powerful tool for accurate, rapid, and real-time prediction of clinical outcomes following MT. By harnessing synthetic data pretraining and efficient in-context learning, TabPFN circumvents the typical need for data-specific retraining and parameter tuning, establishing a new state-of-the-art in stroke prognostication.

Statement of Impact

By delivering millisecond scale predictions directly from tabular data, this framework can be integrated into acute stroke workflows to personalize intervention strategies, reduce decision latency, and ultimately improve patient outcomes.

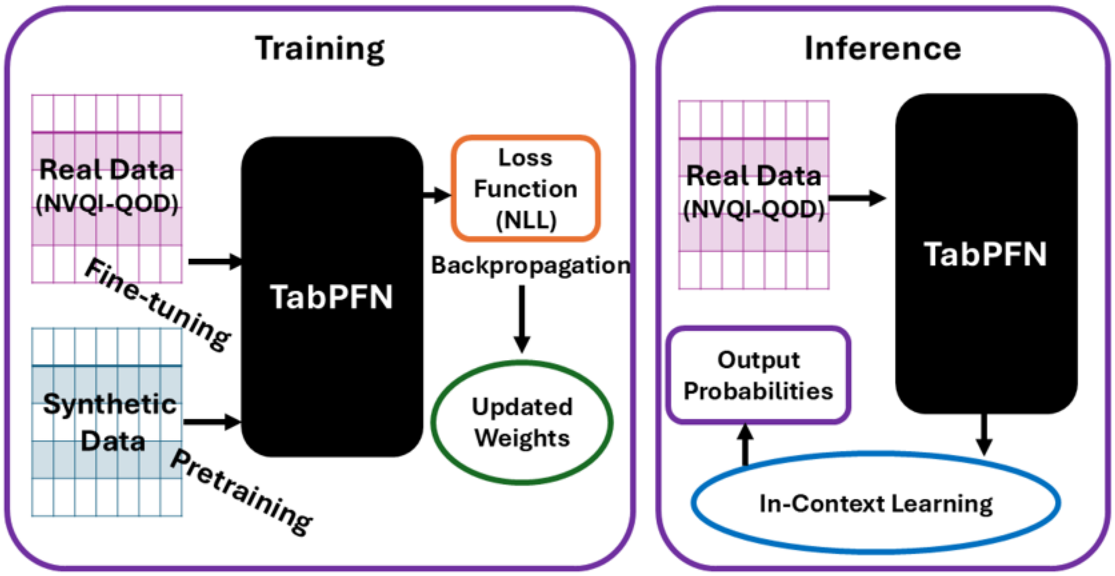

Fig. 1. Workflow of TabPFN. The diagram illustrates the process where the NVQI-QOD dataset is input to the TabPFN mode. The model, pretrained on synthetic data, utilizes in-context learning to generate output predictions directly from the input data.

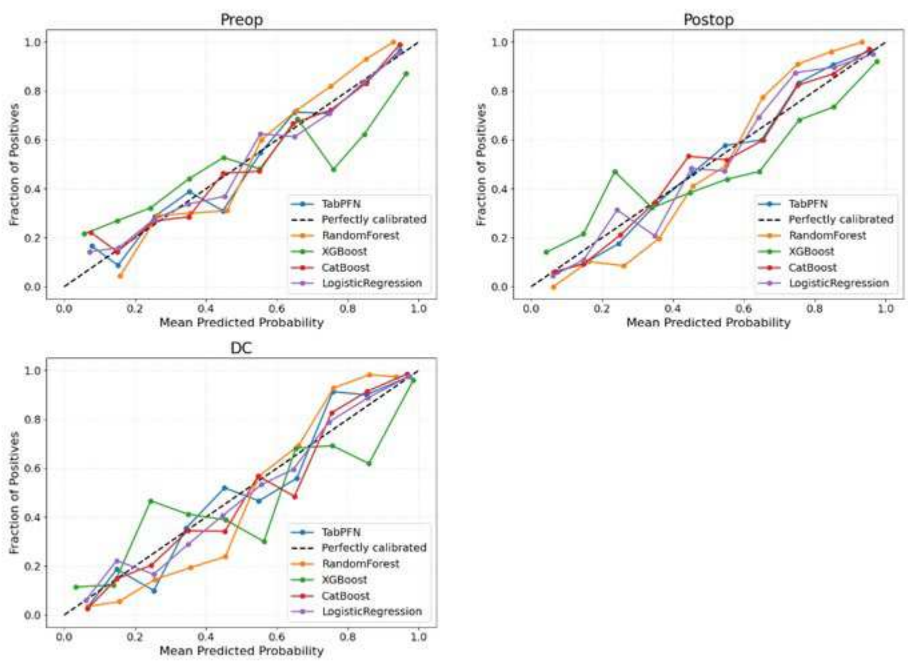

Fig. 2. Calibration plots for outcome prediction at different stages. The calibration curves display the relationship between the mean predicted probability and the fraction of positives for each model – TabPFN, Random Forest, XGBoost, CatBoost, and Logistic Regression – at the Preop, Postop, and DC stages. The dashed line represents the ideal calibration, where predicted probabilities perfectly match the true outcomes.

	ROC AUC	Precision	Recall	F1-Score	NLL	Accuracy
			Preop			
TabPFN	**0.791**	0.713	**0.712**	**0.713**	10.126	**0.719**
Random Forest	0.776	0.700	0.695	0.697	10.598	0.706
XGBoost	0.760	0.692	0.690	0.691	10.879	0.698
CatBoost	0.788	**0.714**	0.710	0.712	**10.116**	**0.719**
Logistic Regression	0.784	0.708	0.705	0.706	10.317	0.714
KNN	0.697	0.648	0.651	0.645	12.748	0.646
SVC	0.776	0.699	0.692	0.694	10.648	0.705
Naïve Bayes	0.732	0.661	0.660	0.647	12.728	0.647
Decision Tree	0.606	0.605	0.606	0.605	13.963	0.613
			Postop			
TabPFN	**0.870**	0.785	0.783	0.784	7.615	0.789
Random Forest	0.863	0.783	0.777	0.780	7.715	0.786
XGBoost	0.848	0.770	0.768	0.769	8.147	0.774
CatBoost	0.869	**0.787**	**0.785**	**0.786**	**7.544**	**0.791**
Logistic Regression	0.866	0.785	**0.785**	0.785	7.605	0.789
KNN	0.755	0.697	0.700	0.691	11.120	0.691
SVC	0.861	0.774	0.772	0.773	7.996	0.778
Naïve Bayes	0.790	0.682	0.611	0.536	15.701	0.564
Decision Tree	0.693	0.692	0.693	0.693	10.889	0.698
			DC			
TabPFN	**0.900**	0.808	**0.810**	**0.809**	6.771	0.812
Random Forest	0.893	**0.810**	0.808	**0.809**	**6.731**	**0.813**
XGBoost	0.885	0.802	0.802	0.802	6.992	0.806
CatBoost	0.898	0.808	0.809	0.808	6.781	0.812
Logistic Regression	0.893	0.803	0.804	0.804	6.952	0.807
KNN	0.790	0.721	0.725	0.715	10.247	0.716
SVC	0.890	0.798	0.797	0.797	7.142	0.802
Naïve Bayes	0.842	0.727	0.709	0.681	11.412	0.683
Decision Tree	0.732	0.732	0.732	0.732	9.463	0.737

Table 1. Model performance comparison across different stages clinical care. The table shows the performance of several machine learning models – TabPFN, Random Forest, XGBoost, CatBoost, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Naïve Bayes, and Decision Tree for mRS classification across Preoperative (Preop), Postoperative (Postop), and Discharge (DC) feature sets. The performance metrics evaluated include ROC AUC, precision, recall, F1-score, negative log-likelihood (NLL), and accuracy. Best values are highlighted in bold.

Keywords

In-Context Learning; TabPFN; Stroke Prognosis; mechanical thrombectomy; Tabular Clinical Data; Bayesian Inference