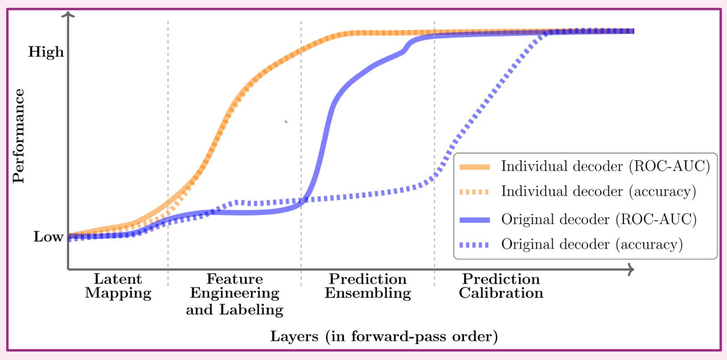
Is One Layer Enough? Understanding Inference Dynamics in Tabular Foundation Models

Amir Balef[1,2,3], Mykhailo Koshil[2,3] and Katharina Eggensperger[2,3]
[1]University of Tübingen, [2]TU Dortmund University
[3]Lamarr Institute for Machine Learning and Artificial Intelligence

Too many plots; Didn't Read (TL;DR)

TFMs achieve strong performance on tabular tasks, but their **inference mechanisms are poorly understood.**
➔ First large-scale mechanistic study of 6 TFMs using layer-wise analyses across **6 experiments** ⚙️, each with **key takeaways** 📝.

❓ RQ1 How does inference unfold across depth in TFMs?



📝 TFMs and LLMs share clear inference dynamics, unfolding across **4 distinct stages**.

❓ RQ2 How do the inference dynamics of TFMs compare to those observed in LLMs?

➔ Rigid Hierarchy.

TFMs are significantly more sensitive to layer swapping than LLMs

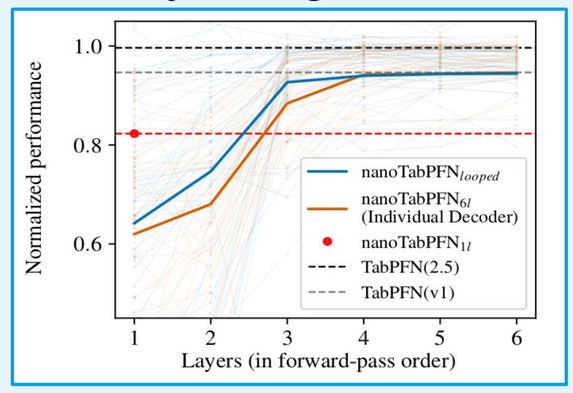
➔ Early layers.

Redundancy dominates middle layers of TFMs; confident predictions emerge earlier than in LLMs.

➔ Final Layer Importance.

Unlike the critical "residual sharpening" in LLMs, the TFM final layer is less vital for performance.

⚡ POC Is one layer enough?



📝 By repeating one layer six times, we achieves performance comparable to a standard six-layer model using **only 20%** of the parameters.

📉 Architectural Suggestions

- **Using stronger encoders** (like in TabICL or LimiX-2M) to create richer initial embeddings
- Reducing model depth by **removing redundant middle layers.**
- Using **recurrently applied shared layers** (looped architectures) to improve parameter efficiency

⚠️ Limitations & 🚀 Future Work

- Prior design: **TabICL priors** may limit expressivity
- Scaling: **Looping not yet tested on larger models**
- Evaluation: Two benchmarks, no ensembling
- **Next: Neuron-level insights** and **larger recurrent models**

📚 References

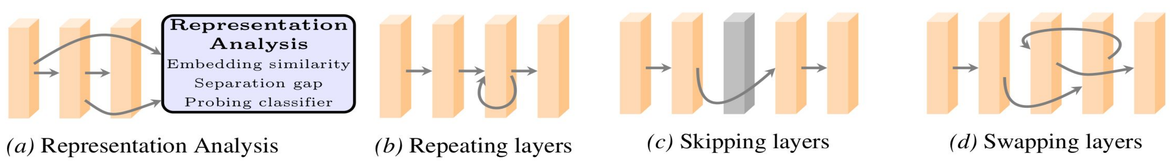
[1] Lad et al. (2025). Remarkable robustness of LLMs: Stages of inference? NeurIPS.
[2] Pfefferle et al. (2025). nano-TabPFN: A Lightweight and Educational Reimplementation of TabPFN. EurIPS Workshop on AI for Tabular Data.
[3] Grinsztajn et al. (2025). TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models. arXiv.
[4] Balef et al. (2025). Towards Understanding Layer Contributions in Tabular In-Context Learning Models. EurIPS Workshop on AI for Tabular Data.

ICML 2026

Experiments

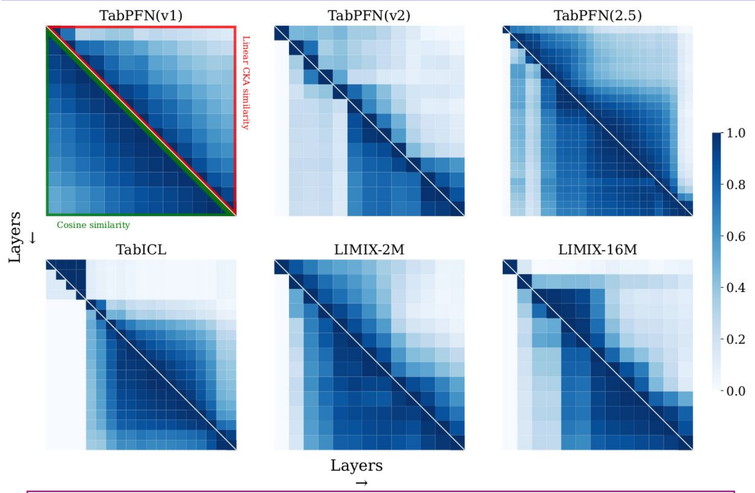
💡 Methodology

- Predictions across inference depth
  - Representation analysis
  - Layer ablation
- Experiments on TabArena and PMLBmini benchmarks



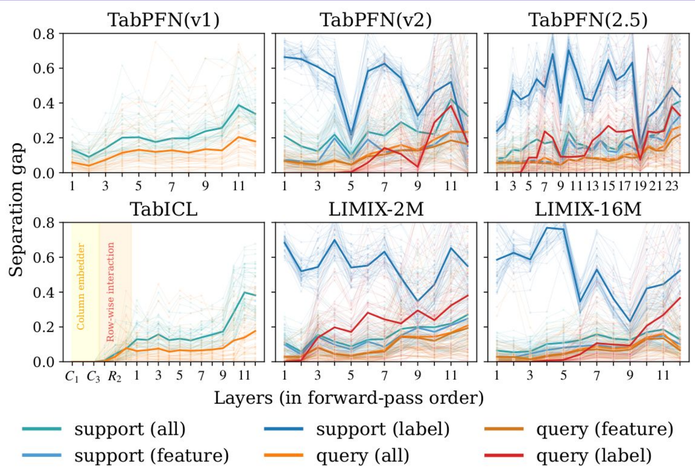
SCAN ME!

⚙️ Exp 1. Embedding Similarity



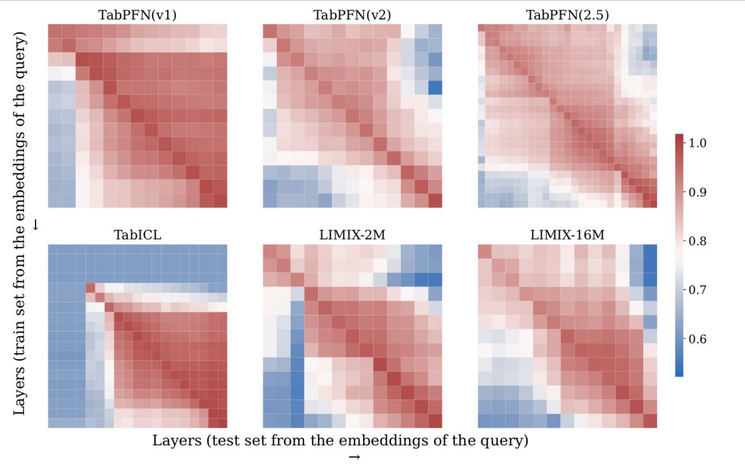
📝 Takeaway 1. TFMs often **form blocks** in which the embeddings remain similar.

⚙️ Exp 2. Separation Gap



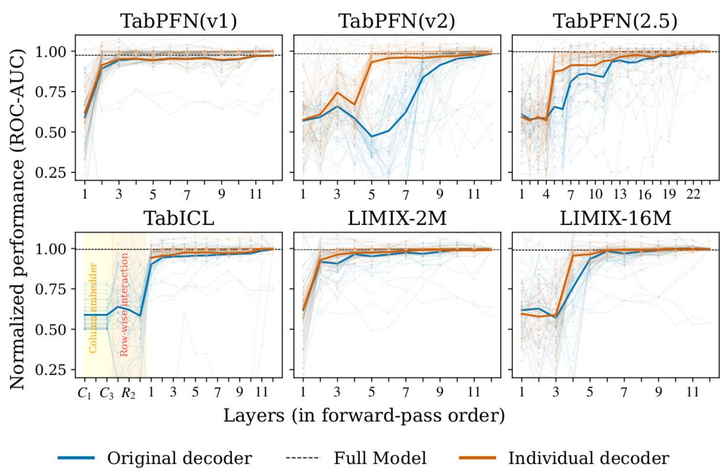
📝 Takeaway 2. TFMs **incrementally** increase the distance between samples from different classes.

⚙️ Exp 3. Probing Classifiers



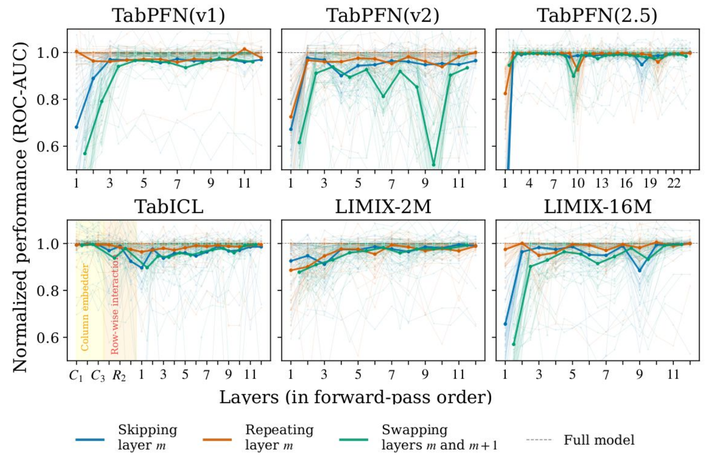
📝 Takeaway 3. Each layer enriches the representation with **new features while preserving earlier ones**.

⚙️ Exp 4. "Tabular" Logit Lens



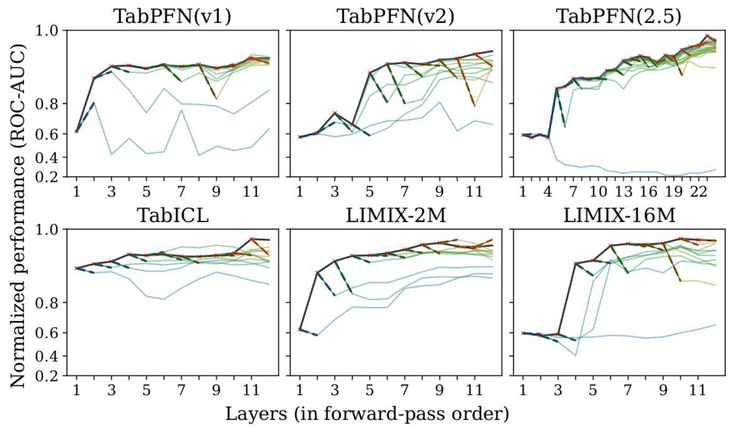
📝 Takeaway 4. Early layers already form representations for reliable prediction, though **not necessarily aligned with the original decoder**.

⚙️ Exp 5. Layer Ablation



📝 Takeaway 5. Early layers contribute the most, while laterlayers perform **iterative refinement** of the representation.

⚙️ Exp 6. Self-Repair



📝 Takeaway 6. Self-repair generally occurs after layer ablations, **except for the first layer**.

Partner institutions:
technische universität dortmund
Fraunhofer IAIS
UNIVERSITÄT BONN
Fraunhofer IML

Institutionally funded by:
Bundesministerium für Forschung, Technologie und Raumfahrt
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen