STT-LLM: Structural-Temporal Tokenization for Adapting LLMs to Longitudinal Clinical Profiles

Maxx Richard Rahman[1,2], Mostafa Hammouda[1,2], Wolfgang Maass[1,2]

[1]German Research Centre for Artificial Intelligence (DFKI), Germany
[2]Saarland University, Germany

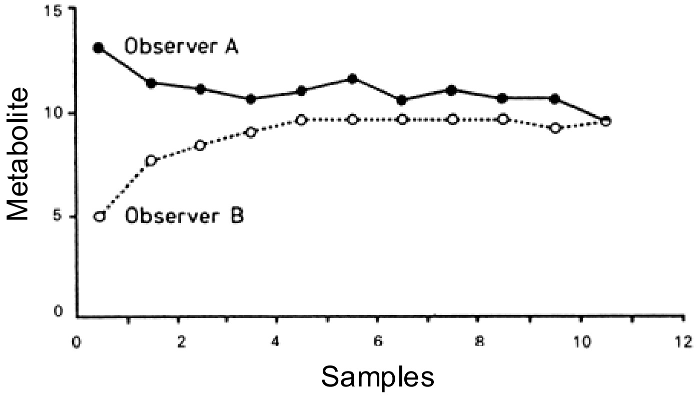
The 43rd International Conference on Machine Learning (ICML 2026)
Seoul, South Korea

Introduction

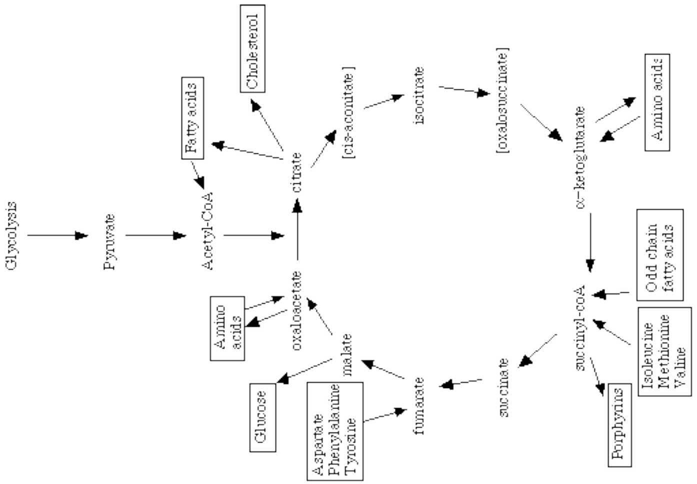
- Longitudinal clinical profiles capture **measurements across multiple time points**. [*Fitzmaurice et al., 2011*]
- Such data are widely used in **clinical monitoring, precision medicine**, and **anti-doping**. [*Singer & Willett, 2009*]
- Recent advances in LLMs show **promise for modelling** sequential data. [*Li et al., 2024*]
- **Challenges:**
  - Complexity of longitudinal data
  - Modality gaps for LLM training
  - Data availability

Aim: Adapt LLMs to understand the **domain structure** & **temporal dynamics** in longitudinal data.

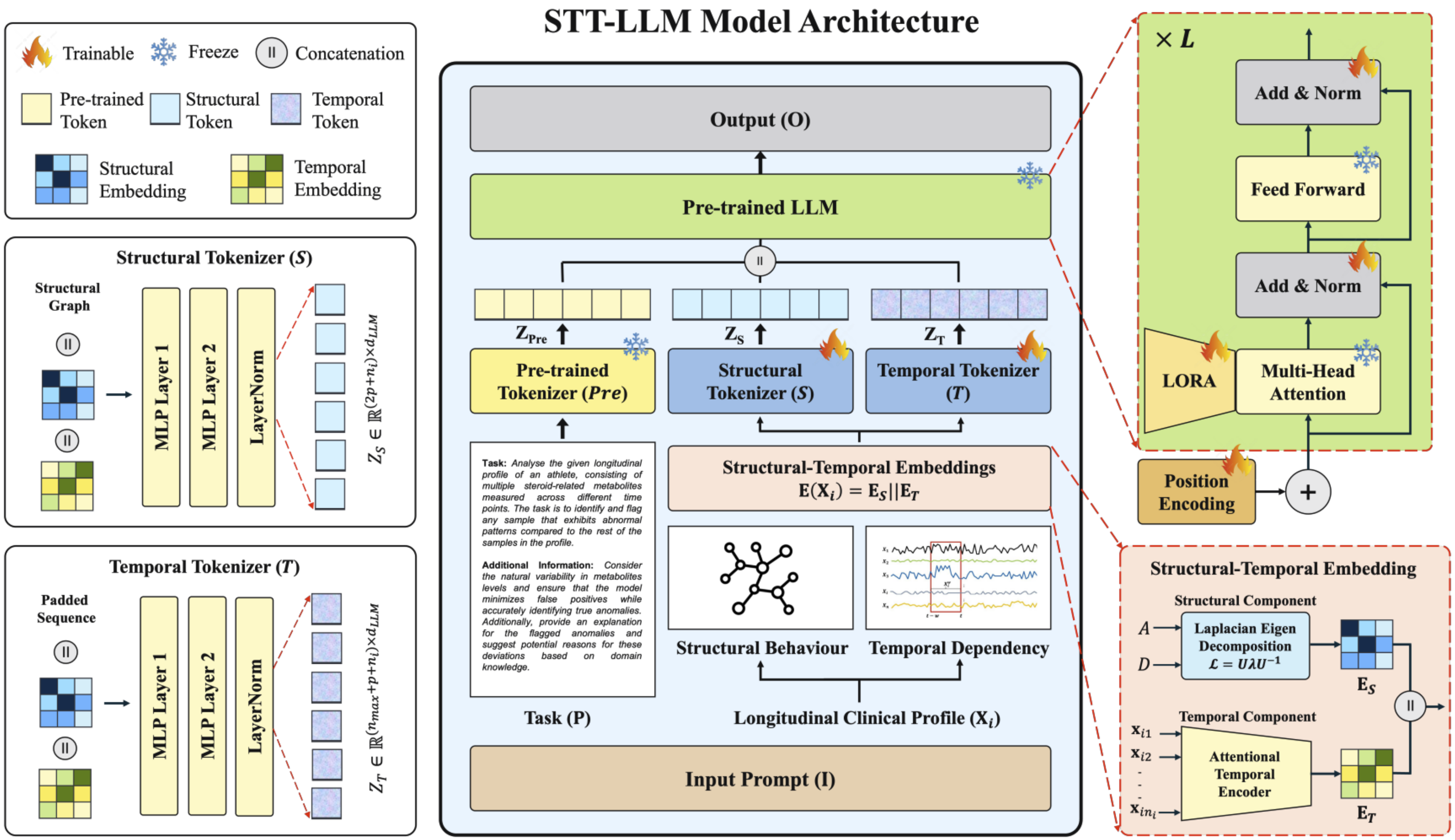
Longitudinal Profile



Metabolic Pathway



STT-LLM: Structural-Temporal Tokenization for Large Language Model



Key contributions:

- **Structural-Temporal Embeddings *E(X)*** to integrate both temporal patterns & structural relationships across metabolites.
- **Structural Tokenizer *S*** to capture topological relationships between metabolites
- **Temporal Tokenizer *T*** to capture sequential dependencies and dynamic pattern over time

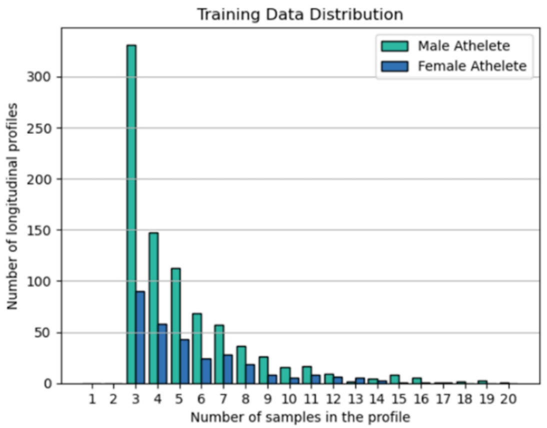
Experiments
Datasets

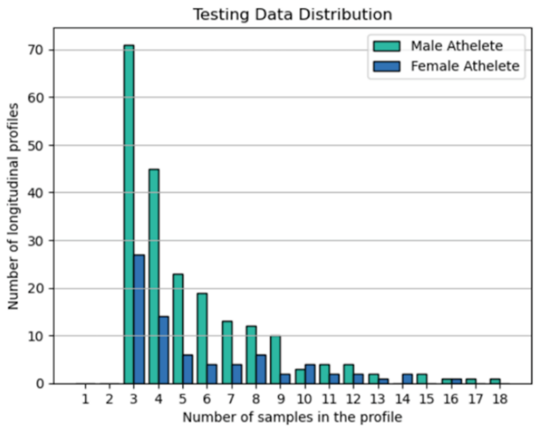
Real-world athlete datasets consisting longitudinal steroid profiles

- 6 steroid metabolites
- <20% anomalous profiles
- Training Dataset: 4000 male & female

Table 1. Summary statistics of all the datasets.

Datasets	Gender	# Profiles	# Samples	Length n_i
Steroid-M	Male	755	4214	3-20
Steroid-F	Female	375	2307	3-20
Steroid-M_{lim}	Male	737	1474	2
Steroid-F_{lim}	Female	293	586	2



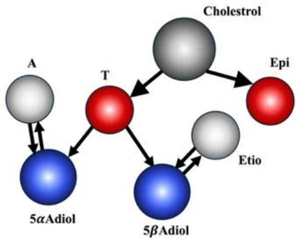


Longitudinal Steroid Profile

Parameter	Description
INC/OOC	Collection in/out competition
T	Testosterone
E	Epitestosterone
Etio	Etiocholanolone
A	Androsterone
5aAdiol	5α-androstane-3α, 17β-diol
5bAdiol	5β-androstane-3α, 17β-diol
T/E	Ratio
A/Etio	Ratio
A/T	Ratio
5aAdiol/5bAdiol	Ratio
5aAdiol/E	Ratio

Primary Parameters

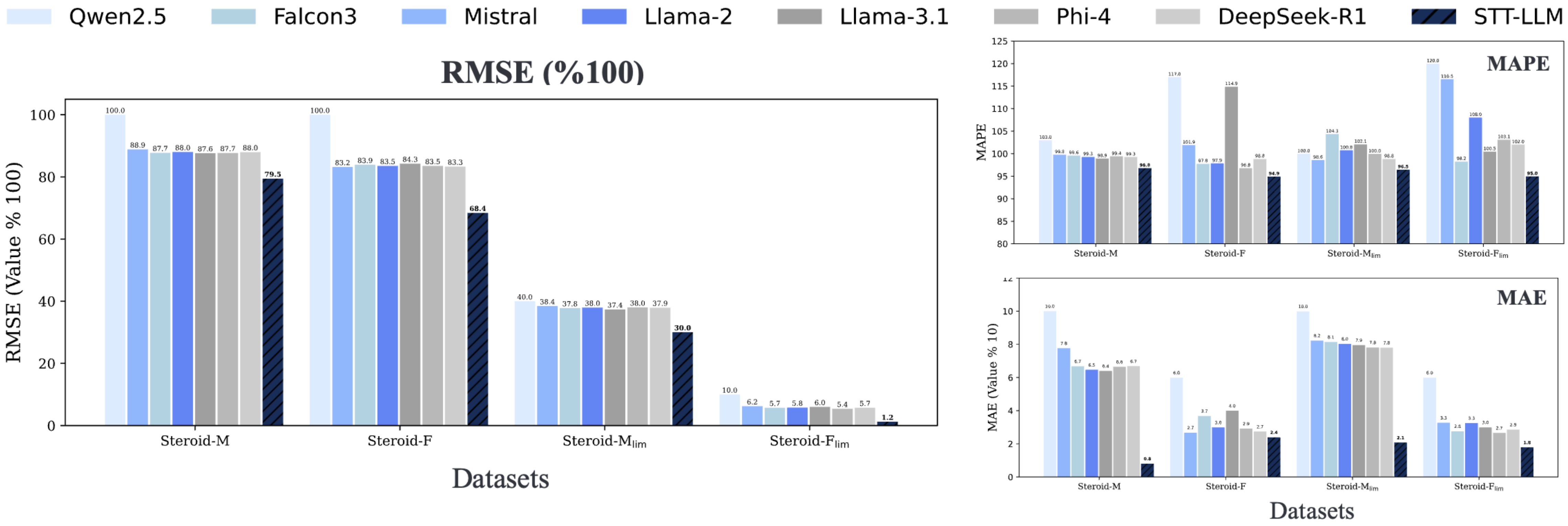
Derived Parameters



Results

Sequence Prediction (Zero-shot setting)

STT-LLM consistently **achieves the lowest error rates** across all datasets and metrics, outperforming baseline LLMs in zero-shot sequence prediction.

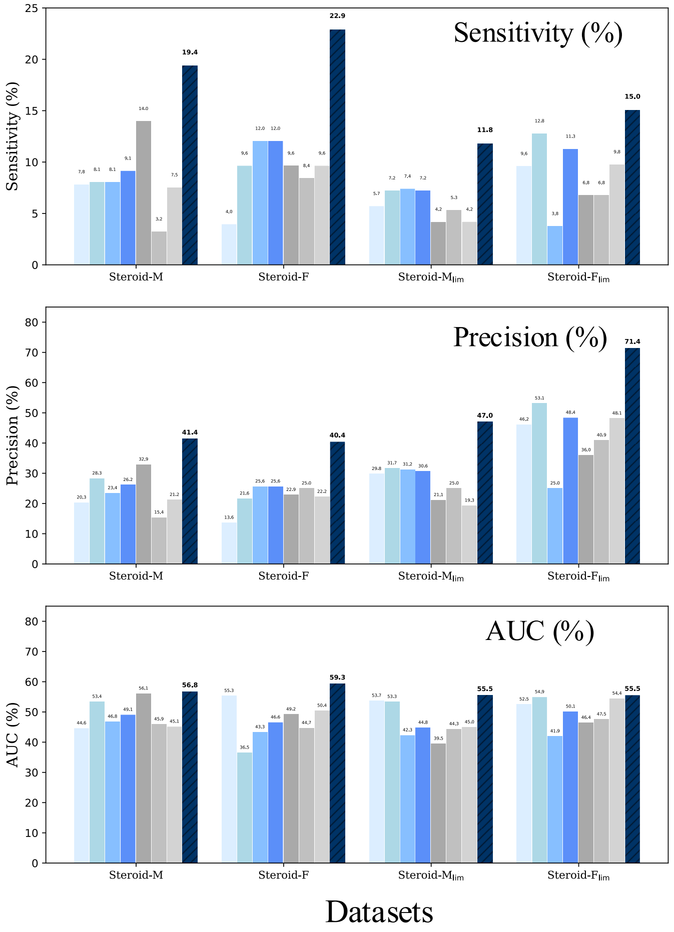


Results

Anomaly Detection (Zero-shot setting)

STT-LLM shows **strong gains** in sensitivity & precision for global anomaly detection across all datasets.

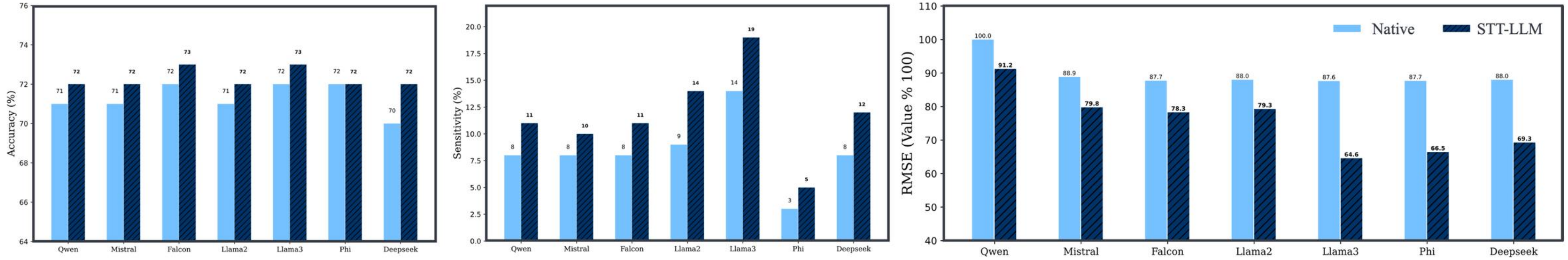
Datasets	Baseline	Local					Global				
		Acc↑	Sens↑	Prec↑	F1↑	AUC↑	Acc↑	Sens↑	Prec↑	F1↑	AUC↑
Steroid-M	Qwen-2.5	**0.96±.01**	0.00±.00	0.00±.00	0.00±.00	0.47±.02	0.71±.02	0.08±.03	0.20±.02	0.11±.02	0.45±.02
	Mistral	0.87±.02	0.05±.01	0.02±.01	0.03±.01	0.43±.02	0.71±.02	0.08±.02	0.23±.03	0.12±.02	0.47±.02
	Falcon-3	0.94±.02	0.01±.00	0.02±.01	0.01±.01	0.46±.02	0.72±.02	0.08±.02	0.28±.03	0.13±.02	0.53±.02
	LLaMA-2	0.90±.02	0.05±.01	0.03±.01	0.04±.01	0.42±.02	0.71±.02	0.09±.02	0.26±.02	0.14±.03	0.49±.02
	LLaMA-3.1	0.87±.02	0.07±.01	0.03±.01	0.05±.01	0.51±.02	0.72±.02	0.14±.02	0.33±.03	0.19±.03	0.56±.02
	Phi-4	0.87±.02	0.08±.01	0.04±.01	0.05±.01	0.50±.02	0.72±.02	0.03±.01	0.15±.02	0.05±.01	0.46±.02
	DeepSeek-R1	0.95±.01	0.02±.01	0.01±.01	0.01±.00	0.39±.02	0.70±.02	0.08±.02	0.21±.02	0.11±.02	0.45±.02
	STT-LLM	0.87±.02	**0.15±.02**	**0.07±.01**	**0.09±.02**	**0.57±.02**	**0.73±.02**	**0.19±.03**	**0.41±.03**	**0.26±.03**	**0.57±.02**
Steroid-F	Qwen-2.5	0.87±.02	0.04±.01	0.02±.01	0.02±.01	0.46±.02	0.73±.02	0.04±.01	0.14±.02	0.06±.01	0.55±.02
	Mistral	**0.96±.01**	0.00±.00	0.00±.00	0.00±.00	0.62±.02	0.73±.02	0.12±.02	0.26±.03	0.16±.02	0.43±.02
	Falcon-3	0.95±.01	0.00±.00	0.00±.00	0.00±.00	0.60±.02	0.72±.02	0.09±.02	0.22±.02	0.13±.02	0.37±.02
	LLaMA-2	0.87±.02	0.06±.01	0.02±.01	0.03±.01	0.55±.02	0.73±.02	0.12±.02	0.26±.02	0.16±.02	0.47±.02
	LLaMA-3.1	0.95±.01	0.01±.00	0.03±.01	0.02±.01	0.57±.02	0.73±.02	0.10±.02	0.23±.03	0.14±.02	0.49±.02
	Phi-4	0.88±.02	0.06±.01	0.02±.01	0.03±.01	0.50±.02	0.74±.02	0.08±.02	0.25±.02	0.13±.02	0.45±.02
	DeepSeek-R1	0.87±.02	0.06±.01	0.02±.01	0.03±.01	0.42±.02	0.73±.02	0.10±.02	0.22±.03	0.13±.02	0.50±.02
	STT-LLM	0.87±.02	**0.08±.01**	**0.03±.01**	**0.05±.01**	0.47±.02	**0.75±.02**	**0.23±.03**	**0.40±.03**	**0.29±.03**	**0.59±.02**
Steroid-M_{lim}	Qwen-2.5	0.86±.02	0.00±.00	0.00±.00	0.00±.00	0.18±.01	0.62±.02	0.06±.02	0.30±.03	0.10±.02	0.54±.02
	Mistral	**0.96±.01**	0.00±.00	0.00±.00	0.00±.00	0.37±.02	0.61±.02	0.07±.02	0.31±.02	0.12±.02	0.42±.02
	Falcon-3	0.88±.02	0.08±.01	0.03±.01	0.05±.01	0.44±.02	0.61±.02	0.07±.02	0.32±.02	0.12±.02	0.53±.02
	LLaMA-2	0.88±.02	0.03±.01	0.01±.01	0.02±.01	0.22±.01	0.61±.02	0.07±.02	0.31±.02	0.12±.02	0.45±.02
	LLaMA-3.1	0.88±.02	0.04±.01	0.02±.01	0.02±.01	0.39±.02	0.60±.02	0.04±.01	0.21±.02	0.07±.02	0.39±.02
	Phi-4	0.89±.02	0.21±.02	0.09±.01	0.12±.02	0.65±.02	0.61±.02	0.05±.01	0.25±.02	0.09±.01	0.44±.02
	DeepSeek-R1	0.87±.02	0.06±.01	0.02±.01	0.03±.01	0.43±.02	0.60±.02	0.04±.01	0.19±.02	0.07±.01	0.45±.02
	STT-LLM	0.88±.02	**0.36±.02**	**0.12±.02**	**0.18±.02**	**0.75±.02**	**0.64±.02**	**0.12±.02**	**0.47±.03**	**0.19±.02**	**0.55±.02**
Steroid-F_{lim}	Qwen-2.5	0.88±.02	0.06±.01	0.06±.01	0.06±.01	0.14±.01	0.54±.02	0.10±.02	0.46±.03	0.16±.02	0.53±.02
	Mistral	0.95±.01	0.01±.00	0.03±.00	0.02±.00	**0.64±.02**	0.51±.02	0.04±.01	0.25±.02	0.07±.02	0.42±.02
	Falcon-3	**0.96±.01**	0.00±.00	0.00±.00	0.00±.00	0.27±.02	0.55±.02	0.13±.02	0.53±.03	0.21±.03	0.55±.02
	LLaMA-2	0.86±.02	0.03±.01	0.01±.01	0.02±.01	0.32±.02	0.54±.02	0.11±.02	0.48±.03	0.18±.02	0.50±.02
	LLaMA-3.1	0.87±.02	0.00±.00	0.00±.00	0.00±.00	0.08±.01	0.52±.02	0.07±.01	0.36±.02	0.11±.01	0.46±.02
	Phi-4	0.87±.02	0.07±.01	0.03±.01	0.04±.01	0.48±.02	0.53±.02	0.07±.01	0.41±.03	0.12±.02	0.48±.02
	DeepSeek-R1	0.86±.02	0.10±.01	0.04±.01	0.06±.01	0.51±.02	0.54±.02	0.10±.02	0.48±.03	0.16±.02	0.54±.02
	STT-LLM	0.87±.02	**0.17±.02**	**0.08±.01**	**0.11±.02**	**0.54±.02**	**0.59±.02**	**0.15±.03**	**0.71±.03**	**0.25±.03**	**0.56±.02**



Results

Overall Performance Comparison

Performance comparison of **LLM-native** and **STT-LLM tokenization** strategies across different general-purpose LLMs



Performance of **non-LLM baselines** compared to STT-LLM on Steroid-M dataset

Non-LLM Baseline	Sequence Prediction			Anomaly Detection (Global)				
	RMSE↓	MAE↓	MAPE↓	Acc↑	Sens↑	Prec↑	F1↑	AUC↑
TabPFN (Grinsztajn et al., 2025)	–	–	–	0.5107	0.5032	0.0124	0.4019	0.4876
MLP-SLAM (Li & Sun, 2024)	–	–	–	0.4928	0.4215	0.0137	0.3284	0.4891
M-GNN (Yuan & Nault, 2025)	–	–	–	0.4716	0.4869	0.0043	0.6292	0.4738
ARIMA (Schaffer et al., 2021)	1675.47	879.96	96.99	–	–	–	–	–
TDDGNN (Luo et al., 2024b)	1675.78	880.61	98.91	–	–	–	–	–
MAGNN (Chen et al., 2023)	1675.70	880.44	98.29	–	–	–	–	–
STT-LLM	**1664.59**	881.20	**96.80**	**0.7338**	0.1935	**0.4138**	0.2637	**0.5675**

Conclusion

Summary of Contributions

- Presented **STT-LLM framework**, a tokenization strategy for longitudinal clinical data.
- Introduced structural–temporal embeddings to capture both biological relationships and temporal dynamics.
- Enabled LLMs to process longitudinal clinical profiles **without modifying the model architecture**.

Key Results

- STT-LLM consistently **outperforms baseline LLMs** in sequence prediction and anomaly detection.
- Ablation studies show the **importance of combining structural** and **temporal** information.
- The framework produces more meaningful and interpretable representations of biomedical profiles.

Take-home Message

- Integrating domain knowledge (biological pathways) with temporal modeling allows LLMs to better analyze longitudinal clinical data.

Thanks for your Attention!

Dr.-Ing. Maxx Richard Rahman

Scientific Researcher,
German Research Centre for Artificial Intelligence (DFKI),
Saarbrücken, Germany
maxx_richard.rahman@dfki.de