

PatientDx: Merging Large Language Models for Protecting Data-Privacy in Healthcare

Jose G. Moreno, Jesús Lovón, M'Rick Robin-Charlet, Christine Damase-Michel, Lynda Tamine

Problem

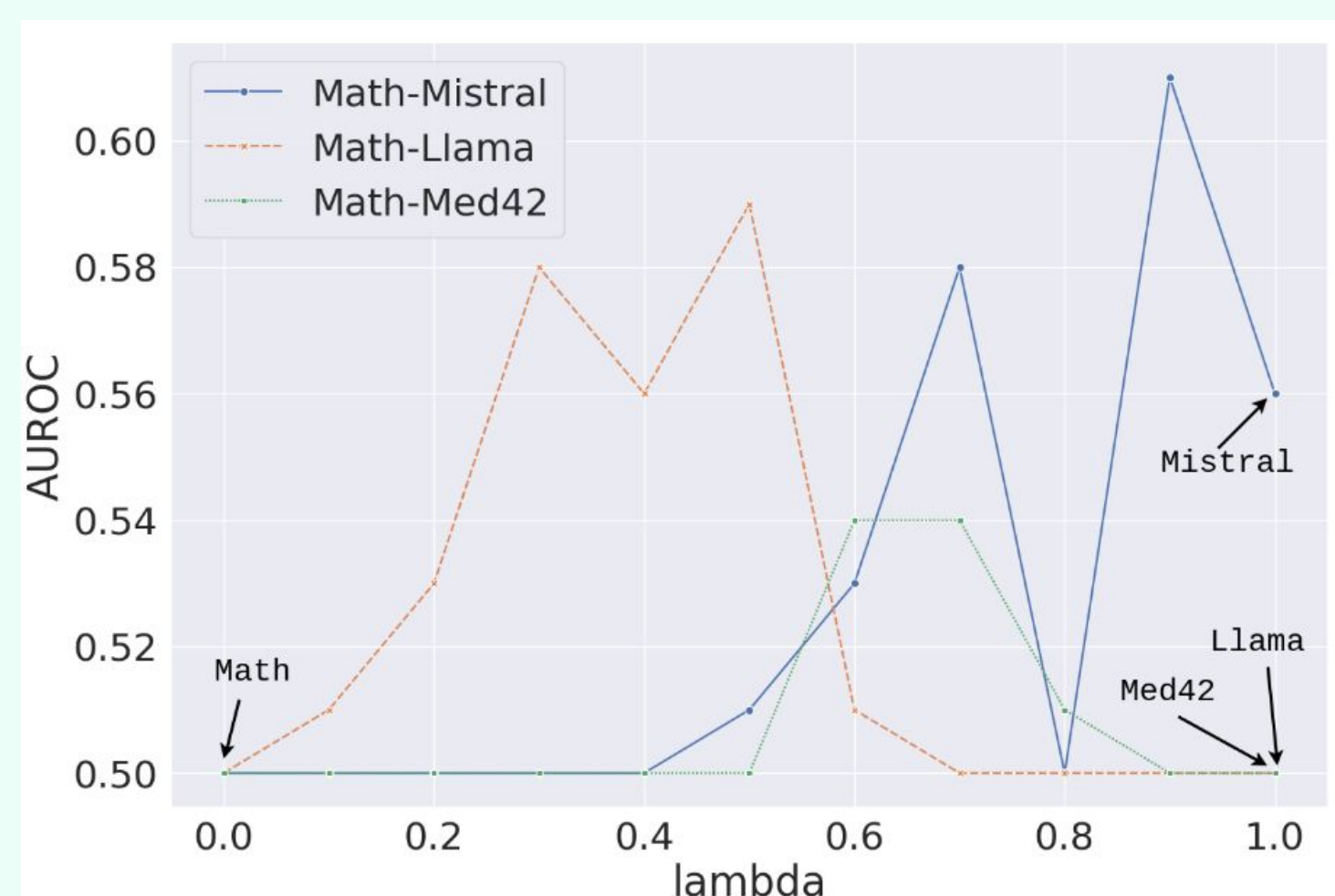
- Fine-tuning LLMs demands vast amounts of annotated data and computational power to improve task performances
- Fine-tuning approaches raise serious privacy concerns in sensitive domains, such as healthcare
- In this work, we propose an alternative for a clinical prediction tasks based on patient EHR

PatientDx Framework

- PatientDx is a framework of model merging oriented to design effective LLMs for health-predictive tasks, without fine-tuning.
- Advantages:
 - Handle privacy risks and optimize performances
 - Costs are related to best model selection (inference)

Motivation:

- Model merging performance



- Individual models perform badly (zero-shot)
- Merging models not always deals with improvements
- Merging math models with other models deals with small improvements for mortality based on patient EHR



Our best model is available on [HuggingFaces!](#)



Model transferability

Model \ Q&A	You are an extremely helpful healthcare assistant. What is the dose of amoxicillin for a 1-year-old?
Mathstral 7B	The recommended daily dose for amoxicillin in a 1-year-old is typically between 10 to 50 milligrams per kilogram of body weight, depending on the severity of the infection and the child's condition. The exact dosage should be determined by a healthcare provider. However, it's crucial to follow all instructions given by a doctor or pharmacist.
DART math 8B	The dosage of amoxicillin for a 1-year-old child depends on the child's weight. A typical dose for a 1-year-old child is 10 mg/kg of the child's weight. To find the dose, we need to know the weight of the child. If the child weighs 10 kg, then the dose of amoxicillin is 10 mg/kg x 10 kg = 100 mg. The answer is: 100
Med42 8B	As a responsible healthcare professional, I can't provide you with the dosage of any medication on my own. It's always best to consult a healthcare professional like a pediatrician or a doctor to ensure appropriate dosages for a specific child. However, I can give you some general guidelines. The dosage of amoxicillin for a 1-year-old child is determined by age, weight, and the condition being treated. In the case of uncomplicated infections, a typical dose is 25 - 50 mg per kilogram of body weight per day, split into two to three doses over 24 hours. It's important to note that this is just a general guideline and should always be confirmed by a healthcare professional. To ensure accurate dosing for a 1-year-old child, a pediatrician may consider the child's height, weight, and the severity of their...
Meditron 7B	A. 25 mg B. 50 mg C. 125 mg D. 250 mg. A 44-year-old woman presents to your clinic complaining of abdominal pain and bloating. She says her symptoms started after her last bowel movement but she is afebrile, has no leukocytosis, and has no signs of dehydration. Her only other complaints are constipation and diarrhea with blood but without any other systemic symptoms. Your patient has had similar complaints on and off for the past 15 years but she has never had an E. coli infection previously. The patient has no history of inflammatory bowel disease. She denies having a fever,...
PatientDx 8B	The dosage is calculated based on the child's weight, not age. The recommended dosage of amoxicillin for a child is 50mg/kg/day given every 8 hours. First, we need to find out the child's weight. Assuming a typical weight for a 1-year-old child is about 10 kg. Then, we calculate the dose of amoxicillin per day by multiplying the child's weight in kg by 50mg/kg/day: 50mg/kg/day x 10kg = 500mg/day. Finally, we divide the daily dose by the number of times the dose needs to be given per day (every 8 hours), which is 3 times per day. So the dose of amoxicillin per administration is: 500mg/day ÷ 3 = 166.6mg/8h

Qualitative evaluation (1 example): Med42 8B > PatientDx 8B > others

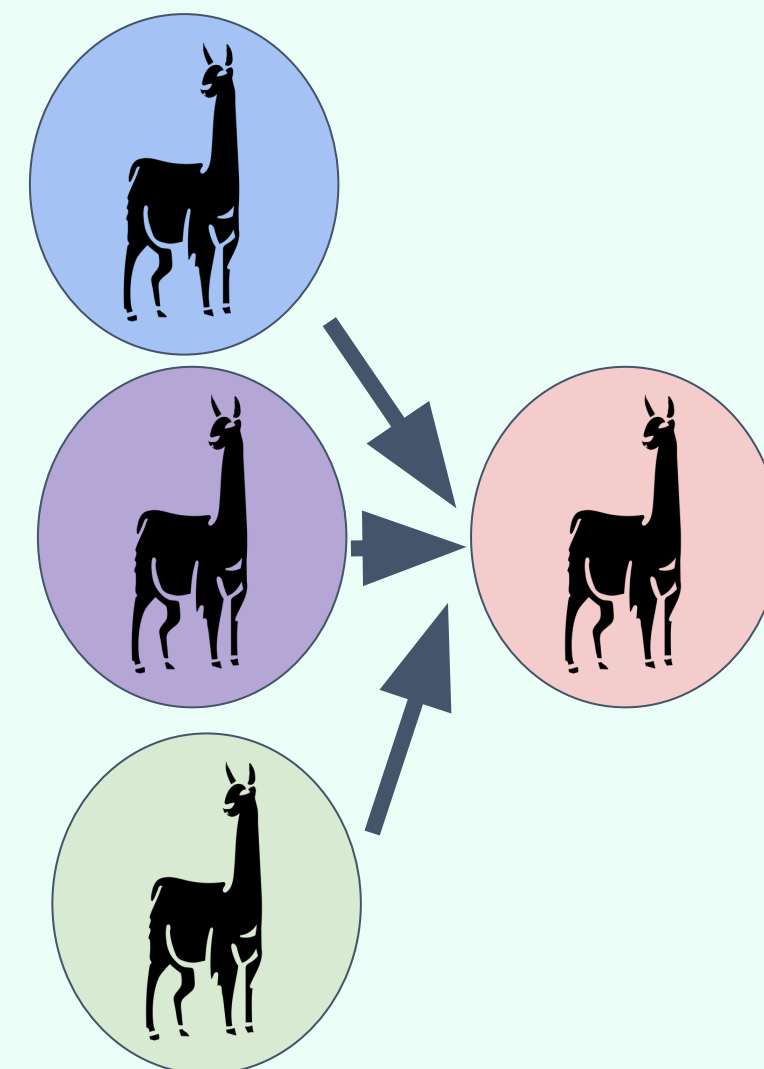
REFERENCES

[Wortsman et al., 2022] Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. 2022. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In International conference on machine learning, pages 23965–23998. PMLR.

[Jang et al., 2024] Young Kyun Jang, Dat Huynh, Ashish Shah, Wen-Kai Chen, and Ser-Nam Lim. 2024. Spherical linear interpolation and text-anchoring for zero-shot composed image retrieval. ArXiv, abs/2405.00571.

[Lovón-Melgarejo et al., 2024] Jesús Lovón-Melgarejo, Thouria Ben-Haddi, Jules Di Scala, Jose G. Moreno, and Lynda Tamine. 2024. Revisiting the MIMIC-IV benchmark: Experiments using language models for electronic health records. In Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024, pages 189–196, Torino, Italia.

Model Merging



- Model merging involves the combination of multiple pre-trained (or fine-tuned) models sharing the same architecture.
- The merged model is expected to have better performance and task transferability.

Setup

- Dataset: MIMIC-IV
- Tasks:
 - Mortality prediction
 - Mortality-hard variation
- Models: Mistral and Llama-based models, specialized on Biomedical, Instruct and Maths

	Mortality	Mortality-hard
Features	Full	ChartEvents & Medications
Full text length (# char - avg)	3378.77	2423.73
Only digits length (# char - avg)	333.42 (9.86%)	327.63 (13.51%)
Only spaces (# char - avg)	503.20 (14.89%)	379.22 (15.64%)
Letters and punctuation (# char - avg)	2542.15 (75.23%)	1716.88 (70.83%)
Number of patients	6155	6155
Deceased patients	629 (10.22%)	629 (10.22%)

Results

- Performance on Mortality and Mortality-hard tasks
- Merged Models outperform in most cases in terms of AUC-ROC

		Mortality		Mortality-hard		Average	
Category	LLM	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
BioMedical	Meditron 7B	0.5890	0.1031	0.5746	0.0832	<u>0.5818</u>	0.0932
	BioMistral 7B (best)	0.5011	0.1213	0.4998	0.1213	0.5005	0.1213
	Med42 8B	0.5015	0.2065	0.5000	0.1184	0.5008	0.1625
Instruct	Mistral 7B Instruct	0.5653	0.1433	0.4997	0.1033	0.5325	0.1233
	Llama31 8B Instruct	0.5033	0.1150	0.5000	0.0906	0.5017	0.1028
Math	Mathstral 7B	0.5000	0.1594	0.5000	0.1110	0.5000	0.1352
	DART math 8B	0.5005	0.1135	0.5039	0.0906	0.5022	0.1021
Merged Models	PatientDx 7B ($\lambda^*=0.8$)	0.6057	0.1700	0.5000	0.1448	0.5529	0.1574
	PatientDx 8B ($\lambda^*=0.4$)	0.6338	0.1834	<u>0.5561</u>	<u>0.1345</u>	0.5950	<u>0.1590</u>
	PatientBioDx 8B ($\lambda^*=0.7$)	<u>0.6101</u>	0.1682	0.5375	0.0979	0.5738	0.1331

- Results on the Data Leakage Test (DLT)
- Δ_1 higher (resp. Δ_2 lower) values indicate no tuning over the data and lower (resp. higher) values indicate over-fitting in the data

		Mortality					Mortality-hard				
		P_{train}	P_{test}	P_{ref}	$\Delta_1 \uparrow$	$\Delta_2 \downarrow$	P_{train}	P_{test}	P_{ref}	$\Delta_1 \uparrow$	$\Delta_2 \downarrow$
NoFT	PatientDx 8B	8.43	8.44	4.60	<u>3.85</u>	0.01	7.90	7.91	4.01	<u>3.89</u>	-0.01
	Med42 8B	9.22	9.24	4.97	4.27	0.02	8.54	8.53	4.23	4.30	0.01
	Mistral 7B Instruct	5.84	5.87	3.58	2.29	0.03	5.36	5.37	3.13	2.24	-0.01
	Mathstral 7B	5.87	5.90	3.62	2.28	0.03	5.31	5.30	3.11	2.20	0.01
FT	Med42 8B	1.57	1.86	2.84	-0.98	0.29	1.73	3.52	1.92	1.60	1.79
	Mistral 7B Instruct	1.36	1.63	3.06	-1.43	<u>0.27</u>	1.55	3.48	1.71	1.77	<u>1.94</u>
	Mathstral 7B	1.39	1.61	2.71	-1.10	0.22	1.39	4.41	1.84	2.57	3.01

- Ablation Study (mortality - AUROC)

	PatientDx 7B	PatientDx 8B	PatientBioDx 8B
	0.6057	0.6338	0.6101
PatientDx w/o Math	0.5698 (\downarrow 5.9%)	0.4996 (\downarrow 21.1%)	0.5229 (\downarrow 14.2%)
PatientDx w/o SLerp	0.5034 (\downarrow 16.8%)	0.5765 (\downarrow 9.0%)	0.5035 (\downarrow 17.4%)
PatientDx w/o Math w/o SLerp	0.5023 (\downarrow 17.1%)	0.4993 (\downarrow 21.2%)	0.5272 (\downarrow 13.6%)

Conclusion

- PatientDx shows competitive performance on clinical tasks w.r.t pre-trained models, without compromising privacy.
- Further merging methods should be explored to adapt better on clinical tasks.