CAN WE GENERATE PORTABLE REPRESENTATIONS FOR CLINICAL TIME SERIES DATA USING LLMS?

Zongliang Ji[*1,3], **Yifei Sun**[*1,3], **Andre Amaral**[1,2], **Anna Goldenberg**[1,3], **Rahul G. Krishnan**[1,3]
[1]University of Toronto, Canada, [2]Sunnybrook Health Sciences Centre, Canada,
[3]Vector Institute, Canada
{jerryji, rahulgk}@cs.toronto.edu

ABSTRACT

Deploying clinical ML is slow and brittle: models that work at one hospital often degrade under distribution shifts at the next. In this work, we study a simple question – can large language models (LLMs) create *portable* patient embeddings i.e. representations of patients enable a downstream predictor built on one hospital to be used elsewhere with minimal-to-no retraining and fine-tuning. To do so, we map from irregular ICU time series onto concise natural language summaries using a frozen LLM, then embed each summary with a frozen text embedding model to obtain a fixed length vector capable of serving as input to a variety of downstream predictors. Across three cohorts (MIMIC-IV, HIRID, PPICU), on multiple clinically grounded forecasting and classification tasks, we find that our approach is simple, easy to use and competitive with in-distribution with grid imputation, self-supervised representation learning, and time series foundation models, while exhibiting smaller relative performance drops when transferring to new hospitals. We study the variation in performance across prompt design, with structured prompts being crucial to reducing the variance of the predictive models without altering mean accuracy. We find that using these portable representations improves few-shot learning and does not increase demographic recoverability of age or sex relative to baselines, suggesting little additional privacy risk. Our work points to the potential that LLMs hold as tools to enable the scalable deployment of production grade predictive models by reducing the engineering overhead[1].

1 INTRODUCTION

Clinical machine learning tools continue to be deployed one site at a time. Teams build a model at Hospital A, run a silent trial, tune thresholds and features, and only then attempt deployment. The process is repeated at Hospital B. Each iteration introduces distribution, population, and incidence shift as laboratory measurement policies, case mix, and disease prevalence change that can degrade model performance. These changes inflate time to deployment, lengthen validation time and calibration cycles, and slow the delivery of benefit to patients.

The status quo in healthcare treats the *model* as the object that is transferrable across institutions. Interoperability standards such as OMOP and FHIR reduce data wrangling but do not guarantee an input representation that will remain predictive when moved to a new site. Domain adaptation and invariant risk minimization typically adapt the model to the target hospital, which preserves accuracy at the cost of additional data, labels, and calibration. Foundation models trained on electronic health records improve in-distribution performance and reduce task-specific engineering, yet they are often coupled to a particular sensor schema or sampling profile, and model transfer across sites still requires tuning. In short, current solutions standardize data formats, sometimes stripping clinically salient context, or recalibrate models.

We adopt a complementary thesis: *portable input representations enable portable models*. i.e. if heterogeneous electronic medical records can be mapped into a semantically aligned, site-agnostic interface, then conventional predictors will require substantially less site-specific adaptation. Rather

[1]Code is available at https://github.com/Jerryji007/Record2Vec-ICLR2026

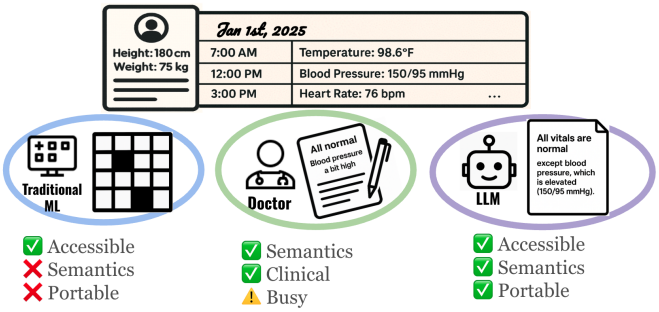

Figure 1: Motivation for Record2Vec. Numeric imputation loses clinical semantics and limits portability; human handoffs preserve meaning but are costly and variable. LLMs can create handoff-style summaries that retain semantics and provide portable inputs for forecasting and classification.

than forcing each hospital's raw schema into a shared syntactic format, we aim to produce a shared semantic representation.

Clinicians already solve an analogous problem. When assuming care for a new patient or starting a new shift, physicians interpret heterogeneous measurements through structured narrative handoffs that foreground clinically salient context while abstracting away measurement idiosyncrasies. We investigate whether large language models (LLMs) can analogously transform irregular multivariate patient histories into consistent, handoff-style summaries that serve as a portable intermediate representation.

Concretely, we instantiate this idea with a summarize-then-embed pipeline. For each patient window, a frozen LLM converts the irregular time-series record into a concise natural-language summary. A frozen text embedding model then maps that summary into a fixed-length vector consumed by standard forecasting and classification models without architectural modification.

This design leverages several properties of language models. First, summarization operates in a semantic space rather than a schema-bound numeric space, allowing normalization of units, resolution of synonyms, and abstraction over site-specific coding artifacts. Second, natural language provides a common representational substrate that aligns heterogeneous sampling policies and missingness patterns into comparable clinical concepts. Third, fixed-length embeddings standardize the downstream interface, simplifying training budgets and enabling zero- or few-shot transfer. Freezing both the summarizer and embedder further constrains overfitting to site-specific artifacts and improves reproducibility by decoupling representation learning from downstream optimization.

Our contributions are threefold:

- We rethink how machine learning in healthcare should operate via a deployment-first framing of prediction that focuses on creating portable input representations for healthcare, aiming to reduce per-site engineering and calibration.
- We present Record2Vec, a practical summarize-then-embed pipeline using frozen language models to transform irregular ICU histories into fixed-length vectors consumed by standard predictors without architectural changes.
- We conduct a multi-site evaluation across three cohorts and multiple tasks showing that the learned representations are competitive in-distribution, more portable across hospitals, more data-efficient, and match or improve on privacy preservation relative to established pipelines.

2 Related Work

Clinical schema and harmonization. Common data models and interoperability standards such as OMOP and HL7 FHIR have substantially reduced extract–transform–load burden and enabled multi–institutional reuse of EHRs through shared schemas, vocabularies, and APIs (Stang et al., 2010; Bender & Sartipi, 2013; Reich et al., 2024; Vorisek et al., 2022; Tabari et al., 2024). While effective for data sharing and governance, these frameworks standardize *format* rather than learn task–optimized, site–invariant representations for predictive modeling.

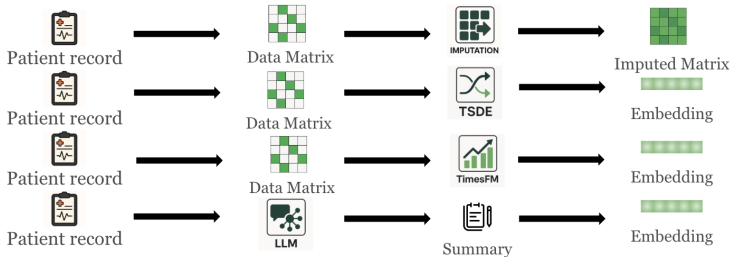

Figure 2: Methods to generate medical-record representations. Top to bottom: imputation pipeline; self-supervised TS representation (TSDE); TS foundation model (TimesFM); and Record2Vec: LLM summary followed by text embedding.

Domain shift, generalization, and adaptation. A large literature addresses distribution shift by adapting model parameters to target environments. Domain–adversarial training encourages site–invariant features via a domain classifier (Ganin & Lempitsky, 2016); correlation alignment matches second–order statistics across domains (Sun & Saenko, 2016). Domain generalization methods such as invariant risk minimization seek predictors stable across environments (Arjovsky et al., 2019; Subbaswamy et al., 2022), and distributionally robust optimization emphasizes worst–case groups to improve reliability under shift (Sagawa et al., 2020). These approaches are model–centric, often requiring environment labels, target–site data, or careful tuning. In contrast, we study *input level* portability by learning a site–agnostic representation that standard predictors can consume without architectural changes.

EHR and self–supervised representation learning. Representation learning on structured EHRs and clinical time series has advanced rapidly. BEHRT pretrains Transformers over longitudinal code sequences to learn patient timelines (Li et al., 2020b; Steinberg et al., 2020). Recent surveys review concept and patient–level embeddings for EHRs (Kauffman et al., 2025). Time–series foundation models pretrained at scale demonstrate strong zero–/few–shot forecasting, notably TimesFM (Das et al., 2024). Diffusion–style self–supervision for general time–series representation learning, such as TSDE, learns versatile embeddings via imputation–interpolation–forecasting masking with dual–encoder Transformers (Senane et al., 2024). These are strong baselines; however, portability and deployment cost across hospitals are rarely treated as first–class evaluation endpoints.

LLMs for clinical summarization and structuring. Large language models have been applied to clinical summarization and structuring tasks. Adapted LLMs can outperform experts on multiple clinical text summarization tasks (Van Veen et al., 2024), and there is growing guidance on prompt design for medical use (Zaghir et al., 2024). Evaluations of medical evidence summarization further characterize strengths and limitations (Tang et al., 2023). Beyond notes, scoping reviews document integrating standardized terminologies like SNOMED CT with LLMs for normalization and coding (Chang & Sung, 2024; Luo et al., 2020). While prior work uses LLMs for text summarization or coding, we repurpose them as an *information transformation layer* for numeric ICU time series, bridging structured and unstructured modalities.

Recent research investigates serializing structured EHR data for LLM processing. Approaches like TabLLM (Hegselmann et al., 2023) and DeLLiriuM (Contreras et al., 2025) fine-tune LLMs on text to perform direct prediction. Other studies (Lee et al., 2025; Gao et al., 2024; Hegselmann et al., 2025) use frozen models to embed raw data serializations. A critical distinction lies in our deployment philosophy. These prior methods optimize the model pipeline and often require the transfer of site-tuned LLMs. In contrast, our approach focuses on preparing a transferable input representation X. We use a standard and locally deployed frozen LLM to summarize dense time-series data. This step normalizes site-specific artifacts at the data level. Consequently, hospitals can generate portable representations that allow any lightweight downstream classifier to function effectively without porting entire end-to-end model pipelines.

3 Methodology and Setup

We design a controlled study to test whether natural language can serve as an *information transformation layer* that yields *portable inputs* for ICU prediction, formalizing sites, inputs, and tasks

across four method families, three ICU cohorts, and seven prediction tasks, and presenting our Record2Vec pipeline with baselines, data, and evaluation.

Problem formalization. Let sites be $S \in \{\text{MIMIC-IV}, \text{HiRID}, \text{PPICU}\}$. For a stay i at site s, the irregular ICU record over a 48-hour window is $\mathcal{R}_i^{(s)} = \{(c, \{(t_k, v_k)\}_{k=1}^{K_c}) : c \in \mathcal{C}^{(s)}\}$, a dictionary from clinical concepts c to time–value pairs (t_k, v_k) observed in the window. We consider methods that transform $\mathcal{R}_i^{(s)}$ into either (i) a regular *grid* $X_i \in \mathbb{R}^{D^{(s)} \times T}$ with $T = 48$ hours, or (ii) a fixed-length *vector* $z_i \in \mathbb{R}^d$. A downstream predictor f_θ maps inputs to task labels y_i across standard ICU outcomes spanning different label-prevalence regimes.

3.1 RECORD2VEC

We treat language as an information transformation layer. Given the 48h irregular record $\mathcal{R}_i^{(s)}$ for stay i at site s, Record2Vec first produces a concise clinical summary and then embeds that summary into a fixed-length vector used by shared decoders.

Summarization. A frozen LLM g_ϕ maps each irregular window and a prompt π to a concise summary, i.e., $\text{text}_i = g_\phi(\mathcal{R}_i^{(s)}, \pi)$. We compare *structured* slot-based prompts that elicit vitals, labs, therapies, trajectories, salient events, and gaps with *free-form* prompts that request a short narrative of states and trends. We evaluate three summarizers representing different deployment regimes: a large general model (Gemini-2.0 Flash Comanici et al. (2025)), a clinically tuned model (MedGemma Sellergren et al. (2025)), and a small open model (Llama-3.1 Dubey et al. (2024)). All LLMs remain frozen without finetuning.

Embedding. A frozen text encoder h_ψ maps the summary to a fixed-length vector used by downstream predictors, $z_i = h_\psi(\text{text}_i) \in \mathbb{R}^d$. Unless noted, we use Qwen3 text-embedding as h_ψ [2] As a no-summarization control, we also embed a canonical serialization of the record directly $z_i^{\text{direct}} = h_\psi(\text{serialize}(\mathcal{R}_i^{(s)}))$.

3.2 BASELINES

We compare Record2Vec to three strong families used for ICU time series, plus the direct-embed control.

- **Grid imputation pipelines.** Irregular records are discretized into hourly grids $X \in \mathbb{R}^{D^{(s)} \times 48}$, then completed via (a) mean fill, (b) right-shift carry-forward, or (c) linear interpolation. Completed grids feed the shared decoders for forecasting, regression, and classification. Per-cohort normalization uses training-split statistics and is applied only to grid methods.
- **Self-supervised representation learning (TSDE).** Time-Series Diffusion Embedding (TSDE) (Senane et al., 2024) learns general-purpose embeddings via masked imputation/interpolation/forecasting objectives. We train TSDE on each cohort's training split to obtain one vector per example; vectors are converted to grids by the shared projection in the downstream model and decoded with the same heads as other methods. Model selection uses only training/validation within the source cohort.
- **Time-series foundation model (TimesFM).** TimesFM (Das et al., 2024) is a decoder-only attention model pretrained on diverse time series for strong zero-shot forecasting. We use TimesFM as a *frozen encoder*: after mean imputing each feature to form regular per-feature series, we extract hidden representations feature-by-feature and average across features to obtain a window-level embedding, which is then mapped to a grid by the shared projection and decoded. No TimesFM finetuning is performed.
- **General healthcare predictive framework (GenHPF-variant).** GenHPF (Hur et al., 2024) divides patient features hierarchically, encodes each medical event into an embedding followed by a step to aggregate them. We create a variant of this method with minor modifications to their textual templates and follow the same downstream learning pipeline. [3]

3.3 DATASETS AND PREPROCESSING

We use three ICU cohorts: MIMIC-IV (Johnson et al., 2023a), HiRID (Yèche et al., 2021), and PPICU. PPICU is an external dataset from a distinct hospital system. In total, the cohorts include

[2] We perform embedding model ablations; the results are recorded in Appendix L.

[3] Details on how mapping was done are shown in Appendix N.

Dataset →	**HiRID**					**MIMIC**					**PPICU**					**Wins**
Tasks → **Method ↓**	Forecast MSE	LoS MAE	Mort AUROC	Drug Recall	Lab Recall	Forecast MSE	LoS MAE	Mort AUROC	Drug Recall	Lab Recall	Forecast MSE	LoS MAE	Mort AUROC	Drug Recall	Lab Recall	
Mean	0.040	0.378	0.914	0.878	0.857	0.035	0.447	0.847	0.838	0.886	0.028	0.528	0.842	0.834	0.847	0
Right shift	0.041	**0.342**	0.923	0.884	0.858	0.037	0.409	0.886	0.841	0.884	0.031	0.490	0.868	0.831	0.846	1
Interpolation	0.435	0.370	0.918	0.879	0.853	0.103	0.430	0.873	0.844	0.893	0.133	0.493	0.857	0.837	0.847	0
TSDE	0.029	0.411	0.923	0.901	0.902	0.030	0.406	**0.915**	0.888	0.899	0.053	0.485	**0.890**	0.899	0.870	1
TimesFM	0.028	0.440	0.826	0.850	0.925	0.030	0.413	0.791	0.806	0.940	0.036	0.494	0.658	0.923	0.925	0
GenHPF	–	–	0.836	0.770	0.741	–	–	0.776	0.720	0.713	–	–	0.780	0.835	0.6773	0
Record2Vec	**0.021**	0.347	**0.930**	**0.911**	**0.931**	**0.027**	**0.328**	0.888	**0.903**	**0.947**	**0.017**	**0.358**	**0.890**	**0.937**	**0.936**	13

Table 1: In-distribution Result (RQ1). Best (per column) in **bold**; second-best underlined. *Wins* counts the number of bests per method across all 15 downstream tasks.

57,212 stays with 60 variables for MIMIC-IV, 32,216 stays with 64 variables for HiRID, and 39,000 stays with 75 variables for PPICU.[4] We segment each stay into non-overlapping windows of 48 hours. The variables cover laboratory tests, vital signs, and clinical interventions.

For the grid baselines, we normalize values within each cohort using statistics computed on the corresponding training split, consistent with prior benchmarks where normalization is crucial for numeric model stability. For text-based transformations, we keep raw magnitudes and the native clinical units to preserve meaning in the summaries i.e we do not perform further modification of the raw data. For language inputs, we create a canonical serialization of the window. For each variable present, the serialization lists the name and the full sequence of timestamps and raw values observed during the window. This serialization is the input to the summarizers and also serves as the no-summarization control described in Section 3.1. Additional details on concept mappings, unit handling, and filtering rules, which were curated with practicing ICU clinicians to ensure correct harmonization of variables, appear in Appendix C.2.

3.4 Tasks and Evaluation

We study five predictive tasks and two privacy probes. The tasks are: multivariate forecasting of all variables over the next 24 hours; remaining length of stay at the window end; in-hospital mortality; two treatment indicators for vasopressor and antibiotic use in the next 24 hours; and a multi-label outcome for whether ten common blood tests will be ordered in the next 24 hours. The privacy probes test demographic recoverability by predicting age (clipped to 18–90) and sex.

We use three settings. In-distribution trains and validates within one cohort and tests on its holdout split. Cross-site trains on a source cohort and tests on a distinct target without target labels; we report target accuracy and the drop relative to in-distribution. Few-shot starts from a source-trained model and fine-tunes on 16, 64, or 512 labeled target examples, then tests on the target split.

We report MAE for length of stay and age, masked MSE for forecasting, and micro-averaged recall (with precision and F1) for classification. Results average four seeds with mean and standard deviation. Budgets, early stopping, capacity, and regularization are matched within input type. TSDE is trained on the training split with validation selection. TimesFM, g_ϕ, and h_ψ remain frozen. Full hyperparameters appear in Appendix D.

4 Results

We organize the evaluation around seven research questions probing in-distribution performance (RQ1), cross-cohort transfer (RQ2), the value of summarization and model choice (RQ3), prompt sensitivity (RQ4), few-shot adaptation (RQ5), privacy (RQ6), and information analysis (RQ7). Unless noted, results tables report means over seeds with the top performer highlighted. For consistency, the results sections report numbers from a single downstream model (PatchTSMixer), except for mortality, where we report the best performance across all models tested in the mortality benchmark (Yèche et al., 2022); we verified the same trends hold with neural network architectures including multi-layer perceptron (MLP), LSTM (Hochreiter & Schmidhuber, 1997), TimeMixer (Wang et al., 2024b) in Appendix O.

4.1 How do the four methods compare in distribution? (RQ1)

Record2Vec achieves the strongest in-distribution results overall, winning 13 of 15 tasks and ranking second on the remaining two.

[4]Differences in EHR systems lead to different variable sets across sites.

Dataset →	**HiRID → PPICU**					**MIMIC → PPICU**					**Wins**
Tasks → **Method ↓**	Forecast MSE	LoS MAE	Mort AUROC	Drug Recall	Lab Recall	Forecast MSE	LoS MAE	Mort AUROC	Drug Recall	Lab Recall	
Mean	0.306	1.09	0.50	0.42	0.77	4.839	0.85	0.50	0.30	0.81	0
Right shift	0.417	1.03	0.50	0.53	0.88	6.344	0.76	0.50	0.23	0.82	0
Interpolation	24.80	0.91	0.50	0.34	0.81	13.41	<u>0.71</u>	0.50	0.21	0.80	0
TSDE	0.209	<u>0.73</u>	0.49	0.91	0.91	0.269	0.76	0.51	<u>0.90</u>	<u>0.90</u>	0
TimesFM	0.217	0.80	0.64	0.93	0.95	0.284	0.93	0.69	**0.95**	**0.95**	2
GenHPF	-	-	0.582	0.699	0.316	-	-	0.425	0.442	0.428	0
Record2Vec Template	<u>0.195</u>	0.98	<u>0.72</u>	<u>0.96</u>	<u>0.96</u>	<u>0.263</u>	0.77	<u>0.71</u>	**0.95**	**0.95**	2
Record2Vec	**0.183**	**0.69**	**0.72**	**0.97**	**0.97**	**0.242**	**0.49**	**0.72**	**0.95**	**0.95**	10

Table 2: Transfer Learning Result (RQ2). Best per column in **bold**; second-best <u>underlined</u>. Wins counts tied bests.

Table 1 summarizes performance within each cohort. A clear pattern emerges: Record2Vec leads on most outcomes in HiRID, sweeps all tasks in MIMIC, and again dominates in PPICU with one exception. The two columns not led by Record2Vec are PPICU mortality, where TSDE is best, and HiRID length of stay, where a simple right-shift imputation edges ahead. Representation-learning baselines (TSDE, TimesFM) are consistently competitive and often place second, while classic grid imputations rarely win and generally trail across cohorts and endpoints. These results indicate that a language-mediated input yields robust utility across forecasting, regression, and classification without tailoring model architectures to a site.

We hypothesize three drivers. First, summarization captures clinically salient semantics—states, trajectories, interventions, salient events—while aligning heterogeneous names, units, and sampling policies into a shared space. This reduces reliance on ad hoc imputations and preserves meaning when measurements are sparse or irregular. Second, the fixed-length embedding offered by Record2Vec stabilizes the downstream interface, limiting sensitivity to missingness patterns and local measurement habits that degrade grid-based inputs. Third, compared with TSDE and TimesFM, which emphasize correlations and trends in numeric streams, Record2Vec adds a layer of clinical context that appears to aid discrimination on classification endpoints while remaining competitive on regression. Overall, the findings support the view that LLM-driven transformation produces more informative and portable inputs than either imputation or purely numeric representation learning.

4.2 Does the LLM based representation improve transferability across ICUs? (RQ2)

Record2Vec transfers best across ICUs, winning 10 of 10 columns in Table 2 with two ties by TimesFM.

When models are trained on HiRID or MIMIC and evaluated on PPICU, Record2Vec consistently ranks first across forecasting, length of stay, mortality, treatment, and lab prediction. In contrast, grid imputations degrade sharply under shift, and several classification scores collapse toward chance. TSDE and TimesFM are stronger than imputation, yet they still trail Record2Vec in most settings. TimesFM ties for the top position in two columns and, together with Record2Vec, benefits from the larger MIMIC source when predicting common interventions and labs at PPICU. The pattern is clear: methods that rely on site-specific numeric regularization or self-supervision on the source cohort do not maintain accuracy when confronted with new variable sets, sampling habits, and missingness regimes.

We attribute these results to three factors. First, natural-language summaries align heterogeneous coding choices, units, and documentation styles into a shared clinical space before embedding. This reduces the burden on downstream decoders to re-learn site-specific conventions and helps preserve signal for tasks like mortality and treatment prediction. Second, the fixed-length embedding produced by Record2Vec offers a stable input interface that is less sensitive to irregular sampling and missingness patterns than grids, which explains the large gap from imputation under distribution shift. Third, compared with TSDE and TimesFM, Record2Vec injects semantic context about states, trends, and salient events extracted by the summarizer. To better understand the gains from summarization, we compared against the strategy used by related work (Hur et al., 2024; Gao et al., 2024; Hegselmann et al., 2025) to map patient data onto a fixed template. When comparing Record2Vec with TSDE, and TFM, against fixed-template baselines, we observe additional performance gains that can be attributed to summarization. We hypothesize these gains primarily come due to the standardization of heterogeneous patient profiles across sites while preserving biologically relevant in-

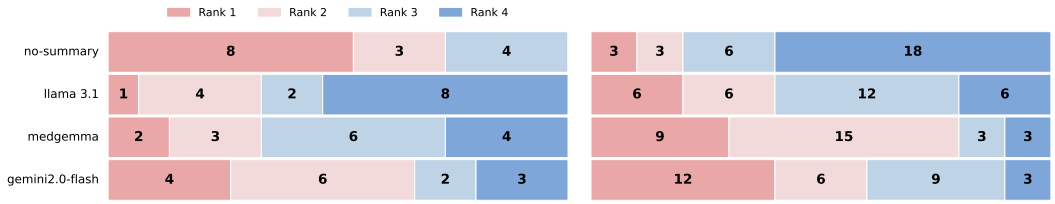

Figure 3: Rank distributions for No-summary vs. three LLM variants across 15 in-distribution tasks (left) and 30 cross-site transfer tasks (right). Methods are ranked based on performance across five downstream tasks: Forecast (MSE), LoS (MAE), Mortality (AUROC), Drug (Recall), and Lab (Recall). See the Appendix M for the detailed values.

formation for making predictions.Foundation pretraining helps TimesFM transfer better than TSDE, but it remains tied to numeric trends alone, whereas Record2Vec retains both numerical and clinical meaning. Together, these properties yield robust portability across hospitals for a broad set of ICU prediction tasks.

4.3 Do summaries help, and which summarizer works best? (RQ3)

Summarize-then-embed improves cross-site transfer while staying competitive in-distribution, and among summarizers Gemini-2.0 Flash and MedGemma perform best, with Llama-3.1 worse than both.

Figure 3 shows rank distributions for four variants: *no-summary* (directly embed the raw serialization), and three LLM summarizers (Llama-3.1, MedGemma, Gemini-2.0 Flash). On the left (in-distribution; 15 tasks), *no-summary* often attains the top rank, with the LLM-based variants close behind. Within a single site, feeding the embedding model the full, unsummarized content appears advantageous and does not require normalization into prose. On the right (cross-site transfer; 30 tasks), the pattern reverses. *No-summary* concentrates in the worst ranks overall, while the three summarize-then-embed variants dominate first and second ranks. Comparing the LLMs, Gemini-2.0 Flash and MedGemma account for most of the top positions in both settings; their gap narrows under transfer. Llama-3.1 accumulates the most last-place ranks among summarizers in both settings.

We think there are three reasons for these outcomes. First, summarization imposes a shared clinical language over heterogeneous variable names, units, and sampling practices, which removes site-specific idiosyncrasies before embedding. This harmonization matters most under shift and explains why Record2Vec transfers better than direct embedding of raw, site-specific streams. Second, within-site, direct embeddings preserve all numeric detail and local conventions, which can make representations highly distinctive and strong on held-out splits from the same cohort; however, precisely those site-specific details impair portability to other hospitals. Third, the choice of summarizer influences both fidelity and standardization. MedGemma's medical pretraining promotes clinically faithful phrasing that is stable across institutions, supporting transfer. Gemini-2.0 Flash's instruction following and planning yield concise, information-dense summaries that are marginally stronger in-distribution and remain competitive under transfer, leading to a small gap between the two. Llama-3.1, with smaller capacity and less domain specialization, tends to produce shorter or less standardized summaries, which hurts both within-site utility and cross-site robustness. **A major practical gain is efficiency: without summarization the raw serialization passes, on average, $\sim$25$\times$ more tokens to the embedder than the summarized version, cutting inference cost proportionally while improving transfer.**

4.4 How sensitive is performance to prompt design? (RQ4)

Across four prompting strategies, performance is broadly similar in-distribution and under transfer, with a slight edge for ICD-style prompts in transfer. Figure 4 plots rank distributions for zero-shot, chain-of-thought, trend-focused, and ICD-focused prompts using Gemini-2.0 Flash. The bars span ranks fairly evenly on both the in-distribution side and the cross-site side, indicating limited sensitivity to prompt choice at the aggregate level. We observe a small shift in transfer where ICD-style prompts collect more top ranks, while the other three strategies remain closely clustered. Within cohorts, no single prompt dominates consistently across forecasting, regression, or classification.

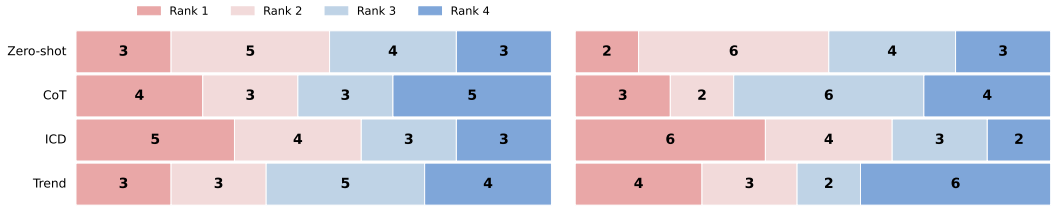

Figure 4: Rank distributions for four prompt variants with Gemini 2.0-flash across 15 in-distribution tasks (left) and 30 cross-site transfer tasks (right). Lower ranks indicate better performance. Methods are ranked based on performance across five downstream tasks: Forecast (MSE), LoS (MAE), Mortality (AUROC), Drug (Recall), and Lab (Recall). See the Appendix M for the detailed values.

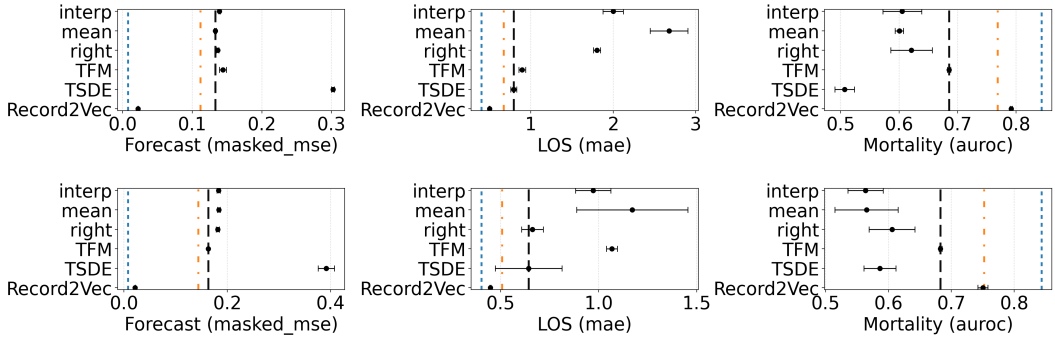

Figure 5: Few-shot finetuning with 16 labeled target samples for mortality prediction (RQ5) shown across six transfer settings. All tasks are reported the same metric as previous section: Forecast: masked mse, LOS: mae, Mortality: AUROC. The first row is the result of Hirid → Ppicu and the second row is Mimic → Ppicu. Reference lines: blue = best in-distribution upper bound, orange = best pre-finetune result, black = best finetuned baseline. Record2Vec surpassed baselines with a large gap, reaching comparable performance to in-distribution.

We think two factors explain this pattern. First, the summarizer is strong enough that high-level clinical content is retained under all four prompts, producing broadly comparable embeddings. Second, the ICD framing may aid transfer by nudging summaries toward standardized terminology that travels better across sites. Overall, these results suggest that prompt choice is not the primary driver of utility in our setup, and that more targeted prompt engineering could be a promising avenue for future improvement. Detailed prompt templates appear in Appendix E.

4.5 Does the LLM based representation improve few-shot downstream learning? (RQ5)

Record2Vec needs fewer samples to achieve generalization on small clinical datasets. Hospitals that are not part of large medical centers often lack sufficient patient data to train high-performing models from scratch. In Appendix I, we compare models trained on 1,000 PPICU samples with Record2Vec models that are pre-trained on HiRID or MIMIC and then fine-tuned on only 16 randomly selected PPICU samples. Beyond the gains from in-distribution training, Record2Vec delivers substantial improvements over its transfer variants and outperforms all competing methods under the same supervision budget. In multiple settings, the adapted Record2Vec models approach the in-distribution reference model trained on 36,019 samples, indicating that very limited supervision can recover most of the performance lost under distribution shift.

These gains likely arise because Record2Vec produces a compact representation that is already semantically aligned across hospitals, reducing the burden on the downstream head to learn site-specific conventions from scratch. Finetuning can therefore act mainly as a light calibration of task boundaries rather than a wholesale re-learning of measurement scales, missingness patterns, and naming differences. In effect, the summary abstracts away local differences while preserving clinically salient signals—states, trends, and recent interventions—that are informative for mortality, length of stay, and near-term actions. This combination of cross-site alignment and retained clinical

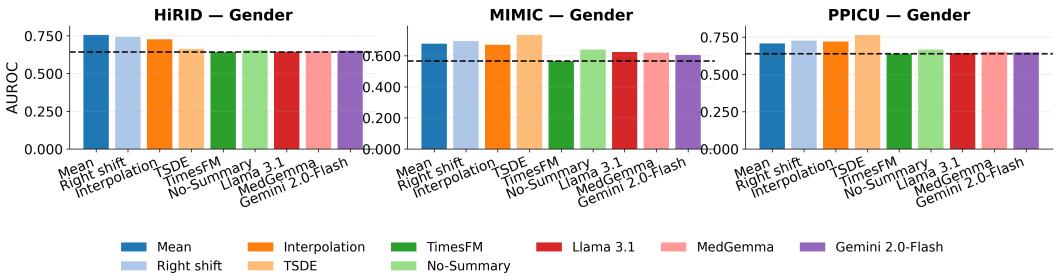

Figure 6: In-distribution privacy prediction on **Gender** across three ICU datasets. Each bar is a different method. The dashed horizontal line in each panel marks the least leaked classical baseline among {TimesFM, TSDE, Mean, Interpolation, Right shift} for that metric/dataset. Our method reached comparable and mostly reduced results in terms of privacy leakage. **Age** results have less than 0.5% gap and can be inferred from Table 4

content makes Record2Vec especially data-efficient in shifted settings, enabling effective few-shot learning when labeled target data are scarce.

4.6 Does the portable representation increase privacy risks? (RQ6)

We find no evidence that Record2Vec increases demographic leakage risk: gender prediction collapses to a constant baseline for all methods, and Record2Vec's age error is similar to or higher than baselines. Record2Vec's MAE is generally on par with competing approaches across cohorts, suggesting that the portable representation does not make age easier to infer. Figure 6 visualizes demographic recoverability from learned representations. For *gender* (binary), models trained on embeddings from grids, TSDE, TimesFM, and Record2Vec all degenerate to predicting a single class on held-out data, indicating near-chance information about gender regardless of method. We additionally examined performance across minority diagnostic subgroups and found no disparate impact, with Record2Vec tracking overall population trends without exhibiting worst-group performance degradation relative to baselines. Detailed numbers appear in Appendix O.

We think this outcome reflects how the pipeline shapes information. Record2Vec emphasizes clinical states, trends, and recent interventions needed for downstream tasks, rather than demographic markers. Because the summarizer and text encoder are frozen and not trained to predict demographics, they do not amplify demographic signal beyond what is already present in the record windows. In practice, the summaries in our setup rarely include explicit age or sex mentions, and the embedding stage is not optimized to capture them, which helps keep demographic recoverability low while preserving clinical utility. However, we emphasize that these results are specific to demographic leakage and do not preclude other privacy risks such as membership inference or embedding inversion.

4.7 What information is obtained or lost in each embedding? (RQ7)

Record2Vec captures semantically meaningful structure that improves in-distribution performance and, importantly, encodes cross-site invariant patterns, while attenuating demographic attributes. We compare the Record2Vec embedding—MedGemma summaries prompted with ICD codes—against imputation-based vectors to quantify task-specific information gain or loss (Figure 7). Unless stated otherwise, "gain" denotes the relative improvement ($\Delta\%$) of a metric aligned with task direction (e.g., lower for error; higher for AUROC/recall) with respect to the imputation baseline.

For in-distribution, the largest gains are observed for *forecasting* and *mortality* prediction, indicating that LLM-derived semantic abstractions, such as ICD-anchored summarization, improve outcome modeling and strengthen associations with significant temporal features. For cross-site transfer, We observe consistent improvements for *lab*, *drug*, *forecast*, *LOS*, and *mortality*, with the most pronounced effect on future drug-use prediction. This implies that *Record2Vec* captures patterns that generalize across institutions and reduces sensitivity to site-specific artifacts. For demographic attenuation, signals tied to *age* and *gender* are generally weakened or unchanged. This suggests

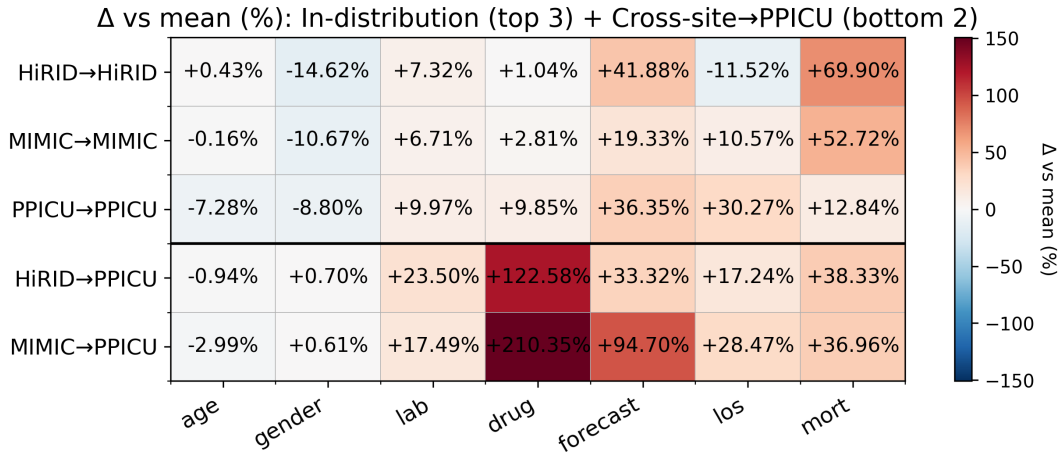

Figure 7: **Information gained vs. lost with Record2Vec (heatmaps; in-distribution and cross-site transfer).** Each cell shows the *task-aligned* relative change (%) of *Record2Vec* over representation vectors. Overall, *Record2Vec* yields the largest gains for *drug*, *forecast*, and *mortality*, including cross-site transfer, while demographic signals (*age*, *gender*) are attenuated or unchanged.

Record2Vec does not amplify, and mostly reduces, demographic attribute leakage relative to raw time series, which is another significant advantage under healthcare settings.

5 Limitations

Our framework requires sending patient records to LLM services. Although all experiments were conducted in a secure, IRB-compliant setting with strict controls, some patients and families may be uncomfortable with external processing. At present, this confines the approach to research use unless on-prem or fully compliant deployments are available. Generating LLM-based representations can be resource-intensive. Many hospitals may lack the compute or budget to support large-scale or continuous inference without additional investment (see Appendix H for a detailed analysis of token counts, latency, and costs). Finally, API-based or large local LLMs introduce non-trivial latency compared to lighter models, which can limit real-time applicability even if offline use remains practical. We acknowledge that our study is primarily empirical; we believe developing the theory for when and where this methodology works would significantly improve the reach of our work. We rely on downstream task performance as a proxy for representation quality and do not currently utilize theoretical tools (such as causal diagrams) to formally derive the conditions for site-invariance. Furthermore, we do not directly quantify end-to-end information loss (e.g., via mutual information or per-feature fidelity checks). While our downstream accuracy suggests that our method preserves important signals, a limitation of current work, it is challenging to characterize exactly what information is retained or lost during summarization.

6 Conclusion

Transforming irregular ICU records by first summarizing with an LLM and then embedding offers a promising route to portable, task-ready representations. Our results indicate competitive in-distribution performance, strong transferability, and improved few-shot learning without increased demographic leakage. We hypothesize that LLMs can leverage their learned priors to produce summaries that are standardized across sites, rich in biological information relevant to downstream decision making and close (in a distributional sense) across sites, enabling the portability of downstream predictive models. Nevertheless, deployment risks remain, particularly around privacy, cost, and latency. Future work should focus on creating compliant on-prem solutions, and prospective evaluations within clinical workflows. Assessing feature-level fidelity and developing theory to better understand and control the properties of LLM-based standardization is an important direction to understand how far this approach can be extended.

REPRODUCIBILITY STATEMENT

We take several steps to enable independent verification of our results. Data preprocessing (filtering, resampling, normalization, masking, and label construction) is documented in Appendix C.2. HiRID and MIMIC-IV are publicly accessible via PhysioNet; PPICU is a privately maintained dataset that will be shared upon reasonable request to the authors once the paper is public and de-anonymized, in accordance with institutional and data-use policies. Training details, including hardware requirements, software environment with pinned versions, all hyperparameters, and early-stopping criteria, are provided in Appendix D. To assess robustness, we tune key hyperparameters and repeat training/evaluation across multiple random seeds; the resulting analyses indicate our conclusions are stable to such variations. The code to reproduce our results is available at `https://github.com/Jerryji007/Record2Vec-ICLR2026`.

We take several steps to enable independent verification of our results. Data preprocessing (filtering, resampling, normalization, masking, and label construction) is documented in Appendix C.2. HiRID and MIMIC-IV are publicly accessible via PhysioNet; PPICU is a privately maintained dataset that will be shared upon reasonable request to the authors once the paper is public and de-anonymized, in accordance with institutional and data-use policies. Training details—including hardware requirements, software environment with pinned versions, all hyperparameters, and early-stopping criteria—are provided in Appendix D. To assess robustness, we tune key hyperparameters and repeat training/evaluation across multiple random seeds; the resulting analyses indicate our conclusions are stable to such variations.

ETHICS STATEMENT

Scope and oversight. Our work is in machine learning for healthcare and involves processing de-identified patient records under institutional oversight. All activities complied with applicable regulations and institutional policies (including HIPAA in the United States and the policies of our institutional review/ethics board). Data use agreements were honored throughout.

Training and authorization. All authors completed required human-subjects and privacy training and are certified/authorized to handle clinical data in secure research environments.

Data protection and privacy. We followed data-minimization and least-privilege principles; access to protected health information (PHI) was restricted to approved personnel and audit-logged. Data were stored on access-controlled servers with encryption in transit and at rest; exports outside secure infrastructure were prohibited. When sharing intermediate artifacts (e.g., for internal review), we ensured de-identification and suppression of quasi-identifiers.

Risk assessment and mitigation. We considered potential harms (privacy leakage, representational bias, inequitable performance) and mitigated them via dataset curation checks, stratified evaluation where applicable, and manual review of outputs used in the paper. No clinical decisions were made using the research system.

LLM usage. We employed LLMs in strictly limited ways (see Section A): (i) locally hosted models on institution-managed GPUs within secure, access-controlled servers; and (ii) third-party services (Gemini and ChatGPT families) *only* for grammar/wording polish of author-written text and for generating small iconography within figures (not full figures and not scientific content). No methods, ideas, analyses, code, or experiments were generated by LLMs, and all suggestions were manually reviewed.

Gemini 2.0 Flash configuration. Under a dedicated research agreement, we used a sandboxed deployment of the Gemini 2.0 Flash family strictly for research purposes. For any interactions related to healthcare data:

- prompts and inputs excluded PHI whenever possible; when unavoidable for evaluation, only de-identified data were used;
- inference occurred in an access-controlled environment with logging;
- data-retention was disabled per our agreement; the model does not retain or use our prompts or outputs for provider-side training or service improvement.

Author responsibility. Beyond copy-editing and minor figure iconography, LLMs were not used. All ideas, study design, algorithm development, experiments, analyses, and substantive writing were performed by the authors, who take full responsibility for the work and its ethical conduct.

ACKNOWLEDGMENTS

RGK is supported by a Canada CIFAR AI Chair and a Canada Research Chair Tier II in Computational Medicine (CRC-2022-00049). This research was supported by an NFRF Special Call NFRFR2022-00526. Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute.

REFERENCES

Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. Large language models are few-shot clinical information extractors. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 1998–2022, 2022.

Emily Alsentzer, John R. Murphy, Willie Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. In *Proceedings of the 2nd Clinical NLP Workshop at NAACL*, pp. 72–78, 2019.

Ali F. Ansari, Luca Stella, Caglayan Turkmen, Xueyan Zhang, Pedro Mercado, Haoyin Shen, Oleksandr Shchur, Syama S. Rangapuram, Sigberto P. Arango, Somesh Kapoor, et al. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*, 2024.

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

Mehala Balamurali. T-distributed stochastic neighbor embedding. In *Encyclopedia of Mathematical Geosciences*, pp. 1527–1535. Springer, 2023.

Parishad BehnamGhader, Vaibhav Adlakha, Marius Mosbach, Dzmitry Bahdanau, Nicolas Chapados, and Siva Reddy. Llm2vec: Large language models are secretly powerful text encoders, 2024. URL https://arxiv.org/abs/2404.05961.

Dirk Bender and Kamran Sartipi. HL7 FHIR: An agile and restful approach to healthcare information exchange. In *2013 IEEE 26th International Symposium on Computer-Based Medical Systems*, pp. 326–331. IEEE, 2013. doi: 10.1109/CBMS.2013.6627810.

Katherine E. Brown, Chao Yan, Zhuohang Li, Xinmeng Zhang, Benjamin X. Collins, You Chen, Ellen W. Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A. Malin. Large language models are less effective at clinical prediction tasks than locally trained machine learning models. *Journal of the American Medical Informatics Association*, 32(5):811–822, 2025.

Michael Burger, Fedor Sergeev, Martin Londschien, Dominique Chopard, Hugo Yèche, Polina Leshetkina, Andreas Morgenroth, Zeynep Babür, Jasmina Bogojeska, Martin Faltys, Rita Kuznetsova, and Gunnar Rätsch. Towards foundation models for critical care time series. *arXiv preprint arXiv:2411.16346*, 2024.

Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting, 2024. URL https://arxiv.org/abs/2310.04948.

Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, pp. 6776–6786, 2018.

Ching Chang, Wei-Yao Wang, Wen-Chih Peng, and Tien-Fu Chen. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. *ACM Trans. Intell. Syst. Technol.*, February 2025. doi: 10.1145/3719207.

Eunsuk Chang and Sumi Sung. Use of snomed ct in large language models: Scoping review. *JMIR Medical Informatics*, 12:e62924, 2024. doi: 10.2196/62924.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1):6085, 2018.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Miguel Contreras, Sumit Kapoor, Jiaqing Zhang, Andrea Davidson, Yuanfang Ren, Ziyuan Guan, Tezcan Ozrazgat-Baslanti, Jessica Sena, Subhash Nerella, Azra Bihorac, et al. Dellirium: A large language model for delirium prediction in the icu using structured ehr. *Research Square*, pp. rs–3, 2025.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*, 2024.

Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. Gru-ode-bayes: Continuous modeling of sporadic time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 7377–7388, 2019.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pp. arXiv–2407, 2024.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

Len Feremans, Boris Cule, and Bart Goethals. PETSC: Pattern-based embedding for time series classification. *Data Mining and Knowledge Discovery*, 36:1015–1061, 2022. doi: 10.1007/s10618-022-00822-7.

Vincent Fortuin, Dmitry Baranchuk, Gunnar Rätsch, and Stephan Mandt. Gp-vae: Deep probabilistic time series imputation. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Yaroslav Ganin and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Yanjun Gao, Skatje Myers, Shan Chen, Dmitriy Dligach, Timothy A Miller, Danielle Bitterman, Matthew Churpek, and Majid Afshar. When raw data prevails: Are large language model embeddings effective in numerical data representation for medical machine learning applications? *arXiv preprint arXiv:2408.11854*, 2024.

Martin Griot, Charlotte Hemptinne, Jean Vanderdonckt, and Deniz Yuksel. Large language models lack essential metacognition for reliable medical reasoning. *Nature Communications*, 16(1):642, 2025.

Nate Gruver, Marc Finzi, Xuechen Qiu, and Andrew G. Wilson. Large language models are zero-shot time series forecasters. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 19622–19635, 2023.

Mehak Gupta, Brian M. Gallamoza, Nicholas Cutrona, Prakash Dhakal, Romain Poulain, and Reza Beheshti. An extensive data processing pipeline for mimic-iv. In *Proceedings of Machine Learning for Health (ML4H) at NeurIPS*, pp. 311–325, 2022.

Thomas Hartvigsen, Walter Gerych, Jidapa Thadajarassiri, Xiangnan Kong, and Elke Rundensteiner. Stop&hop: Early classification of irregular time series. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 696–705, 2022. doi: 10.1145/3511808.3557460.

Hrayr Harutyunyan, Hrant Khachatrian, David C. Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1):96, 2019.

Rui He and Jason N. Chiang. Simultaneous forecasting of vital sign trajectories in the icu. *Scientific Reports*, 15:14996, 2025.

Stefan Hegselmann, Alejandro Buendia, Hunter Lang, Monica Agrawal, Xiaoyi Jiang, and David Sontag. Tabllm: Few-shot classification of tabular data with large language models. In *International conference on artificial intelligence and statistics*, pp. 5549–5581. PMLR, 2023.

Stefan Hegselmann, Georg von Arnim, Tillmann Rheude, Noel Kronenberg, David Sontag, Gerhard Hindricks, Roland Eils, and Benjamin Wild. Large language models are powerful ehr encoders. *arXiv e-prints*, pp. arXiv–2502, 2025.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Kyunghoon Hur, Jungwoo Oh, Junu Kim, Jiyoun Kim, Min Jae Lee, Eunbyeol Cho, Seong-Eun Moon, Young-Hak Kim, Louis Atallah, and Edward Choi. Genhpf: General healthcare predictive framework for multi-task multi-source learning. *IEEE Journal of Biomedical and Health Informatics*, 28(1):502–513, January 2024. ISSN 2168-2208. doi: 10.1109/jbhi.2023.3327951. URL http://dx.doi.org/10.1109/JBHI.2023.3327951.

Stephanie L. Hyland, Martin Faltys, Matthias Hüser, Xinrui Lyu, Thomas Gumbsch, Cristóbal Esteban, Christian Bock, Michael Moor, Gunnar Rätsch, Finale Doshi-Velez, and Thomas Merz. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, 26(3):364–373, 2020.

Zongliang Ji, Anna Goldenberg, and Rahul G Krishnan. Measurement scheduling for icu patients with offline reinforcement learning. *arXiv preprint arXiv:2402.07344*, 2024.

Zongliang Ji, Andre Carlos Kajdacsy-Balla Amaral, Anna Goldenberg, and Rahul G Krishnan. Exosito: Explainable off-policy learning with side information for intensive care unit blood test orders. *arXiv preprint arXiv:2504.17277*, 2025.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-llm: Time series forecasting by reprogramming large language models, 2024. URL https://arxiv.org/abs/2310.01728.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Marzyeh Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):160035, 2016.

Alistair E. W. Johnson, Lucas Bulgarelli, Li-wei Shen, Anthony Gayles, Amogh Shammout, Steven Horng, Tom J. Pollard, Sharon Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1–11, 2023a.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, 2023b.

Witesyavwirwa Vianney Kambale, David Krame Kadurha, Ali Deeb, Fadi Al Machot, Taha Bernabia, and Kyandoghere Kyamakya. Transformers in time series forecasting: A brief analysis of the autoformer transfer learning performance. In *2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 1–6, 2023. doi: 10.1109/ICECS58634.2023.10382781.

Justin Kauffman, Riccardo Miotto, Eyal Klang, Anthony Costa, Beau Norgeot, Marinka Zitnik, Shameer Khader, Fei Wang, Girish N. Nadkarni, and Benjamin S. Glicksberg. Embedding methods for electronic health record research. *Annual Review of Biomedical Data Science*, 8(1):563–590, 2025. doi: 10.1146/annurev-biodatasci-103123-094729.

Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 6696–6707, 2020.

Rita Kuznetsova, Alizée Pace, Manuel Burger, Hugo Yèche, and Gunnar Rätsch. On the importance of step-wise embeddings for heterogeneous clinical time-series. In *Machine Learning for Health (ML4H)*, pp. 268–291. PMLR, 2023.

Hunter Lang, Monica N Agrawal, Yoon Kim, and David Sontag. Co-training improves prompt-based learning for large language models. In *International Conference on Machine Learning*, pp. 11985–12003. PMLR, 2022.

Jinhyuk Lee, Zhuyun Dai, Xiaoqi Ren, Blair Chen, Daniel Cer, Jeremy R. Cole, Kai Hui, Michael Boratko, Rajvi Kapadia, Wen Ding, Yi Luan, Sai Meher Karthik Duddu, Gustavo Hernandez Abrego, Weiqiang Shi, Nithi Gupta, Aditya Kusupati, Prateek Jain, Siddhartha Reddy Jonnalagadda, Ming-Wei Chang, and Iftekhar Naim. Gecko: Versatile text embeddings distilled from large language models, 2024. URL https://arxiv.org/abs/2403.20327.

Simon A Lee, Sujay Jain, Alex Chen, Kyoka Ono, Arabdha Biswas, Ákos Rudas, Jennifer Fang, and Jeffrey N Chiang. Clinical decision support using pseudo-notes from multiple streams of ehr data. *npj Digital Medicine*, 8(1):394, 2025.

Hailin Li. Multivariate time series clustering based on common principal component analysis. *Neurocomputing*, 349:239–247, 2019. doi: 10.1016/j.neucom.2019.03.060.

Yikuan Li, Shishir Rao, Jose Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Behrt: Transformer for electronic health records. *Scientific Reports*, 10(1):7155, 2020a.

Yikuan Li, Sudha Rao, Jose R. A. Solares, Abdelkader Hassaine, Rema Ramakrishnan, Dexter Canoy, Ying Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Behrt: Transformer for electronic health records. *Scientific Reports*, 10(1):7155, 2020b. doi: 10.1038/s41598-020-62922-y.

Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning, 2024a. URL https://arxiv.org/abs/2403.07300.

Pengjie Liu, Huazheng Guo, Tongtong Dai, Ninghao Li, Jianfeng Bao, Xiang Ren, Yifan Jiang, and Shu-Tao Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 2025.

Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jiening Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. Large language models are few-shot health learners, 2023. URL https://arxiv.org/abs/2305.15525.

Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Autotimes: Autoregressive time series forecasters via large language models. *arXiv preprint arXiv:2402.02370*, 2024b.

Vadim Liventsev and Tobias Fritz. Intensive care as one big sequence modeling problem. *arXiv preprint arXiv:2402.17501*, 2024.

Renqian Luo, Luyu Sun, Yuxian Xia, Tao Qin, Shuo Zhang, Hoifung Poon, and Tie-Yan Liu. Biogpt: Generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, 23(6):bbac409, 2022.

Yuanfang Luo, Samantha Henry, Yanshan Wang, Feifan Shen, Ozlem Uzuner, and Anna Rumshisky. The 2019 n2c2/umass lowell shared task on clinical concept normalization. *Journal of the American Medical Informatics Association*, 27(8):1529–1537.e1, 2020. doi: 10.1093/jamia/ocaa166.

Matthew B. A. McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, and Marzyeh Ghassemi. Reproducibility in machine learning for health research: still a ways to go. *Science Translational Medicine*, 13(586):eabb1655, 2021.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013. URL https://arxiv.org/abs/1301.3781.

Michael Moor, Oishi Banerjee, Zeshan S. H. Abad, Harlan M. Krumholz, Jure Leskovec, Eric J. Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023.

Mohammad Amin Morid, Olivia R. Liu Sheng, and Joseph Dunbar. Time series prediction using deep learning methods in healthcare. *CoRR*, abs/2108.13461, 2021. URL https://arxiv.org/abs/2108.13461.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*, 2022.

Harsha Nori, Nicholas King, Scott M. McKinney, David Carignan, and Eric Horvitz. Capabilities of gpt-4 on medical challenge problems, 2023.

A. Núñez Reiz, M. A. Armengol de la Hoz, and M. Sánchez García. Big data analysis and machine learning in intensive care units. *Medicina Intensiva (English Edition)*, 43(7):416–426, 2019.

Mathieu Oliver, Jérôme Allyn, Romain Carencotte, Nadhira Allou, and Cyril Ferdynus. Introducing the blendedicu dataset, the first harmonized, international intensive care dataset. *Journal of Biomedical Informatics*, 146:104502, 2023.

Srikant Panda, Hitesh Laxmichand Patel, Shahad Al-Khalifa, Amit Agarwal, Hend Al-Khalifa, and Sharefah Al-Ghamdi. Daiq: Auditing demographic attribute inference from question in llms. *arXiv preprint arXiv:2508.15830*, 2025.

Chengpeng Peng, Xiaoman Yang, Aoxiao Chen, et al. A study of generative large language model for medical research and healthcare. *npj Digital Medicine*, 6:210, 2023. doi: 10.1038/s41746-023-00958-w.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proc. EMNLP*, pp. 1532–1543, 2014. doi: 10.3115/v1/D14-1162.

Tom J. Pollard, Alistair E. W. Johnson, Jesse D. Raffa, Leo Anthony Celi, Roger G. Mark, and Omar Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific Data*, 5(1):1–13, 2018.

Alvin Rajkomar, Eyal Oren, Kai Chen, Peng Dai, Naim Hajaj, Peter J. Liu, Melody Sun, Pushpak Sharma, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1:18, 2018.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*, 4(1):86, 2021.

Christian Reich, Anna Ostropolets, Patrick Ryan, Peter Rijnbeek, Martijn Schuemie, Alexander Davydov, Dmitry Dymshyts, and George Hripcsak. Ohdsi standardized vocabularies—a large-scale centralized reference ontology for international data harmonization. *Journal of the American Medical Informatics Association*, 31(3):583–590, 2024. doi: 10.1093/jamia/ocad247.

Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations (ICLR)*, 2020. URL https://openreview.net/forum?id=ryxGuJrFvS.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cían Hughes, Charles Lau, et al. Medgemma technical report. *arXiv preprint arXiv:2507.05201*, 2025.

Zineb Senane, Lele Cao, Valentin Leonhard Buchner, Yusuke Tashiro, Lei You, Pawel Andrzej Herman, Mats Nordahl, Ruibo Tu, and Vilhelm Von Ehrenheim. Self-supervised learning of time series representation via diffusion process and imputation-interpolation-forecasting mask. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2560–2571, 2024.

Benjamin Shickel, Patrick J. Tighe, Azra Bihorac, and Parisa Rashidi. Deep ehr: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5):1589–1604, 2018.

Satya N. Shukla and Benjamin M. Marlin. Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations (ICLR)*, 2021.

Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3285–3292, 2019. doi: 10.1109/BigData47090.2019.9005997.

Karan Singhal, Shekoofeh Azizi, Tao Tu, et al. Large language models encode clinical knowledge. *Nature*, 620:172–180, 2023. doi: 10.1038/s41586-023-06291-2.

Karan Singhal, Shekoofeh Azizi, Tao Tu, et al. Toward expert-level medical question answering with large language models. *Nature Medicine*, 30:225–234, 2024. doi: 10.1038/s41591-024-03423-7.

Congzheng Song and Ananth Raghunathan. Information leakage in embedding models. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pp. 377–390. ACM, 2020. doi: 10.1145/3372297.3417270.

Paul E. Stang, Patrick B. Ryan, Judith A. Racoosin, J. Marc Overhage, Abraham G. Hartzema, Christian Reich, Emily Welebob, Thomas Scarnecchia, and Janet Woodcock. Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership. *Annals of Internal Medicine*, 153(9):600–606, 2010. doi: 10.7326/0003-4819-153-9-201011020-00010.

Ethan Steinberg, Kenneth Jung, Jason A. Fries, Conor K. Corbin, Stephen R. Pfohl, and Nigam H. Shah. Language models are an effective representation learning technique for electronic health record data. *Journal of Biomedical Informatics*, 113:103637, 2020. doi: 10.1016/j.jbi.2020.103637.

Hongjin Su, Weijia Shi, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A Smith, Luke Zettlemoyer, and Tao Yu. One embedder, any task: Instruction-finetuned text embeddings. *arXiv preprint arXiv:2212.09741*, 2022.

Adarsh Subbaswamy, Bryant Chen, and Suchi Saria. A unifying causal framework for analyzing dataset shift-stable learning algorithms. *Journal of Causal Inference*, 10(1):64–89, 2022.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision – ECCV 2016 Workshops*, volume 9915 of *Lecture Notes in Computer Science*, pp. 443–450, Cham, 2016. Springer. doi: 10.1007/978-3-319-49409-8_35.

Chenxi Sun, Hongyan Li, Yaliang Li, and Shenda Hong. Test: Text prototype aligned embedding to activate llm's ability for time series, 2024. URL https://arxiv.org/abs/2308.08241.

Harini Suresh, Nathan Hunt, Alistair Johnson, Leo Anthony Celi, Peter Szolovits, and Marzyeh Ghassemi. Clinical intervention prediction and understanding with deep networks. In *Proceedings of the Machine Learning for Healthcare Conference (MLHC)*, pp. 322–337, 2017.

Parinaz Tabari, Gennaro Costagliola, Mattia De Rosa, and Martin Boeker. State-of-the-art fast healthcare interoperability resources (fhir)–based data model and structure implementations: Systematic scoping review. *JMIR medical informatics*, 12(1):e58445, 2024.

Sabera Talukder, Yisong Yue, and Georgia Gkioxari. Totem: Tokenized time series embeddings for general time series analysis, 2025. URL https://arxiv.org/abs/2402.16412.

Mingtian Tan, Mike A Merrill, Vinayak Gupta, Tim Althoff, and Tom Hartvigsen. Are language models actually useful for time series forecasting? *arXiv preprint arXiv:2401.12345*, 2024.

Qingxiong Tan, Mang Ye, Andy Jinhua Ma, Baoyao Yang, Terry Cheuk-Fung Yip, Grace Lai-Hung Wong, and Pong C. Yuen. Explainable uncertainty-aware convolutional recurrent neural network for irregular medical time series. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10):4665–4679, 2021. doi: 10.1109/TNNLS.2020.3025813.

Liyan Tang, Zhaoyi Sun, Betina Idnay, Jordan G. Nestor, Ali Soroush, Pierre A. Elias, Ziyang Xu, Ying Ding, Greg Durrett, Justin F. Rousseau, Nigam H. Shah, and Fei Wang. Evaluating large language models on medical evidence summarization. *npj Digital Medicine*, 6(1):158, 2023. doi: 10.1038/s41746-023-00896-7.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 24804–24816, 2021.

Peter J. Thoral, Jan M. Peppink, Rutger H. Driessen, Eric J. G. Sijbrands, Erwin J. O. Kompanje, Lewis J. Kaplan, Helen Bailey, Jozef Kesecioglu, Maurizio Cecconi, Matthew Churpek, and Paul Elbers. Sharing icu patient data responsibly under the sccm/esicm joint initiative: Amsterdamumcdb. *Critical Care*, 25(1):307, 2021.

Nenad Tomasev, Xavier Glorot, Jack W. Rae, Michal Zielinski, Harry Askham, Allan Saraiva, Kayleigh Hajewski, and et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572(7767):116–119, 2019.

Patara Trirat, Yooju Shin, Junhyeok Kang, Youngeun Nam, Jihye Na, Minyoung Bae, Joeun Kim, Byunghyun Kim, and Jae-Gil Lee. Universal time-series representation learning: A survey, 2024. URL https://arxiv.org/abs/2401.03717.

Robin van de Water, Hendrik Schmidt, Paul Elbers, Patrick Thoral, Bert Arnrich, and Patrick Rockenschaub. Yet another icu benchmark: A flexible multi-center framework for clinical ml, 2024. URL https://arxiv.org/abs/2306.05109.

Dave Van Veen, Chris Van Uden, Leon Blankemeier, Jean-Benoit Delbrouck, Amin Aali, Christian Bluethgen, and Ankit Pareek. Adapted large language models can outperform medical experts in clinical text summarization, 2023.

Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerová, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, John Pauly, and Akshay S. Chaudhari. Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine*, 30(4):1134–1142, 2024. doi: 10.1038/s41591-024-02855-5.

Carina Nina Vorisek, Moritz Lehne, Sophie Anne Ines Klopfenstein, Paula Josephine Mayer, Alexander Bartschke, Thomas Haese, and Sylvia Thun. Fast healthcare interoperability resources (fhir) for interoperability in health research: systematic review. *JMIR medical informatics*, 10(7): e35724, 2022.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Improving text embeddings with large language models, 2024a. URL https://arxiv.org/abs/2401.00368.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. *arXiv preprint arXiv:2405.14616*, 2024b.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven C. H. Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2022.

Chaoyi Wu, Peng Qiu, Jing Liu, Hao Gu, Ning Li, Yiyuan Zhang, Yixuan Wang, and Weidi Xie. Towards evaluating and building versatile large language models for medicine. *npj Digital Medicine*, 8(1):58, 2025.

Xiaoyang Yang, Aonan Chen, Neda PourNejatian, et al. A large language model for electronic health records. *npj Digital Medicine*, 5:194, 2022. doi: 10.1038/s41746-022-00742-2.

Hugo Yèche, Rita Kuznetsova, Marc Zimmermann, Matthias Hüser, Xinrui Lyu, Martin Faltys, and Gunnar Rätsch. Hirid-icu-benchmark: A comprehensive machine learning benchmark on high-resolution icu data. In *NeurIPS 2021 Datasets and Benchmarks Track*, 2021.

Hugo Yèche, Rita Kuznetsova, Marc Zimmermann, Matthias Hüser, Xinrui Lyu, Martin Faltys, and Gunnar Rätsch. Hirid-icu-benchmark: A comprehensive machine learning benchmark on high-resolution icu data. In *NeurIPS Datasets and Benchmarks*, 2021.

Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. Time-series generative adversarial networks. *Advances in neural information processing systems*, 32, 2019.

Zhihan Yue, Yue Wang, Yujing Sun, Yujin Zheng, Qing He, Jian Zeng, and Rong Jin. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8982–8990, 2022.

Hugo Yèche, Rita Kuznetsova, Marc Zimmermann, Matthias Hüser, Xinrui Lyu, Martin Faltys, and Gunnar Rätsch. Hirid-icu-benchmark – a comprehensive machine learning benchmark on high-resolution icu data, 2022. URL https://arxiv.org/abs/2111.08536.

Jamil Zaghir, Marco Naguib, Mina Bjelogrlić, Aurélie Névéol, Xavier Tannier, and Christian Lovis. Prompt engineering paradigms for medical applications: Scoping review and recommendations for better practices. *Journal of Medical Internet Research*, 26:e60501, 2024. doi: 10.2196/60501.

Georgios Zerveas, Srivatsan Jayaraman, Dheevatsa Patel, Arshan Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. *arXiv preprint arXiv:2010.02803*, 2021.

Xi Zhang, Yuan Pu, Yuki Kawamura, Andrew Loza, Yoshua Bengio, Dennis L. Shung, and Alexander Tong. Trajectory flow matching with applications to clinical time series modelling. In *Advances in Neural Information Processing Systems*, volume 37, pp. 107198–107224, 2024a. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/c1e67cde895c3c91edb43569ad0df260-Paper-Conference.pdf.

Xiyuan Zhang, Ranak R. Chowdhury, Rajesh K. Gupta, and Jingbo Shang. Large language models for time series: A survey. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024b.

Tian Zhou, PeiSong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all:power general time series analysis by pretrained lm, 2023. URL https://arxiv.org/abs/2302.11939.

Zhihao Zhu, Ninglu Shao, Defu Lian, Chenwang Wu, Zheng Liu, Yi Yang, and Enhong Chen. Understanding privacy risks of embeddings induced by large language models. *arXiv preprint arXiv:2404.16587*, 2024.

A USAGE OF LARGE LANGUAGE MODELS

We used large language models (LLMs) in two strictly limited ways:

1. **Locally hosted models.** For development convenience, we ran LLMs on institution-managed GPUs within access-controlled, offline (or firewall-restricted) servers. These models were used only for debugging assistance and rapid prototyping under our secure computing environment; no research data were exported outside these servers.
2. **Third-party services for copy-editing and minor figure elements.** We used the Gemini and ChatGPT families solely to polish grammar and wording of author-written text and to generate small iconography embedded within figures (e.g., simple symbols), *not* full figures and not scientific content. We did not use these services to generate methods, ideas, analyses, code, or experiments. No sensitive or identifying data were included in prompts. All suggested edits were manually reviewed by the authors.

Beyond the above, LLMs were *not* involved in the conception of ideas, study design, algorithm development, experiments, analyses, or writing of substantive content. All contributions—conceptual, methodological, experimental, and explanatory—are entirely by the authors.

B EXTENDED RELATED WORK

Privacy and attribute inference in embeddings. Learned representations may inadvertently reveal sensitive attributes. Recent analyses show that embeddings can leak information via inversion, membership, or attribute–inference attacks, even when protected features are not explicit (Song & Raghunathan, 2020; Zhu et al., 2024). Emerging auditing frameworks evaluate demographic attribute inference from model outputs and intermediate representations (Panda et al., 2025). We therefore adopt demographic recoverability (age and sex) from learned embeddings as a practical proxy for privacy risk, situating our evaluation alongside this auditing literature.

ICU Time Series and Clinical Benchmarks Public ICU datasets drove clinical ML, notably MIMIC and eICU (Johnson et al., 2016; Pollard et al., 2018; Johnson et al., 2023b). Early pipelines were bespoke (Harutyunyan et al., 2019; McDermott et al., 2021), prompting standardized benchmarks: canonical MIMIC tasks (Harutyunyan et al., 2019), the HiRID ICU benchmark with high frequency signals and evaluation code (Yèche et al., 2021), and YAIB harmonizing MIMIC, eICU, HiRID, and AmsterdamUMCdb (van de Water et al., 2024; Thoral et al., 2021). Aggregations such as BlendedICU broaden multi center evaluation (Oliver et al., 2023), and foundation pretraining over heterogeneous ICU series is emerging (Burger et al., 2024). Deep sequential models improve early warning and trajectory modeling (Hyland et al., 2020; Tomasev et al., 2019; Ji et al., 2024; 2025), and EHR sequence models reach expert level outcomes (Rajkomar et al., 2018). Persistent challenges include explainability, irregular sampling, and missingness (Morid et al., 2021), addressed by interpretable modeling and transformer forecasting in clinical settings (Tan et al., 2021; Hartvigsen et al., 2022; Zhang et al., 2024a). Surveys stress the influence of curation, task design, and preprocessing (Núñez Reiz et al., 2019; Shickel et al., 2018; Suresh et al., 2017). Recent pipelines underscore standardization needs (Gupta et al., 2022; He & Chiang, 2025).

Imputation and Representation Learning for Time Series. Irregular sampling and missingness are central in clinical time series. Model based imputers encode absence patterns or learn imputation with prediction (Che et al., 2018; Cao et al., 2018), while generative methods improve accuracy through diffusion and realistic dynamics (Tashiro et al., 2021; Yoon et al., 2019). Continuous time and structure aware encoders address nonuniform sampling and uncertainty (Fortuin et al., 2020; Rubanova et al., 2019; De Brouwer et al., 2019; Kidger et al., 2020; Shukla & Marlin, 2021). Self supervised objectives and transformer encoders learn general purpose representations (Eldele et al., 2021; Yue et al., 2022; Woo et al., 2022; Zerveas et al., 2021; Franceschi et al., 2019). Diffusion style self supervision couples imputation, interpolation, and forecasting masks for transferable embeddings (Senane et al., 2024). Decoder only foundation pretraining yields strong forecasting across domains (Das et al., 2024). Earlier work spans dimensionality reduction and deep recurrent networks (Li, 2019; Balamurali, 2023; Siami-Namini et al., 2019). Recent ideas explore discrete tokenization and approximating LLM embedding spaces (Talukder et al., 2025; Sun et al., 2024). In healthcare, pattern based embeddings and semantic grouping improve interpretability (Feremans et al., 2022;

Kuznetsova et al., 2023). Surveys advocate universal representations robust to noise, sparsity, and shift (Trirat et al., 2024).

Large Language Models in Healthcare and Time Series. Large language models have advanced clinical NLP and representation learning. Domain pretrained transformers on coded EHRs improve disease prediction (Li et al., 2020a; Rasmy et al., 2021), clinical scale models achieve strong extraction and inference (Yang et al., 2022; Peng et al., 2023), and general medical systems perform well on exams and challenges (Singhal et al., 2023; 2024; Nori et al., 2023). Early adaptations demonstrated clinical question answering and discharge summarization (Agrawal et al., 2022; Lang et al., 2022; Van Veen et al., 2023). For patient level prediction, audits report underperformance and safety concerns (Brown et al., 2025; Wu et al., 2025; Griot et al., 2025; Moor et al., 2023). Bridging modalities includes in modality pretraining, cross modality transfer and reprogramming, prompt tuning, cross modal fine tuning, autoregressive prediction, and alignment with language space (Kambale et al., 2023; Zhou et al., 2023; Chang et al., 2025; Jin et al., 2024; Cao et al., 2024; Liu et al., 2024a;b; 2025). Textualization strategies often trail specialized forecasters (Ansari et al., 2024; Gruver et al., 2023; Tan et al., 2024). Few shot models show gains, and surveys emphasize formatting, scaling, and evaluation (Liu et al., 2023; Zhang et al., 2024b; Liventsev & Fritz, 2024). Advances in embeddings and domain specific language models motivate LLM derived representations for ICU data (Pennington et al., 2014; Mikolov et al., 2013; Reimers & Gurevych, 2019; Su et al., 2022; Muennighoff et al., 2022; Lee et al., 2024; BehnamGhader et al., 2024; Wang et al., 2024a; Luo et al., 2022; Alsentzer et al., 2019).

Recent studies have also investigated serializing structured EHR data for LLM processing. Approaches like TabLLM (Hegselmann et al., 2023) and DeLLiriuM (Contreras et al., 2025) fine-tune LLMs on serialized records for tabular classification or specific risk predictions. Others, such as Lee et al. (2025), Gao et al. (2024), and Hegselmann et al. (2025), explore using frozen LLMs to embed raw data serializations directly. While these methods demonstrate the utility of LLMs for medical data, they primarily focus on sparse longitudinal records or single-site benchmarks. Our work distinguishes itself by addressing dense, irregular ICU time-series, where we find that direct serialization yields excessive sequence lengths and brittleness to distribution shifts, necessitating a summarization step for effective cross-site portability.

Additionally, domain-specific foundation models like Med-BERT (Rasmy et al., 2021) and GenHPF (Hur et al., 2024) apply Transformer-based pretraining to structured EHR codes. While effective for longitudinal disease prediction within a specific health system, these models rely on fixed vocabularies that can be sensitive to schema variations and distribution shifts across institutions. Furthermore, they often struggle to capture the high-frequency numeric dynamics of intensive care data compared to generalist time-series models or semantic LLM representations, which offer greater flexibility and transferability in diverse deployment settings.

C Dataset and processing

C.1 General information

MIMIC (80,749 samples). The processed ICU time series comprises 60 features: `ALP`, `ALT`, `Bicarbonate`, `Bilirubin`, `BloodUreaNitrogen`, `Calcium`, `Creatinine`, `CreatinineKinase`, `Hemoglobin`, `INR`, `Lactate`, `Magnesium`, `PaCO2`, `PaO2`, `Phosphate`, `Platelets`, `Potassium`, `Sodium`, `Troponin`, `WBC`, `ph`, `AirwayPressure`, `DiastolicBloodPressure`, `FiO2`, `GCS`, `HeartRate`, `ICDSC`, `MeanBloodPressure`, `MinuteVentilation`, `PEEP`, `RespiratoryRate`, `SAS`, `SystolicBloodPressure`, `Temperature`, `TidalVolume`, `UrineOutput`, `Analgesia`, `Antiarrhythmics`, `Antibiotics`, `Anticoagulants`, `Antiepileptics`, `Antihypertensives`, `Antipsychotics`, `CTScan`, `CaReplacement`, `Dialysis`, `Diuretics`, `EnteralNutrition`, `ICPMonitor`, `KReplacement`, `MRI`, `MgReplacement`, `PPI`, `Paralysis`, `TPN`, `Transfusions`, `UltraSound`, `Vasopressors`, `Ventilation`, `Xray`.

HiRID (36,019 samples). The processed ICU time series comprises 64 features: `ALP`, `ALT`, `AST`, `Bicarbonate`, `Bilirubin`, `BloodUreaNitrogen`, `Calcium`, `Chloride`, `CreatineKinase`, `Creatinine`, `Glucose`, `Hemoglobin`, `INR`, `Lactate`, `Magnesium`, `PaCO2`, `PaO2`, `Phosphate`, `Platelets`, `Potassium`, `Sodium`, `Troponin`, `WBC`, `ph`,

AirwayPressure, AirwayPressurePeak, DiastolicBloodPressure, FiO2, FluidBalance, GCS, HeartRate, MeanBloodPressure, PEEP, RespiratoryRate, SAS, Saturation, SystolicBloodPressure, Temperature, TidalVolume, UrineOutput, Analgesia, Antiarrhythmics, Antibiotics, Anticoagulants, Antiepileptics, Antihypertensives, Aspirin, CaReplacement, Dialysis, Diuretics, ICPMonitor, Insulin, KReplacement, LiverToxicDrug, MgReplacement, Neuroleptics, Paralysis, Saline, Sedation, Steroids, TPN, Transfusions, Vasopressors, Ventilation.

PPICU (47,119 samples). The processed ICU time series comprises 75 features: ALP, ALT, AST, Bicarbonate, Bilirubin, BloodUreaNitrogen, Calcium, Chloride, Creatinine, CreatinineKinase, GGT, Glucose, Hemoglobin, INR, Lactate, Magnesium, PaCO2, PaO2, Phosphate, Platelets, Potassium, Sodium, Troponin, WBC, ph, AirwayPressure, AirwayPressureIP, AirwayPressurePeak, DiastolicBloodPressure, FiO2, FluidInput, FluidOutput, GCS, HeartRate, ICDSC, MeanBloodPressure, MinuteVentilation, PAVSupport, PC, PEEP, PlateauPressure, RespiratoryRate, SAS, Saturation, SystolicBloodPressure, Temperature, TidalVolume, UrineOutput, Analgesia, Antiarrhythmic, Antiarrhythmics, Antibiotics, Anticoagulants, Antiepileptics, Antihypertensives, Aspirin, Dialysis, Diuretics, EKG, EVD, ICPMonitor, InhaledVasodilator, Insulin, Isoproterenol, MgReplacement, Neuroleptics, OsmoticTherapy, PPI, PS, Paralysis, Sedation, Steroids, TPN, Transfusions, Vasopressors.

C.2 Data Preprocessing Details

We provide reproducible preprocessing pipelines for all ICU datasets considered in this work. This appendix summarizes the steps at a high level, without referencing specific files or directories.

Common setup. We organize raw and processed data under a central data directory. Access to each dataset follows the relevant data use agreements. Unless stated otherwise, per-stay sequences are constructed at hourly resolution, and features are normalized using statistics computed on the training split. All clinical features are selected in discussion with experienced practicing ICU clinicians.

HiRID

We first obtain authorized access and download the raw HiRID release. Because HiRID associates each patient with a single ICU stay, we group all clinical and medication records by patient identifier and construct a sparse, event-centric representation per patient. We then curate clinically meaningful features and extend the feature dictionary to include start/end times and summary statistics for quality control and normalization. Using this curated set, we produce patient-level targets and serialize time indices, value indices, and per-event counts for downstream modeling. Finally, we standardize signals and combine them on a per-feature basis to ensure consistent scaling across patients and over time.

MIMIC-IV

We begin by obtaining credentialed access and downloading the raw MIMIC-IV data, then convert the raw tables into a columnar layout to improve I/O efficiency. Next, we filter to ICU stays that exceed a minimum duration threshold, retaining this subset for further processing. In collaboration with clinicians, we develop a mapping from measurement identifiers to a study feature set; this mapping is intentionally decoupled so it can be revised without reprocessing earlier stages. We restrict event tables to the curated features and carry forward only necessary metadata into a project-specific processed area. Core administrative and ICU stay tables are merged to derive age, gender, mortality, and length-of-stay labels, excluding pediatric subjects for adult-only cohorts. We then materialize a subject-centric layout (diagnoses, events, procedures, stays), preferring numeric event values where available, and write feature-separated storage to enable efficient access and validation across tasks.

PPICU

For PPICU, we convert all raw sources into a uniform, analysis-friendly tabular format and separate multi-signal streams into per-feature tables for efficient access; laboratory measurements are processed analogously. We merge demographic and outcome information to construct a source stay table with identifiers and key labels (age, gender, length of stay, mortality). The feature space is restricted to the analysis cohort, with per-feature summary statistics computed for clinician review; clinician-verified filters are then applied to remove implausible or out-of-range values. From the merged feature data, we generate per-stay tables sorted by measurement time and align feature names to a harmonized schema. A global time index keyed by stay identifier is built with arrays of time points at the chosen resolution. Each stay is finally converted into both a sparse dictionary-backed representation and a dense, zero-filled discrete representation, and we emit the resulting artifacts alongside train/validation/test splits.

D Training Details

D.1 System Configuration

All experiments were run on NVIDIA L40S GPUs. Unless noted, training and inference can be paused and fully completed on a single L40S (24 GB); large-scale reference runs used 4×L40S for throughput.

D.2 TSDE Embedding Extraction

The TSDE model is trained in a self-supervised manner on the same training split as downstream tasks. We mask the final 24 h of each sequence and train the model to predict the masked segment. After convergence, we extract the hidden representation as the TSDE embedding for each record.

D.3 TFM Embedding Extraction

We use the time-series foundation model `google/timesfm-1.0-200m`. We run inference on the full HiRID/MIMIC/PPICU datasets and extract representations from the last two transformer blocks. The penultimate layer yields consistently better downstream performance and is used in all main results.

D.4 Hyperparameter Details

For baselines, we tune batch size and learning rate with early stopping and report the best configuration per method:

- **right_shift:** batch size 512, learning rate 1×10^{-5}
- **mean:** batch size 512, learning rate 4×10^{-5}
- **interpolation:** batch size 256, learning rate 5×10^{-5}

Record2Vec models are trained under a common setting: batch size 512, learning rate 1×10^{-5}. All methods are trained with seeds $\{42, 84, 1005, 2025\}$; we report the mean and standard deviation across seeds.

D.5 Few-shot Finetuning (16-sample setting)

We study an extreme low-data regime with only 16 labeled target samples and no held-out validation set. Finetuning is performed with AdamW and a warmup–cosine schedule; the learning rate scales with the effective batch size:

$$\text{LR}_{\text{scaled}} = \text{BASE\_LR} \times \frac{\text{batch\_size}}{\text{REF\_FINETUNE\_BS}}, \qquad \text{BASE\_LR} = 1\times10^{-6},\ \text{REF\_FINETUNE\_BS} = 16.$$

Weight decay is 0.01 with $\beta_1 = 0.9,\ \beta_2 = 0.999$. Let S be the total number of optimization steps ($S = \text{num\_epochs} \times \text{steps\_per\_epoch}$) and $S_{\text{w}} = \max\{10, \lfloor 0.03\, S \rfloor\}$ the warmup steps. In the final

experiment, $S = 100 \times 16$. The per-step multiplier is

$$\lambda(t) = \begin{cases} \dfrac{t+1}{S_{\mathrm{w}}}, & t < S_{\mathrm{w}}, \\ \mathrm{MIN\_LR\_RATIO} + \big(1 - \mathrm{MIN\_LR\_RATIO}\big)\dfrac{1 + \cos\!\big(\pi\, \frac{t - S_{\mathrm{w}}}{S - S_{\mathrm{w}}}\big)}{2}, & t \geq S_{\mathrm{w}}, \end{cases}$$

with MIN_LR_RATIO $= 0.10$. Unless otherwise specified, weight updates apply to both the backbone and the task decoder.

Concrete finetuning setup (Shared across all methods).

- **Batch size:** batch_size $= 16$, $\mathrm{LR}_{\text{scaled}} = 1 \times 10^{-6}$.
- **Optimizer:** AdamW(lr $= \mathrm{LR}_{\text{scaled}}$, weight_decay $= 0.01$, $\beta = (0.9, 0.999)$).
- **Scheduler:** Warmup (first $\max\{10, 0.03S\}$ steps) then cosine decay to $0.1\times$ the peak LR.

E Prompts used for LLM Summarization and a Case Study

In RQ5, we discussed the effect of prompts on the quality of embedding. The patient data is summarized on a split-by-feature manner, raw text as following: `ALP has value 66 in hour 31. ALT has value 22 in hour 31. AST has value 17 in hour 31. Bilirubin has value 4 in hour 31. Blood Urea Nitrogen has value 5.6 in hour 31. Calcium has value 1.99 in hour 31. Chloride has value 115 in hour 31. Creatinine has value 103 in hour 31. Creatinine Kinase has value 316 in hour 31. Hemoglobin has value 134 in hour 31. INR has value 0.99 in hour 31. Lactate has value 1.4 in hour 31. Magnesium has value 1.03 in hour 31. Phosphate has value 0.78 in hour 31. Platelets has value 303 in hour 31. Potassium has value 3.4 in hour 31. Sodium has value 150 in hour 31. Troponin has value 8 in hour 31. WBC has value 8.4 in hour 31. Airway Pressure has value 7.36 in hour 31, value 7.21 in hour 32, value 7.13 in hour 33, value 7.1 in hour 34, value 7.36 in hour 35, value 6.29 in hour 36, value 5.08 in hour 37, value 5.3 in hour 38. Diastolic Blood Pressure has value 59 in hour 31, value 56 in hour 32, value 77 in hour 33, value 80 in hour 34, value 75 in hour 35, value 82 in hour 36, value 75 in hour 37, value 78 in hour 38, value 69 in hour 39, value 68 in hour 40, value 59 in hour 41, value 67 in hour 42, value 65 in hour 44. FiO2 has value 30 in hour 31, value 30 in hour 32, value 30 in hour 33, value 30 in hour 34, value 30 in hour 35, value 30 in hour 36, value 30 in hour 37, value 30 in hour 38, value 21 in hour 39, value 21 in hour 40, value 21 in hour 41, value 21 in hour 42, value 21 in hour 43, value 21 in hour 44. FluidInput has value 80 in hour 34, value 160 in hour 39. FluidOutput has value 100 in hour 39. GCS has value 8 in hour 32, value 8 in hour 33, value 8 in hour 34, value 8 in hour 35, value 7 in hour 36, value 8 in hour 37, value 14 in hour 39, value 14 in hour 40, value 14 in hour 44. Heart Rate has value 97 in hour 31, value 101 in hour 32, value 94 in hour 33, value 93 in hour 34, value 99 in hour 35, value 104.5 in hour 36, value 105 in hour 37, value 102 in hour 38, value 109 in hour 39, value 107 in hour 40, value 106 in hour 41, value 102 in hour 42, value 105 in hour 43. Mean Blood Pressure has value 66 in hour 31, value 62 in hour 32, value 83 in hour 33, value 85 in hour 34, value 81 in hour 35, value 87 in hour 36, value 82 in hour 37, value 85 in hour 38, value 80 in hour 39, value 75 in hour 40, value 68 in hour 41, value 74 in hour 42, value 72 in hour 44. Minute Ventilation has value 4 in hour 31, value 3.7 in hour 32, value 3.84 in hour 33, value 4.05`

in hour 34, value 6.54 in hour 35, value 5.51 in hour 36, value 4.56 in hour 37, value 4.4 in hour 38. PC has value 0 in hour 31, value 0 in hour 32, value 0 in hour 33, value 0 in hour 34, value 0 in hour 35, value 0 in hour 36, value 0 in hour 37, value 0 in hour 38. PEEP has value 5.88 in hour 31, value 5.88 in hour 32, value 5.88 in hour 33, value 5.88 in hour 34, value 5.88 in hour 35, value 4.78 in hour 36, value 3.68 in hour 37, value 3.68 in hour 38. Respiratory Rate has value 12.5 in hour 31, value 7 in hour 32, value 7 in hour 33, value 6.5 in hour 34, value 8 in hour 35, value 8.38 in hour 36, value 7.5 in hour 37, value 4 in hour 38, value 20 in hour 39, value 21 in hour 40, value 16 in hour 41, value 15 in hour 42, value 22 in hour 43, value 26 in hour 44, value 20 in hour 47. SAS has value 5 in hour 32, value 5 in hour 33, value 5 in hour 34, value 5 in hour 35, value 5 in hour 36, value 4 in hour 37, value 4 in hour 39, value 4 in hour 40, value 4 in hour 44. Saturation has value 96 in hour 31, value 95 in hour 32, value 99 in hour 33, value 100 in hour 34, value 99 in hour 35, value 99.5 in hour 36, value 100 in hour 37, value 100 in hour 38, value 96 in hour 39, value 93 in hour 40, value 94 in hour 41, value 97 in hour 42, value 96 in hour 43, value 96 in hour 44, value 95 in hour 45, value 94 in hour 46, value 94 in hour 47. Systolic Blood Pressure has value 94 in hour 31, value 84 in hour 32, value 102 in hour 33, value 106 in hour 34, value 106 in hour 35, value 105 in hour 36, value 108 in hour 37, value 109 in hour 38, value 110 in hour 39, value 102 in hour 40, value 101 in hour 41, value 97 in hour 42, value 98 in hour 44. Temperature has value 36.4 in hour 33, value 35.9 in hour 37. Tidal Volume has value 230 in hour 31, value 215 in hour 32, value 235 in hour 33, value 267.5 in hour 34, value 115 in hour 35, value 300 in hour 36, value 255 in hour 37, value 480 in hour 38. Urine Output has value 350 in hour 31, value 100 in hour 32, value 35 in hour 33, value 60 in hour 34, value 55 in hour 35, value 60 in hour 36, value 60 in hour 37, value 300 in hour 38, value 100 in hour 39, value 100 in hour 40, value 100 in hour 44. Patient receives Analgesia at hour 32, 33, 34, 35, 36, 37, 38, 39. Patient receives Anticoagulants at hour 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47. Patient receives Neuroleptics at hour 38, 39, 40, 41, 42, 43, 44, 45, 46, 47. Patient receives PS at hour 31, 32, 33, 34, 35, 36, 37, 38. Patient receives Sedation at hour 32, 33, 34, 35, 36, 37, 38, 39. Patient receives TPN at hour 32.

The following is the detailed prompts and samples of summarization

ZERO-SHOT

> You are a clinical agent that analyze and then provide the most concise summarization on ICU time series data for forecasting.

'**Summary:**\n\nPatient exhibits fluctuating hemodynamics and respiratory status. Key observations:\n* **Hemodynamics:** Initial hypotension (MBP 66, SBP 94) improved, then showed variability (MBP 62-87, SBP 84-110). DBP increased significantly from hour 32-36 (56-82) then decreased. Heart rate is consistently elevated (97-109).\n* **Respiratory:** Airway pressure decreased (7.36 -> 5.08) then increased. PEEP decreased (5.88 -> 3.68). Respiratory rate varied significantly (7-26). FiO2 decreased from 30 to 21. Saturation generally stable (95-100), with a dip to 93 at hour 40.\n* **Renal:** Creatinine is high (103). Urine output is variable (35-350). BUN is slightly elevated (5.6).\n* **Neurologic:** GCS initially low (8), improved to 14, then decreased.\n* **Labs:** Sodium is high (150). Other labs (ALP, ALT, AST, Bilirubin, CK, Troponin, WBC) are relatively stable or within normal limits.\n* **Interventions:** Continuous anticoagulation, analgesia, sedation, and neuroleptics administered. TPN given at hour 32.\n\n**Overall:** Patient shows signs of initial instability with subsequent fluctuations in hemodynamics, respiratory parameters, and neurological status. High sodium and creatinine are notable.'

ICD

You are a clinical analysis agent. Summarize ICU time-series patient data for forecasting using this structure:

- Trend — overall direction of vitals, labs, therapies, and organ support.
- Seasonality — repeating cycles (e.g., circadian).
- Irregularities — acute deviations or events.

Map each diagnosis to its affected organ system (cardiac, respiratory, hepatic, renal, neurologic, etc.). For every system, assign a severity score from 1 (least affected) to 10 (most severe) based on data patterns and level of support required. Output only the summary in clear clinical prose, concluding with a semicolon-separated list of organ systems and scores (e.g., "Cardiovascular 7/10; Respiratory 8/10; Hepatic 3/10"). Do not explain your reasoning.

```
'Overall, the patient exhibits a complex trajectory with periods of relative stability
punctuated by acute deteriorations and improvements. Trends show initial hemodynamic
instability with low blood pressure, followed by a period of hypertension and tachycardia,
then a return towards lower pressures. Respiratory status is marked by a significant decrease
in airway pressure and PEEP, alongside variable respiratory rates and tidal volumes,
suggesting evolving ventilation needs. Neurologic status initially shows depressed GCS, which
improves significantly before a potential decline. Renal function shows initial oliguria
followed by periods of increased output, but creatinine remains elevated. Liver function tests
are relatively stable. There is no clear circadian pattern evident in the provided data.
Irregularities include a sharp drop in respiratory rate and tidal volume around hour 35, a
subsequent increase in tidal volume at hour 38, a significant improvement in GCS around hour
39, and fluctuations in fluid balance. The patient received multiple therapies including
analgesia, anticoagulants, neuroleptics, pressors (PS), and sedation, alongside
TPN.\n\nCardiovascular 7/10; Respiratory 8/10; Hepatic 3/10; Renal 6/10; Neurologic 6/10'
```

TREND

Examine the data closely and describe the trend changes step by step over time. For example: from [start] to [midpoint], what happened? Then from [midpoint] to [end], what happened? After describing each phase, conclude with an overall summary in natural language. Summarize as many feature as possible starting from the most significant ones in concise words. Only include your description and summarization.

```
"**Phase 1: Hour 31 to Hour 38**\n\nFrom hour 31 to hour 38, the patient showed signs of
initial instability followed by some improvement. Airway Pressure decreased steadily from 7.36
to 5.08, then slightly increased to 5.3. Diastolic Blood Pressure initially decreased from 59
to 56, then significantly increased to 80, fluctuating between 75 and 78 before rising again
to 82. FiO2 remained stable at 30 until hour 39. GCS remained low at 8, then dropped to 7
before improving to 14 by hour 39. Heart Rate increased from 97 to 105, then slightly
decreased to 102. Mean Blood Pressure decreased from 66 to 62, then increased significantly to
87, fluctuating between 82 and 85. Minute Ventilation increased from 4 to 6.54, then decreased
to 4.56 and 4.4. PEEP decreased from 5.88 to 3.68. Respiratory Rate decreased significantly
from 12.5 to 4, then increased sharply to 20. Saturation improved from 96 to 100, then
decreased to 96. Systolic Blood Pressure decreased from 94 to 84, then increased significantly
to 109. Temperature decreased from 36.4 to 35.9. Tidal Volume decreased from 230 to 115, then
increased dramatically to 480. Urine Output decreased significantly from 350 to 35, then
increased to 300. The patient received analgesia, anticoagulants, PS, and sedation during this
period.\n\n**Phase 2: Hour 39 to Hour 47**\n\nFrom hour 39 to hour 47, the patient's condition
showed further fluctuations and potential deterioration. Airway Pressure decreased from 5.3 to
6.29, then decreased to 5.08. Diastolic Blood Pressure decreased from 82 to 69, then 68, 59,
67, 65. FiO2 decreased from 21 to 21. GCS remained at 14. Heart Rate increased from 102."
```

COT

You are a healthcare agent that summarizes ICU patients' time series status for future time series forecasting. Analyze this step by step.

1. Analyze the time series data to identify key trends.
2. Based on the identified trends, determine potential clinical implications.
3. Summarize the findings and suggest possible interventions.

Summarize as many feature as possible starting from the most significant ones in concise words and only respond with your summarization.

Dataset	Embedding	Silhouette
HiRID	Raw time-series	0.372
HiRID	CoT embedding	0.452
HiRID	ICD embedding	0.408
HiRID	zero_shot embedding	0.638
HiRID	Trend embedding	0.540
MIMIC	Raw time-series	0.332
MIMIC	CoT embedding	0.618
MIMIC	ICD embedding	0.322
MIMIC	zero_shot embedding	0.472
MIMIC	Trend embedding	0.681
PPICU	Raw time-series	0.486
PPICU	CoT embedding	0.573
PPICU	ICD embedding	0.385
PPICU	zero_shot embedding	0.375
PPICU	Trend embedding	0.485

Table 3: Silhouette scores for different embeddings across datasets (higher is better).

'**Summary of ICU Patient Status (Hour 31-47):**\n\n* **Hemodynamics:** Initial hypotension (SBP 94, MBP 66) improved significantly (SBP 110, MBP 87) but shows recent instability (DBP fluctuations, MBP decrease in hour 40-41). Tachycardia (HR 97-109) persists. Airway pressure decreased significantly (7.36 to 5.08) then increased slightly, suggesting potential improvement in lung mechanics followed by possible deterioration or change in ventilation strategy.\n* **Respiratory Status:** Respiratory rate is variable (initially low 7-6.5, then increased to 20-26). FiO2 decreased from 30 to 21, indicating potential improvement in oxygenation. Saturation generally stable (96-100) but decreased recently (93-94). Tidal volume shows significant fluctuations (230-480).\n* **Renal Function:** Creatinine is elevated (103), BUN is borderline high (5.6). Urine output is highly variable (350-35, then 100-300).\n* **Metabolic/Electrolytes:** Hypernatremia (Na 150). Hypokalemia (K 3.4). Hypocalcemia (Ca 1.99). Hypophosphatemia (Phos 0.78). Lactate is borderline elevated (1.4).\n* **Neurological Status:** GCS initially low (8), improved to 14, then decreased again.\n* **Liver Function:** Liver enzymes (ALP, ALT, AST) are mildly elevated. Bilirubin is borderline high (4).\n* **Coagulation:** INR is normal (0.99).\n* **Other:** Platelets are normal (303). WBC is normal (8.4). Troponin is elevated (8).\n\n**Clinical Implications:**\n\n* Hemodynamic instability requires close monitoring and potential vasopressor adjustment.\n* Variable respiratory status suggests ongoing lung injury or changing ventilation needs.\n* Renal dysfunction (elevated creatinine, variable urine output) warrants monitoring and potential intervention.\n* Electrolyte imbalances (hypernatremia, hypokalemia, hypocalcemia, hypophosphatemia) need correction.\n*'

F Embedding Quality Assessment

We assess representation quality by applying the unsupervised dimensionality-reduction method UMAP to the raw time series and four *gemini-2.0-flash*–based summaries (*CoT*, *ICD*, *zero_shot*, *Trend*). Although UMAP is fit without labels, we color points by mortality status (0/1) and quantify cluster separation using the Silhouette score (higher is better), as a proxy for discriminability on the downstream *mortality classification* task. Silhouette scores are reported in Table 3. Results are averaged over target embedding dimensions of 128, 256, and 512.

G Privacy Result Table

Table referenced in RQ6 is shown in Table 4.

H Token Count Comparison (Qwen3-Embedding-8B)

We report tokenization statistics for two PPICU text corpora using the `Qwen3-Embedding-8B` tokenizer (special tokens excluded) across raw text and summarization text. Table 5 summarizes per-string token counts. Average tokens for raw text is 6,106.5 and average tokens (Summarization from ICD) is 234.0. Token reduction: 5,872.48 (2,509.3% vs. B).

Dataset →	**HiRID**		**MIMIC**		**PPICU**	
Tasks →	Age	Gender	Age	Gender	Age	Gender
Method ↓	MAE	AUROC	MAE	AUROC	MAE	AUROC
Mean	0.805017	0.754615	0.803168	0.678225	0.76247	0.707425
Right shift	0.803392	0.742055	0.798042	0.694878	0.767785	0.726552
Interpolation	0.800455	0.726648	0.799755	0.670795	0.76095	0.720293
TSDE	0.798725	0.662075	0.797682	0.733638	0.76362	0.764377
TimesFM	0.798495	0.64247	0.795795	0.566465	0.759935	0.63862
No-Summary	0.801658	0.653393	0.803828	0.640195	0.812675	0.666798
Llama 3.1	0.801658	0.64383	0.805292	0.623775	0.81619	0.643475
MedGemma	0.801207	0.645748	0.803142	0.61944	0.814385	0.652317
Gemini 2.0-Flash	0.801828	0.650965	0.805535	0.605605	0.813318	0.646813
Dataset →	**HiRID → PPICU**		**MIMIC → PPICU**		**PPICU → PPICU**	
Tasks →	Age	Gender	Age	Gender	Age	Gender
Method ↓	MAE	AUROC	MAE	AUROC	MAE	AUROC
Mean	0.805353	0.63862	0.78551	0.48239	0.00	0.000
Right shift	0.791562	0.63101	0.77174	0.63016	0.00	0.000
Interpolation	0.78547	0.6317	0.778125	0.63008	0.00	0.000
TSDE	0.7615	0.452603	0.755202	0.457027	0.00	0.000
TimesFM	0.760583	0.63862	0.75289	0.6113	0.00	0.000
No-Summary	0.811862	0.643015	0.80723	0.635738	0.00	0.000
Llama 3.1	0.81223	0.63101	0.80953	0.63016	0.00	0.000
MedGemma	0.812695	0.64306	0.805072	0.633265	0.00	0.000
Gemini 2.0-Flash	0.813743	0.642868	0.809338	0.63083	0.00	0.000

Table 4: RQ6 (Privacy): Age and gender prediction. Age is evaluated with MAE; Gender with AUROC.

Table 5: Per-string tokenization summary (Qwen3-Embedding-8B; no special tokens).

	count	total	mean	std	max
File A (`Raw Text`)	46,818	285,894,402	6,106.51	3,328.77	13,233.00
File B (`ICD Summarization`)	47,119	11,027,274	234.03	60.52	567.00

We add a more detailed report of latency, GPU time, token usage and approximate costs across all of methods in Record2Vec in Table 6. We report inference + embedding latency, GPU time per batch, token counts and costs per patient. Specifically, the language models are served on 4 NVIDIA L40S GPUs using vLLM inference; the embedding model is served on 1 L40S for the summarization-based methods, while the no-summarization baseline uses 2 L40S. Additionally, we compare the inference time using a deterministic template (no summarization). Balancing the results and costs, we find Gemini-2.0-flash results the least latency/costs while maintaining high performance, with a cost of 0.7 dollar per 1000 samples and 0.26s latency per patient.

I 16-SAMPLE FEW-SHOT FINETUNE RESULTS

In this section, we report the detailed mean±std results under our 16-sample few-shot finetuning scenario as Table 7. This is complementary to Figure 5.

J 1000 SAMPLES TRAINING RESULTS

In the section, we record the comparison between results of pre-training using 1000 in-distribution samples and 16-sample few-shot finetuning result of Record2Vec. This is a simulation in real world settings for smaller local hospitals, where they have less sample to train a robust model. The 1000 samples are chosen at random and shared across methods to ensure fair comparison.

Model	Latency (per sample)	GPU time (per batch)	Token counts	Cost (per 1k samples)
MedGemma + Qwen	$0.784 + 0.026\,\mathrm{s}$	$3959.41 + 119.61\,\mathrm{ms}$	234.03 tks/sample	–
Llama + Qwen	$0.25406 + 0.023\,\mathrm{s}$	$1299.4 + 117.24\,\mathrm{ms}$	254.17 tks/sample	–
Gemini + Qwen	$0.28 + 0.036\,\mathrm{s}$	–	293.44 tks/sample	0.7 $/1k samples
Qwen (no summarization)	$2.9675\,\mathrm{s}$	$23608.14\,\mathrm{ms}$	6106.56 tks/sample	–

Table 6: Latency, GPU time, token usage, and approximate cost of different summarization/inference pipelines.

Dataset →	**HiRID → PPICU**			**MIMIC → PPICU**			**Wins**
Task → **Method** ↓	Forecast MSE↓	LoS MAE↓	Mortality AUROC↑	Forecast MSE↓	LoS MAE↓	Mortality AUROC↑	
Mean	$\underline{0.133 \pm 0.001}$	2.678 ± 0.229	0.600 ± 0.007	0.183 ± 0.003	1.171 ± 0.284	0.565 ± 0.050	0
Right shift	0.136 ± 0.002	1.803 ± 0.041	0.621 ± 0.036	0.181 ± 0.003	0.662 ± 0.056	0.606 ± 0.036	0
Interpolation	0.139 ± 0.001	1.999 ± 0.123	0.605 ± 0.033	0.183 ± 0.003	0.972 ± 0.090	0.563 ± 0.028	0
TFM	0.144 ± 0.005	0.896 ± 0.042	$\underline{0.686 \pm 0.003}$	$\underline{0.163 \pm 0.001}$	1.067 ± 0.028	$\underline{0.683 \pm 0.002}$	0
TSDE	0.302 ± 0.002	$\underline{0.795 \pm 0.036}$	0.507 ± 0.017	0.392 ± 0.016	$\underline{0.643 \pm 0.170}$	0.586 ± 0.025	0
Record2Vec	$\mathbf{0.022 \pm 0.000}$	$\mathbf{0.500 \pm 0.005}$	$\mathbf{0.792 \pm 0.002}$	$\mathbf{0.0215 \pm 0.000}$	$\mathbf{0.448 \pm 0.003}$	$\mathbf{0.750 \pm 0.008}$	6

Table 7: Transfer performance from HiRID/MIMIC to PPICU. Values are mean±std. Best per column in **bold**; second-best underlined. Wins count the number of best results per method.

K Strong mortality transfer results

In this section, we report transfer learning results for strong mortality baselines: LSTM, GRU, TCN, Transformer as mentioned in Yèche et al. (2022). We observe that none surpass our simple downstream classifier, PatchTSMixer, which achieves an AUROC of 0.72. (Table 9). This suggests that there is still a need to develop models better suited to transfer-learning settings.

L Embedder Ablations

We performed additional ablations on the choice of embedding model and pooling / normalization strategy. Overall, we observe three main trends: (i) stronger embedding models generally yield better performance, although the gains are modest across current SOTA models; (ii) a weaker, non-SOTA embedder degrades performance but still remains above our non-summarization baseline; and (iii) our method is largely robust to changes in pooling and normalization, with the recommended configuration from the model documentation giving the most consistent performance across benchmarks.

On our base embedder Qwen3-Embedding-8B, we tested the following pooling + normalization pairs: *mean + L2* (our default), *mean + none*, *CLS + L2*, *last + L2*, and *max + L2*. In addition, we replaced Qwen3 with the current MTEB leader `nvidia/llama-embed-nemotron-8b` and the prior SOTA `gte-Qwen2-7B-instruct`. The in-distribution ablation results on PPICU are reported in Table 10, and the HiRID→PPICU transfer ablations are shown in Table 11.

M Detailed results for comparing prompting strategies and summarization methods

In this section, we provide the detailed performance values used to generate the rank distribution plots in Figure 3 and Figure 4. These figures provide an intuitive visualization of how different summarization and prompting strategies affect relative performance across our benchmark.

The rankings are determined by comparing the performance of the competing methods on five specific downstream tasks, using the same metrics reported in the main results (Tables 1 and 2). We utilize Mean Squared Error (MSE) for forecasting, Mean Absolute Error (MAE) for length-of-stay prediction, AUROC for mortality prediction, and Recall for both treatment planning (Drug) and measurement ordering (Lab). For the summarization analysis in Figure 3, ranks are computed across four methods: No-summary, Llama 3.1, MedGemma, and Gemini 2.0 Flash. For the prompting analysis in Figure 4, we compare four variants: Zero-shot, In-Context Learning (ICD), Chain-of-Thought (CoT), and Trend.

Task	Small PPICU	Few-shot Finetune (Record2Vec)
Mortality	0.6843	0.72
LoS	1.03	0.49
Forecast	0.62	0.183
Feature	0.901	0.97
Lab	0.94	0.97

Table 8: Performance comparison between models trained on a small PPICU subset vs. few-shot finetuned models.

	LSTM	GRU	Transformer	TCN
H→P Baseline	0.640	0.687	0.579	0.719
H→P Record2Vec	0.69788	0.6838	0.679	0.684
M→P Baseline	0.5235	0.5476	0.564	0.719
M→P Record2Vec	0.722	0.685	0.682	0.695

Table 9: Model performance for HiRID→PPICU (H→P) and MIMIC→PPICU (M→P), comparing baseline vs. Record2Vec representations.

The rank counts are aggregated across two distinct experimental settings. The "In-Distribution" results (shown in the left subfigures) aggregate rankings over 15 total tasks, corresponding to five tasks evaluated across three source datasets. The "Transfer Learning" results (shown in the right subfigures) aggregate rankings over 30 total tasks, covering five tasks across six distinct transfer directions. For the Figure 4, we only considered three sets of transfers (total 15 numbers). Table 12, Table 13 & 14, Table 15, and Table 16 present the exact performance numbers used to construct these rank distributions, allowing for direct inspection of the underlying values.

N GenHPF Modification and Replication Details

ICU data contain a far greater number of features observed over a much shorter time horizon, with multiple events often occurring simultaneously. When converting ICU data into a textual representation following GenHPF (Hur et al., 2024), we adopt a simple hierarchy: features are first divided into laboratory, vital, and binary (therapy / intervention) groups, and each feature is concatenated with its group prefix. For example, the feature `Hemoglobin` is encoded as `lab_Hemoglobin`. The feature groupings used for the three datasets are summarized in Table 17.

O Detailed Results

We provide detailed results and tables in this section for a comprehensive comparison between LLM-based summarize-then-embed pipeline with the three baseline methods.

Ablations	**Forecast**	**LoS**	**Mort**	**Drug**	**Lab**
Qwen3-mean-l2	0.021	0.347	0.90	0.911	0.931
Qwen3-mean-none	0.027	0.371	0.88	0.871	0.919
Qwen3-cls-l2	0.024	0.530	0.89	0.900	0.940
Qwen3-last-l2	0.022	0.423	0.89	0.920	0.923
gte-Qwen2-instruct	0.028	0.480	0.88	0.886	0.918
llama-embed-nemotron	0.021	0.379	0.89	0.9097	0.929
Baseline	0.040	0.378	0.52	0.878	0.857

Table 10: Ablations on PPICU (in-distribution).

Ablations	**Forecast**	**LoS**	**Mort**	**Drug**	**Lab**
Qwen3-mean-l2	0.183	0.69	0.72	0.97	0.97
Qwen3-mean-none	0.209	0.66	0.71	0.94	0.96
Qwen3-cls-l2	0.230	0.78	0.50	0.995	0.98
Qwen3-last-l2	0.190	0.78	0.73	0.94	0.96
gte-Qwen2-instruct	0.250	0.82	0.50	0.93	0.96
llama-embed-nemotron	0.190	0.72	0.713	0.96	0.998
Baseline	0.306	1.09	0.50	0.42	0.77

Table 11: Ablations for HiRID→PPICU transfer.

Table 12: Summarization methods performance comparison across 15 in-distribution tasks (Figure 3 left subfigure)

	HIRID					MIMIC					PPICU				
Model	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab
no-summary	0.021	0.3538	0.9	0.899	0.931	0.027	0.328	0.82	0.886	0.947	0.0217	0.376	0.64	0.9308	0.923
llama 3.1	0.0214	0.381	0.8797	0.899	0.925	0.0288	0.406	0.8131	0.886	0.938	0.029	0.3752	0.63	0.937	0.9271
medgemma	0.0237	0.3682	0.83	0.911	0.925	0.0299	0.4005	0.7788	0.8946	0.9432	0.029	0.3626	0.6346	0.925	0.936
gemini	0.028	0.347	0.8302	0.9064	0.9307	0.03	0.3328	0.77	0.903	0.942	0.017	0.358	0.6377	0.9308	0.9328

Table 13: Summarization methods performance comparison across 30 transfer learning tasks (Part 1: Transfer Pairs 1-3)

	H → P					M → P					H → M				
Model	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab
no-summary	0.21	0.98	0.66	0.92	0.96	0.263	0.77	0.7	0.92	0.92	0.1284	0.693	0.73	0.881	0.832
llama 3.1	0.1908	0.71	0.735	0.9243	0.9718	0.249	0.5522	0.7097	0.9415	0.9285	0.134	0.6464	0.82	0.878	0.851
medgemma	0.21	0.6981	0.74	0.97	0.98	0.239	0.5627	0.7145	0.9428	0.95	0.0922	0.537	0.81	0.891	0.83
gemini	0.183	0.7017	0.7001	0.92	0.9737	0.258	0.49	0.73	0.95	0.9372	0.089	0.6661	0.8044	0.852	0.842

Table 14: Summarization methods performance comparison across 30 transfer learning tasks (Part 2: Transfer Pairs 4-6)

	P → M					M → H					P → H				
Model	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab
no-summary	0.15971	0.518	0.75	0.913	0.8669	0.135	0.5805	0.8277	0.816	0.8857	0.0954	0.494	0.76	0.892	0.7895
llama 3.1	0.141	0.443	0.752	0.877	0.8669	0.1741	0.572	0.83	0.903	0.9085	0.096	0.4393	0.7657	0.835	0.734
medgemma	0.1281	0.4493	0.81	0.88	0.9054	0.135	0.5805	0.8277	0.8653	0.877	0.089	0.4682	0.7725	0.8637	0.8042
gemini	0.127	0.4886	0.7942	0.901	0.919	0.167	0.5864	0.8037	0.8239	0.916	0.083	0.429	0.81	0.8627	0.812

Table 15: Prompting methods performance comparison across 15 in-distribution tasks (Figure 4 left subfigure)

	HIRID					MIMIC					PPICU				
Model	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab
zero-shot	0.0244	0.354	0.8501	0.911	0.931	0.0282	0.3951	0.77	0.8894	0.9431	0.017	0.376	0.6367	0.9346	0.923
CoT	0.0268	0.381	0.8311	0.9078	0.925	0.0294	0.328	0.8101	0.886	0.938	0.029	0.358	0.6387	0.937	0.936
ICD	0.021	0.3517	0.83	0.899	0.9305	0.027	0.3844	0.8018	0.903	0.947	0.0284	0.3645	0.64	0.925	0.9234
Trend	0.028	0.347	0.9	0.9042	0.9263	0.03	0.406	0.82	0.8881	0.9469	0.0249	0.3713	0.63	0.9324	0.9308

Table 16: Prompting methods performance comparison across 15 transfer learning tasks (Figure 4 right subfigure)

	H → P					M → P					H → M				
Model	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab	Forecast	LOS	Mort	Drug	Lab
zero-shot	0.183	0.6994	0.6888	0.9359	0.9758	0.2536	0.57	0.7045	0.949	0.95	0.1082	0.6467	0.73	0.8878	0.83
CoT	0.1996	0.69	0.66	0.92	0.9667	0.2461	0.49	0.7106	0.92	0.9375	0.089	0.693	0.7328	0.868	0.832
ICD	0.21	0.6989	0.6916	0.933	0.98	0.239	0.5199	0.73	0.926	0.9498	0.134	0.6564	0.82	0.891	0.851
Trend	0.1927	0.71	0.74	0.97	0.96	0.258	0.5595	0.7	0.95	0.92 — 0.1248	0.537	0.7629	0.852	0.847	

Dataset	Lab features	Vital features	Binary features
HiRID	ALP, ALT, AST, Bicarbonate, Bilirubin, BloodUreaNitrogen, Calcium, Chloride, CreatineKinase, Creatinine, Glucose, Hemoglobin, INR, Lactate, Magnesium, PaCO2, PaO2, Phosphate, Platelets, Potassium, Sodium, Troponin, WBC, ph	AirwayPressure, AirwayPressurePeak, DiastolicBloodPressure, FiO2, FluidBalance, GCS, HeartRate, MeanBloodPressure, PEEP, RespiratoryRate, SAS, Saturation, SystolicBloodPressure, Temperature, TidalVolume, UrineOutput, ICPMonitor, Ventilation	Analgesia, Antiarrhythmics, Antibiotics, Anticoagulants, Antiepileptics, Antihypertensives, Aspirin, CaReplacement, Dialysis, Diuretics, ICPMonitor, Insulin, KReplacement, LiverToxicDrug, MgReplacement, Neuroleptics, Paralysis, Saline, Sedation, Steroids, TPN, Transfusions, Vasopressors, Ventilation
MIMIC	ALP, ALT, Bicarbonate, Bilirubin, BloodUreaNitrogen, Calcium, Creatinine, CreatinineKinase, Hemoglobin, INR, Lactate, Magnesium, PaCO2, PaO2, Phosphate, Platelets, Potassium, Sodium, Troponin, WBC, ph	AirwayPressure, DiastolicBloodPressure, FiO2, GCS, HeartRate, ICDSC, MeanBloodPressure, MinuteVentilation, PEEP, RespiratoryRate, SAS, SystolicBloodPressure, Temperature, TidalVolume, UrineOutput	Analgesia, Antiarrhythmics, Antibiotics, Anticoagulants, Antiepileptics, Antihypertensives, Antipsychotics, CTScan, CaReplacement, Dialysis, Diuretics, EnteralNutrition, ICPMonitor, KReplacement, MRI, MgReplacement, PPI, Paralysis, TPN, Transfusions, UltraSound, Vasopressors, Ventilation, Xray
PPICU	ALP, ALT, AST, Bicarbonate, Bilirubin, BloodUreaNitrogen, Calcium, Chloride, Creatinine, CreatinineKinase, GGT, Glucose, Hemoglobin, INR, Lactate, Magnesium, PaCO2, PaO2, Phosphate, Platelets, Potassium, Sodium, Troponin, WBC, ph	AirwayPressure, AirwayPressureIP, AirwayPressurePeak, DiastolicBloodPressure, FiO2, FluidInput, FluidOutput, GCS, HeartRate, ICDSC, MeanBloodPressure, MinuteVentilation, PAVSupport, PC, PEEP, PlateauPressure, RespiratoryRate, SAS, Saturation, SystolicBloodPressure, Temperature, TidalVolume, UrineOutput	Analgesia, Antiarrhythmic, Antiarrhythmics, Antibiotics, Anticoagulants, Antiepileptics, Antihypertensives, Aspirin, Dialysis, Diuretics, EKG, EVD, ICPMonitor, InhaledVasodilator, Insulin, Isoproterenol, MgReplacement, Neuroleptics, OsmoticTherapy, PPI, PS, Paralysis, Sedation, Steroids, TPN, Transfusions, Vasopressors

Table 17: Feature groupings for converting ICU data into hierarchical textual representations. Each feature is prefixed with its group label (e.g., `lab_Hemoglobin`, `vital_HeartRate`, `bin_Vasopressors`) before tokenization.

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	mae	mse	mae	mse	mae	mse
Llama-3.1 CoT +TFM lstm	0.799 ± 4.12e − 4	1 ± 1.893e − 4	0.802 ± 2.517e − 5	1 ± 1.732e − 5	0.808 ± 1.323e − 3	1 ± 3.87e − 4
Llama-3.1 CoT +TFM mlp	0.799 ± 5.4e − 4	1 ± 1.888e − 4	0.802 ± 6.557e − 5	1 ± 7.572e − 5	0.811 ± 1.514e − 3	1.001 ± 1.177e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.808 ± 1.848e − 3	1.018 ± 3.714e − 3	0.806 ± 1.544e − 3	1.01 ± 4.083e − 3	0.818 ± 5.124e − 3	1.015 ± 7.344e − 3
Llama-3.1 CoT +TFM timemixer	0.8 ± 1.198e − 3	1 ± 2.955e − 4	0.802 ± 1.229e − 4	1 ± 1.528e − 5	0.809 ± 2.309e − 5	1 ± 1e − 5
Llama-3.1 CoT lstm	0.799 ± 4.594e − 4	1 ± 1.09e − 4	0.802 ± 8.139e − 5	1 ± 3.948e − 5	0.808 ± 7.551e − 4	1 ± 1.603e − 4
Llama-3.1 CoT mlp	0.801 ± 1.066e − 3	1.001 ± 4.061e − 4	0.802 ± 8.124e − 5	1 ± 3.742e − 5	0.811 ± 8.279e − 4	1.001 ± 3.531e − 4
Llama-3.1 CoT patchtsmixer	0.802 ± 4.342e − 3	1.009 ± 5.105e − 3	0.805 ± 1.877e − 3	1.004 ± 2.605e − 3	0.814 ± 3.97e − 3	1.007 ± 2.545e − 3
Llama-3.1 CoT timemixer	0.8 ± 1.173e − 4	1 ± 2.363e − 5	0.802 ± 6.351e − 5	1 ± 5.774e − 6	0.81 ± 1.245e − 3	1.001 ± 4.287e − 4
Llama-3.1 ICD +TFM lstm	0.799 ± 2.33e − 3	1 ± 7.342e − 4	0.802 ± 1.25e − 4	1 ± 5.292e − 5	0.808 ± 4.464e − 4	1 ± 2.193e − 4
Llama-3.1 ICD +TFM mlp	0.8 ± 8.113e − 4	1 ± 6.769e − 4	0.802 ± 1.704e − 4	1 ± 2.887e − 5	0.811 ± 6.393e − 4	1.001 ± 4.518e − 4
Llama-3.1 ICD +TFM patchtsmixer	0.803 ± 2.735e − 3	1.015 ± 6.16e − 3	0.807 ± 2.349e − 3	1.01 ± 3.026e − 3	0.818 ± 1.976e − 3	1.013 ± 1.36e − 3
Llama-3.1 ICD +TFM timemixer	0.799 ± 5.401e − 4	1 ± 9.165e − 5	0.802 ± 1.528e − 5	1 ± 0e0	0.809 ± 1.202e − 3	1 ± 2.977e − 4
Llama-3.1 ICD lstm	0.8 ± 1.023e − 3	1 ± 2.974e − 4	0.802 ± 1.299e − 4	1 ± 1.291e − 5	0.808 ± 3.73e − 4	1 ± 2.109e − 4
Llama-3.1 ICD mlp	0.801 ± 6.392e − 4	1.001 ± 7.034e − 4	0.802 ± 1.688e − 4	1 ± 6.85e − 5	0.811 ± 1.794e − 3	1.001 ± 7.748e − 4
Llama-3.1 ICD patchtsmixer	0.8 ± 4.883e − 3	1.007 ± 2.973e − 3	0.803 ± 9.965e − 4	1.003 ± 1.069e − 3	0.811 ± 3.361e − 3	1.005 ± 1.447e − 3
Llama-3.1 ICD timemixer	0.799 ± 1.554e − 3	1 ± 1.954e − 4	0.802 ± 1.087e − 4	1 ± 1.155e − 5	0.809 ± 2.757e − 4	1 ± 7.632e − 5
Llama-3.1 Trend +TFM lstm	0.799 ± 1.36e − 3	1 ± 1.986e − 4	0.802 ± 1.007e − 4	1 ± 9.815e − 5	0.808 ± 1.002e − 3	1 ± 1.05e − 4
Llama-3.1 Trend +TFM mlp	0.8 ± 7.948e − 4	1.001 ± 5.086e − 4	0.802 ± 2.108e − 4	1 ± 6.506e − 5	0.81 ± 9.158e − 4	1.001 ± 7e − 5
Llama-3.1 Trend +TFM patchtsmixer	0.805 ± 1.46e − 3	1.013 ± 1.356e − 3	0.806 ± 3.252e − 4	1.008 ± 2.559e − 3	0.816 ± 2.077e − 3	1.013 ± 1.48e − 3
Llama-3.1 Trend +TFM timemixer	0.799 ± 1.756e − 3	1 ± 3.102e − 4	0.802 ± 9.165e − 5	1 ± 5.774e − 6	0.809 ± 3.219e − 4	1 ± 1.026e − 4
Llama-3.1 Trend lstm	0.799 ± 7.871e − 4	1 ± 1.794e − 4	0.802 ± 2.5e − 5	1 ± 4.573e − 5	0.809 ± 1.105e − 3	1 ± 3.561e − 4
Llama-3.1 Trend mlp	0.801 ± 1.622e − 3	1.001 ± 8.542e − 4	0.802 ± 1.209e − 4	1 ± 1.414e − 5	0.81 ± 1.283e − 3	1.001 ± 2.9e − 4
Llama-3.1 Trend patchtsmixer	0.804 ± 4.987e − 3	1.007 ± 2.488e − 3	0.804 ± 1.247e − 3	1.004 ± 1.947e − 3	0.815 ± 2.49e − 3	1.007 ± 1.752e − 3
Llama-3.1 Trend timemixer	0.8 ± 1.757e − 3	1.001 ± 5.354e − 4	0.802 ± 5.058e − 5	1 ± 5e − 6	0.809 ± 1.221e − 3	1.001 ± 3.344e − 4
Llama-3.1 zero_shot +TFM lstm	0.799 ± 7.994e − 4	1 ± 3.576e − 4	0.802 ± 1.904e − 4	1 ± 6.11e − 5	0.808 ± 7.767e − 4	1 ± 4.842e − 4
Llama-3.1 zero_shot +TFM mlp	0.799 ± 1.048e − 3	1 ± 8.607e − 4	0.802 ± 7.767e − 5	1 ± 1.914e − 4	0.809 ± 6.928e − 5	1.001 ± 8.336e − 4
Llama-3.1 zero_shot +TFM patchtsmixer	0.806 ± 4.216e − 3	1.018 ± 6.421e − 3	0.807 ± 2.386e − 3	1.011 ± 3.676e − 3	0.82 ± 5.632e − 3	1.02 ± 1.084e − 2
Llama-3.1 zero_shot +TFM timemixer	0.8 ± 2.454e − 4	1 ± 5.568e − 5	0.802 ± 3.464e − 5	1 ± 5.774e − 6	0.809 ± 9.822e − 4	1 ± 2.2e − 4
Llama-3.1 zero_shot lstm	0.799 ± 6.685e − 4	1 ± 2.419e − 4	0.802 ± 1.103e − 4	1 ± 3.594e − 5	0.809 ± 9.524e − 4	1 ± 2.787e − 4
Llama-3.1 zero_shot mlp	0.801 ± 5.996e − 4	1.001 ± 3.377e − 4	0.802 ± 1.721e − 4	1 ± 6.272e − 5	0.812 ± 2.202e − 4	1.002 ± 3.597e − 4
Llama-3.1 zero_shot patchtsmixer	0.802 ± 3.087e − 3	1.006 ± 1.034e − 3	0.805 ± 2.653e − 3	1.005 ± 3.432e − 3	0.816 ± 4.356e − 3	1.009 ± 6.16e − 3
Llama-3.1 zero_shot timemixer	0.8 ± 1.042e − 3	1 ± 2.07e − 4	0.802 ± 2.217e − 5	1 ± 5e − 6	0.809 ± 8.542e − 4	1 ± 2.232e − 4
TFM lstm	0.794 ± 1.222e − 3	0.995 ± 2.307e − 4	0.793 ± 1.426e − 4	0.978 ± 1.36e − 4	0.752 ± 2.269e − 4	0.934 ± 7.789e − 5
TFM mlp	0.795 ± 1.084e − 3	0.995 ± 2.241e − 4	0.794 ± 1.422e − 4	0.978 ± 2.623e − 4	0.753 ± 8.388e − 4	0.935 ± 1.005e − 3
TFM patchtsmixer	0.798 ± 2.747e − 3	1.007 ± 1.228e − 3	0.796 ± 9.829e − 4	0.985 ± 1.854e − 3	0.76 ± 2.015e − 3	0.947 ± 2.213e − 3
TFM timemixer	0.795 ± 1.268e − 3	0.994 ± 1.484e − 4	0.794 ± 1.97e − 4	0.978 ± 7.767e − 5	0.753 ± 8.918e − 4	0.934 ± 1.767e − 4
TSDE lstm	0.794 ± 1.283e − 3	0.994 ± 2.651e − 4	0.793 ± 2.316e − 4	0.977 ± 3.348e − 4	0.752 ± 1.179e − 4	0.934 ± 3.616e − 4
TSDE mlp	0.795 ± 1.53e − 3	0.995 ± 8.894e − 4	0.793 ± 1.951e − 4	0.977 ± 2.168e − 4	0.753 ± 5.773e − 4	0.934 ± 1.016e − 3
TSDE patchtsmixer	0.799 ± 4.027e − 3	1.012 ± 3.068e − 3	0.798 ± 1.979e − 3	0.987 ± 3.758e − 3	0.764 ± 5.447e − 3	0.951 ± 5.592e − 3
TSDE timemixer	0.795 ± 1.15e − 3	0.994 ± 1.347e − 4	0.793 ± 1.949e − 4	0.978 ± 6.702e − 5	0.753 ± 9.068e − 4	0.934 ± 1.663e − 4
gemini-2.0-flash CoT +TFM lstm	0.798 ± 7.966e − 4	1 ± 1.955e − 4	0.802 ± 1.419e − 4	1 ± 1.25e − 4	0.807 ± 2.479e − 4	0.999 ± 3.291e − 4
gemini-2.0-flash CoT +TFM mlp	0.8 ± 8.967e − 4	1 ± 2.458e − 4	0.802 ± 8.021e − 5	1 ± 1.656e − 4	0.81 ± 1.556e − 3	1.001 ± 1.334e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.802 ± 3.295e − 3	1.015 ± 5.567e − 3	0.806 ± 2.782e − 3	1.011 ± 4.901e − 3	0.818 ± 8.316e − 3	1.017 ± 1.244e − 2
gemini-2.0-flash CoT +TFM timemixer	0.8 ± 1.567e − 3	1 ± 3.384e − 4	0.802 ± 1.253e − 4	1 ± 1.155e − 5	0.809 ± 9.109e − 4	1 ± 2.371e − 4
gemini-2.0-flash CoT lstm	0.8 ± 7.464e − 4	1 ± 2.128e − 4	0.802 ± 1.795e − 4	1 ± 7.719e − 5	0.809 ± 1.255e − 3	1 ± 2.233e − 4
gemini-2.0-flash CoT mlp	0.801 ± 8.308e − 4	1.001 ± 4.742e − 4	0.802 ± 6.928e − 5	1 ± 9.309e − 5	0.812 ± 1.04e − 3	1.001 ± 4.701e − 4
gemini-2.0-flash CoT patchtsmixer	0.803 ± 5.182e − 3	1.007 ± 3.614e − 3	0.804 ± 1.278e − 3	1.003 ± 8.689e − 4	0.814 ± 5.678e − 3	1.006 ± 3.776e − 3
gemini-2.0-flash CoT timemixer	0.799 ± 4.21e − 4	1 ± 6.683e − 5	0.802 ± 1.17e − 4	1 ± 9.574e − 6	0.809 ± 1.109e − 4	1 ± 2.646e − 5
gemini-2.0-flash ICD +TFM lstm	0.798 ± 6.035e − 4	1 ± 3.291e − 4	0.802 ± 1.286e − 4	0.999 ± 1.311e − 4	0.808 ± 7.289e − 4	1 ± 3.786e − 5
gemini-2.0-flash ICD +TFM mlp	0.801 ± 1.713e − 3	1.001 ± 9.697e − 4	0.802 ± 1.443e − 4	1 ± 1.752e − 4	0.81 ± 9.042e − 4	1 ± 4.102e − 4
gemini-2.0-flash ICD +TFM patchtsmixer	0.806 ± 6.972e − 3	1.019 ± 9.015e − 3	0.805 ± 1.407e − 3	1.009 ± 3.682e − 3	0.819 ± 6.351e − 3	1.016 ± 5.938e − 3
gemini-2.0-flash ICD +TFM timemixer	0.799 ± 5.112e − 4	1 ± 9.713e − 5	0.802 ± 7.234e − 5	1 ± 1.155e − 5	0.81 ± 2.857e − 4	1.001 ± 8.718e − 5
gemini-2.0-flash ICD lstm	0.799 ± 1.504e − 3	1 ± 3.638e − 4	0.802 ± 5.909e − 5	1 ± 2.38e − 5	0.809 ± 1.431e − 3	1 ± 3.902e − 4
gemini-2.0-flash ICD mlp	0.802 ± 7.823e − 4	1.001 ± 3.651e − 4	0.802 ± 1.741e − 4	1 ± 5.377e − 5	0.811 ± 9.275e − 4	1.001 ± 4.109e − 4
gemini-2.0-flash ICD patchtsmixer	0.8 ± 2.129e − 3	1.005 ± 1.943e − 3	0.805 ± 2.47e − 3	1.005 ± 7.609e − 4	0.815 ± 4.102e − 3	1.007 ± 1.912e − 3
gemini-2.0-flash ICD timemixer	0.8 ± 1.003e − 3	1 ± 2.123e − 4	0.802 ± 2.708e − 5	1 ± 5e − 6	0.809 ± 6.5e − 4	1 ± 1.513e − 4
gemini-2.0-flash Trend +TFM lstm	0.798 ± 6.421e − 4	1 ± 9.019e − 5	0.802 ± 1.212e − 4	1 ± 6.083e − 5	0.807 ± 4.572e − 4	1 ± 1.253e − 4
gemini-2.0-flash Trend +TFM mlp	0.8 ± 1.213e − 3	1 ± 9.142e − 4	0.802 ± 5.292e − 5	1 ± 1.389e − 4	0.81 ± 9.521e − 4	1 ± 1.044e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.809 ± 2.167e − 3	1.021 ± 4.501e − 3	0.806 ± 1.91e − 3	1.011 ± 5.336e − 3	0.816 ± 2.824e − 3	1.017 ± 4.701e − 3
gemini-2.0-flash Trend +TFM timemixer	0.801 ± 7.143e − 4	1.001 ± 2.248e − 4	0.802 ± 1.039e − 4	1 ± 5.774e − 6	0.809 ± 4.07e − 4	1 ± 8.888e − 5
gemini-2.0-flash Trend lstm	0.8 ± 2.157e − 3	1.001 ± 5.701e − 4	0.802 ± 7.089e − 5	1 ± 5.477e − 5	0.809 ± 1.029e − 3	1 ± 3.321e − 4
gemini-2.0-flash Trend mlp	0.8 ± 7.542e − 4	1.001 ± 5.427e − 4	0.802 ± 1.263e − 4	1 ± 3.317e − 5	0.811 ± 9.28e − 4	1.001 ± 4.237e − 4
gemini-2.0-flash Trend patchtsmixer	0.804 ± 5.877e − 3	1.009 ± 8.576e − 3	0.804 ± 2.883e − 3	1.006 ± 2.217e − 3	0.815 ± 3.544e − 3	1.007 ± 2.735e − 3
gemini-2.0-flash Trend timemixer	0.8 ± 1.186e − 3	1 ± 2.575e − 4	0.802 ± 9.274e − 5	1 ± 5e − 6	0.81 ± 8.4e − 4	1.001 ± 2.402e − 4
gemini-2.0-flash zero_shot +TFM lstm	0.798 ± 4.85e − 4	1 ± 3.037e − 4	0.802 ± 6.245e − 5	1 ± 5.508e − 5	0.808 ± 6.14e − 4	0.999 ± 2.887e − 5
gemini-2.0-flash zero_shot +TFM mlp	0.8 ± 1.932e − 3	1.001 ± 1.177e − 3	0.803 ± 1.767e − 4	1 ± 1.801e − 4	0.81 ± 1.093e − 3	1.001 ± 3.439e − 4
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.803 ± 2.508e − 3	1.014 ± 4.794e − 3	0.806 ± 3.949e − 3	1.009 ± 5.323e − 3	0.817 ± 2.7e − 3	1.013 ± 1.66e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.8 ± 1.125e − 3	1 ± 2.359e − 4	0.802 ± 3.205e − 4	1 ± 5.774e − 6	0.81 ± 8.261e − 4	1.001 ± 2.601e − 4
gemini-2.0-flash zero_shot lstm	0.799 ± 1.286e − 3	1 ± 2.53e − 4	0.802 ± 8.583e − 5	1 ± 3.594e − 5	0.809 ± 8.283e − 4	1 ± 1.135e − 4
gemini-2.0-flash zero_shot mlp	0.801 ± 7.43e − 4	1.001 ± 3.448e − 4	0.802 ± 5.944e − 5	1 ± 5.354e − 5	0.811 ± 1.49e − 3	1.001 ± 8.31e − 4
gemini-2.0-flash zero_shot patchtsmixer	0.802 ± 2.607e − 3	1.006 ± 3.524e − 3	0.806 ± 4.285e − 3	1.005 ± 3.005e − 3	0.813 ± 1.74e − 3	1.006 ± 1.151e − 3
gemini-2.0-flash zero_shot timemixer	0.8 ± 7.074e − 4	1 ± 1.916e − 4	0.802 ± 1.711e − 4	1 ± 8.165e − 6	0.809 ± 1.455e − 3	1 ± 3.885e − 4
interp lstm	0.795 ± 7.933e − 4	0.994 ± 4.948e − 4	0.793 ± 2.49e − 4	0.977 ± 2.41e − 4	0.754 ± 9.792e − 4	0.934 ± 4.055e − 4
interp mlp	0.796 ± 7.269e − 4	0.996 ± 6.276e − 4	0.793 ± 3.807e − 4	0.978 ± 7.938e − 4	0.754 ± 9.573e − 4	0.935 ± 3.97e − 4
interp patchtsmixer	0.8 ± 2.155e − 3	1.013 ± 3.985e − 3	0.8 ± 3.223e − 3	0.989 ± 3.73e − 3	0.761 ± 7.619e − 3	0.947 ± 8.651e − 3
interp timemixer	0.795 ± 1.088e − 3	0.995 ± 1.42e − 4	0.793 ± 8.104e − 5	0.978 ± 3.109e − 5	0.753 ± 8.277e − 4	0.935 ± 1.857e − 4
mean lstm	0.795 ± 6.998e − 4	0.994 ± 7.05e − 4	0.793 ± 2.029e − 4	0.977 ± 4.399e − 4	0.754 ± 7.075e − 4	0.935 ± 2.757e − 4
mean mlp	0.796 ± 5.046e − 4	0.994 ± 7.89e − 5	0.793 ± 3.061e − 4	0.978 ± 3.035e − 4	0.754 ± 1.199e − 3	0.934 ± 1.003e − 3
mean patchtsmixer	0.805 ± 2.516e − 3	1.014 ± 5.21e − 3	0.803 ± 4.826e − 3	0.997 ± 5.996e − 3	0.762 ± 8.643e − 3	0.956 ± 8.724e − 3
mean timemixer	0.795 ± 1.078e − 3	0.995 ± 1.274e − 4	0.793 ± 1.079e − 4	0.978 ± 4.193e − 5	0.753 ± 8.874e − 4	0.935 ± 1.971e − 4
No_sum_CoT CoT +TFM lstm	0.8 ± 4.027e − 4	1 ± 7.5e − 4	0.802 ± 9.292e − 5	0.999 ± 2.042e − 4	0.807 ± 2.136e − 4	0.998 ± 8.445e − 4
No_sum_CoT CoT +TFM mlp	0.8 ± 1.12e − 3	1.001 ± 7.881e − 4	0.802 ± 8.888e − 5	1 ± 1.68e − 4	0.81 ± 8.729e − 4	1 ± 1.419e − 4
No_sum_CoT CoT +TFM patchtsmixer	0.801 ± 1.353e − 3	1.012 ± 3.439e − 3	0.809 ± 3.78e − 3	1.012 ± 5.47e − 3	0.816 ± 4.633e − 3	1.013 ± 2.616e − 3
No_sum_CoT CoT +TFM timemixer	0.8 ± 1.564e − 3	1 ± 4.1e − 4	0.802 ± 1.06e − 4	1 ± 1e − 5	0.81 ± 5.75e − 4	1.001 ± 1.735e − 4
No_sum_CoT lstm	0.799 ± 7.968e − 4	1 ± 1.644e − 4	0.802 ± 1.173e − 4	1 ± 4.272e − 5	0.808 ± 4.664e − 4	1 ± 9.815e − 5
No_sum_CoT mlp	0.802 ± 1.165e − 3	1.001 ± 5.263e − 4	0.802 ± 1.573e − 4	1 ± 7.659e − 5	0.811 ± 1.329e − 3	1.001 ± 7.081e − 4
No_sum_CoT patchtsmixer	0.8 ± 7.784e − 3	1.008 ± 3.88e − 3	0.805 ± 2.973e − 3	1.004 ± 2.265e − 3	0.813 ± 5.557e − 3	1.007 ± 3.969e − 3
No_sum_CoT timemixer	0.799 ± 8.839e − 4	1 ± 1.209e − 4	0.802 ± 1.038e − 4	1 ± 5.774e − 6	0.81 ± 1.31e − 3	1.001 ± 3.4e − 4
No_sum_ICD ICD +TFM lstm	0.797 ± 4.851e − 4	0.999 ± 2.136e − 4	0.802 ± 1.082e − 4	0.999 ± 1.85e − 4	0.808 ± 8.193e − 4	0.999 ± 9.914e − 4
No_sum_ICD ICD +TFM mlp	0.8 ± 5.373e − 4	1.001 ± 5.696e − 4	0.802 ± 1.572e − 4	1 ± 2.651e − 4	0.81 ± 2.277e − 3	1 ± 3.089e − 4
No_sum_ICD ICD +TFM patchtsmixer	0.808 ± 1.307e − 3	1.021 ± 3.624e − 3	0.805 ± 7.229e − 4	1.007 ± 2.108e − 3	0.821 ± 4.215e − 3	1.018 ± 5.508e − 3
No_sum_ICD ICD +TFM timemixer	0.8 ± 1.246e − 3	1 ± 2.969e − 4	0.802 ± 5.033e − 5	1 ± 0e0	0.809 ± 3.927e − 4	1 ± 8.387e − 5
No_sum_ICD lstm	0.8 ± 1.627e − 3	1.001 ± 4.364e − 4	0.802 ± 1.964e − 4	1 ± 1.056e − 4	0.809 ± 5.778e − 4	1 ± 2.636e − 4
No_sum_ICD mlp	0.801 ± 7.965e − 4	1.001 ± 1.358e − 4	0.802 ± 8.578e − 5	1 ± 5.323e − 5	0.811 ± 1.656e − 3	1.001 ± 5.301e − 4
No_sum_ICD patchtsmixer	0.798 ± 3.902e − 3	1.005 ± 1.06e − 3	0.803 ± 1.226e − 3	1.004 ± 1.687e − 3	0.814 ± 4.127e − 3	1.007 ± 3.227e − 3
No_sum_ICD timemixer	0.799 ± 3.359e − 4	1 ± 6.481e − 5	0.802 ± 4.646e − 5	1 ± 5.774e − 6	0.809 ± 5.38e − 4	1 ± 1.349e − 4
No_sum_Trend Trend +TFM lstm	0.799 ± 6.553e − 4	0.999 ± 4.143e − 4	0.802 ± 8.083e − 5	0.999 ± 1.908e − 4	0.807 ± 5.631e − 4	0.999 ± 5.116e − 4
No_sum_Trend Trend +TFM mlp	0.8 ± 9.115e − 4	1 ± 6.274e − 4	0.802 ± 2.307e − 4	1 ± 1.861e − 4	0.81 ± 2.249e − 3	1.001 ± 1.942e − 3
No_sum_Trend Trend +TFM patchtsmixer	0.806 ± 8.208e − 4	1.02 ± 7.728e − 3	0.806 ± 7.015e − 4	1.011 ± 1.952e − 3	0.814 ± 1.981e − 3	1.011 ± 1.88e − 3
No_sum_Trend Trend +TFM timemixer	0.8 ± 1.685e − 3	1 ± 4.328e − 4	0.802 ± 1e − 4	1 ± 5.774e − 6	0.809 ± 6.892e − 4	1 ± 1.815e − 4

Table 18: In-distribution results - Age. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	mae	mse	mae	mse	mae	mse
No_sum_Trend lstm	$0.799 \pm 1.177e-3$	$1 \pm 2.004e-4$	$0.802 \pm 5.188e-5$	$1 \pm 1.5e-5$	$0.808 \pm 6.139e-4$	$1 \pm 2.061e-4$
No_sum_Trend mlp	$0.801 \pm 1.357e-3$	$1.001 \pm 6.168e-4$	$0.802 \pm 2.016e-4$	$1 \pm 1.005e-4$	$0.81 \pm 1.434e-3$	$1 \pm 5.804e-4$
No_sum_Trend patchtsmixer	$0.803 \pm 2.482e-3$	$1.006 \pm 2.681e-3$	$0.805 \pm 1.96e-3$	$1.004 \pm 7.381e-4$	$0.812 \pm 1.637e-3$	$1.005 \pm 1.892e-3$
No_sum_Trend timemixer	$0.8 \pm 6.989e-4$	$1 \pm 1.409e-4$	$0.802 \pm 4.933e-5$	$1 \pm 5.774e-6$	$0.809 \pm 1.931e-4$	$1 \pm 5.888e-5$
medgemma CoT +TFM lstm	$0.798 \pm 3.301e-4$	$1 \pm 5.014e-4$	$0.802 \pm 1.617e-4$	$0.999 \pm 1.401e-4$	$0.808 \pm 1.51e-4$	$1 \pm 2.768e-4$
medgemma CoT +TFM mlp	$0.8 \pm 8.638e-4$	$1.001 \pm 4.661e-4$	$0.802 \pm 1.155e-5$	$1 \pm 1.582e-4$	$0.81 \pm 1.223e-3$	$1 \pm 1.119e-3$
medgemma CoT +TFM patchtsmixer	$0.807 \pm 3.471e-3$	$1.024 \pm 1.068e-2$	$0.805 \pm 8.303e-4$	$1.009 \pm 2.689e-3$	$0.815 \pm 1.743e-3$	$1.016 \pm 4.846e-3$
medgemma CoT +TFM timemixer	$0.8 \pm 1.343e-3$	$1.001 \pm 3.816e-4$	$0.802 \pm 8.021e-5$	$1 \pm 5.774e-6$	$0.809 \pm 1.636e-3$	$1.001 \pm 5.227e-4$
medgemma CoT lstm	$0.799 \pm 1.998e-3$	$1 \pm 5.651e-4$	$0.802 \pm 5.62e-5$	$1 \pm 5.774e-5$	$0.809 \pm 1.522e-3$	$1 \pm 3.647e-4$
medgemma CoT mlp	$0.802 \pm 1.233e-3$	$1.001 \pm 7.105e-4$	$0.802 \pm 1.242e-4$	$1 \pm 2.16e-5$	$0.811 \pm 1.828e-3$	$1.001 \pm 1.194e-3$
medgemma CoT patchtsmixer	$0.801 \pm 2.015e-3$	$1.005 \pm 1.861e-3$	$0.805 \pm 1.934e-3$	$1.005 \pm 2.257e-3$	$0.815 \pm 3.297e-3$	$1.006 \pm 2.075e-3$
medgemma CoT timemixer	$0.8 \pm 5.616e-4$	$1 \pm 1.266e-4$	$0.802 \pm 1.491e-4$	$1 \pm 8.165e-6$	$0.809 \pm 1.859e-3$	$1.001 \pm 6.096e-4$
medgemma ICD +TFM lstm	$0.799 \pm 7.548e-4$	$1 \pm 2.272e-4$	$0.802 \pm 1.249e-4$	$1 \pm 2.627e-4$	$0.808 \pm 7.91e-4$	$1 \pm 2.363e-4$
medgemma ICD +TFM mlp	$0.8 \pm 1.484e-3$	$1.001 \pm 1.013e-3$	$0.802 \pm 2.113e-4$	$1 \pm 2.715e-4$	$0.81 \pm 1.608e-3$	$1.001 \pm 9.855e-4$
medgemma ICD +TFM patchtsmixer	$0.807 \pm 3.378e-3$	$1.019 \pm 4.825e-3$	$0.807 \pm 1.582e-3$	$1.01 \pm 5.048e-4$	$0.816 \pm 5.003e-3$	$1.019 \pm 3.843e-3$
medgemma ICD +TFM timemixer	$0.799 \pm 4.107e-4$	$1 \pm 8.544e-5$	$0.802 \pm 5.292e-5$	$1 \pm 5.774e-6$	$0.809 \pm 1.185e-3$	$1 \pm 2.358e-4$
medgemma ICD lstm	$0.799 \pm 1.006e-3$	$1 \pm 2.429e-4$	$0.802 \pm 1.48e-4$	$1 \pm 1.291e-5$	$0.809 \pm 8.411e-4$	$1 \pm 1.823e-4$
medgemma ICD mlp	$0.802 \pm 4.164e-4$	$1.001 \pm 3.267e-4$	$0.802 \pm 2.65e-4$	$1 \pm 1.431e-4$	$0.81 \pm 1.05e-3$	$1.001 \pm 4.768e-4$
medgemma ICD patchtsmixer	$0.802 \pm 1.518e-3$	$1.007 \pm 2.384e-3$	$0.804 \pm 1.19e-3$	$1.003 \pm 9.347e-4$	$0.818 \pm 4.967e-3$	$1.01 \pm 4.364e-3$
medgemma ICD timemixer	$0.799 \pm 6.545e-4$	$1 \pm 1.09e-4$	$0.802 \pm 5.477e-5$	$1 \pm 5e-6$	$0.809 \pm 8.681e-4$	$1 \pm 1.848e-4$
medgemma Trend +TFM lstm	$0.798 \pm 5.125e-4$	$1 \pm 1.361e-4$	$0.802 \pm 1.159e-4$	$1 \pm 5.859e-5$	$0.807 \pm 2.721e-4$	$1 \pm 2.403e-4$
medgemma Trend +TFM mlp	$0.8 \pm 1.693e-3$	$1.001 \pm 1.002e-3$	$0.802 \pm 1.531e-4$	$1 \pm 4.041e-5$	$0.81 \pm 2.434e-3$	$1.001 \pm 1.474e-3$
medgemma Trend +TFM patchtsmixer	$0.806 \pm 9.166e-4$	$1.018 \pm 1.32e-3$	$0.806 \pm 2.89e-3$	$1.011 \pm 4.614e-3$	$0.815 \pm 2.136e-3$	$1.015 \pm 2.962e-3$
medgemma Trend +TFM timemixer	$0.799 \pm 1.216e-3$	$1 \pm 2.237e-4$	$0.802 \pm 9.539e-5$	$1 \pm 1e-5$	$0.809 \pm 1.571e-3$	$1.001 \pm 5.096e-4$
medgemma Trend lstm	$0.799 \pm 5.958e-4$	$1 \pm 2.047e-4$	$0.802 \pm 7.32e-5$	$1 \pm 3.096e-5$	$0.808 \pm 1.07e-3$	$1 \pm 3.157e-4$
medgemma Trend mlp	$0.801 \pm 1.054e-3$	$1.001 \pm 7.023e-4$	$0.802 \pm 1.398e-4$	$1 \pm 1.742e-4$	$0.811 \pm 6.312e-4$	$1.001 \pm 4.551e-4$
medgemma Trend patchtsmixer	$0.799 \pm 2.469e-3$	$1.004 \pm 1.798e-3$	$0.804 \pm 7.189e-4$	$1.003 \pm 9.537e-4$	$0.813 \pm 2.898e-3$	$1.006 \pm 2.387e-3$
medgemma Trend timemixer	$0.799 \pm 8.691e-4$	$1 \pm 1.797e-4$	$0.802 \pm 1.252e-4$	$1 \pm 1.291e-5$	$0.808 \pm 6.085e-4$	$1 \pm 8.756e-5$
medgemma zero_shot +TFM lstm	$0.799 \pm 6.902e-4$	$1 \pm 3.201e-4$	$0.802 \pm 4.163e-5$	$1 \pm 1.967e-4$	$0.807 \pm 1.253e-4$	$1 \pm 2.721e-4$
medgemma zero_shot +TFM mlp	$0.8 \pm 6.208e-4$	$1 \pm 6.301e-4$	$0.802 \pm 1.007e-4$	$1 \pm 7.572e-5$	$0.81 \pm 1.547e-3$	$1.001 \pm 1.437e-3$
medgemma zero_shot +TFM patchtsmixer	$0.806 \pm 3.184e-3$	$1.019 \pm 6.825e-3$	$0.805 \pm 1.699e-3$	$1.008 \pm 6.408e-4$	$0.818 \pm 3.971e-3$	$1.016 \pm 6.382e-3$
medgemma zero_shot +TFM timemixer	$0.799 \pm 1.214e-3$	$1 \pm 2.055e-4$	$0.802 \pm 2.517e-5$	$1 \pm 0e0$	$0.809 \pm 1.219e-3$	$1 \pm 2.479e-4$
medgemma zero_shot lstm	$0.799 \pm 1.36e-3$	$1 \pm 3.178e-4$	$0.802 \pm 1.184e-4$	$1 \pm 2.944e-5$	$0.808 \pm 3.326e-4$	$1 \pm 1.36e-4$
medgemma zero_shot mlp	$0.8 \pm 9.336e-4$	$1 \pm 5.176e-4$	$0.802 \pm 1.412e-4$	$1 \pm 6e-5$	$0.811 \pm 1.18e-3$	$1.001 \pm 3.718e-4$
medgemma zero_shot patchtsmixer	$0.801 \pm 1.17e-3$	$1.006 \pm 1.53e-3$	$0.803 \pm 2.27e-3$	$1.005 \pm 1.433e-3$	$0.814 \pm 1.797e-3$	$1.005 \pm 1.552e-3$
medgemma zero_shot timemixer	$0.799 \pm 7.274e-4$	$1 \pm 1.086e-4$	$0.802 \pm 1.065e-4$	$1 \pm 1.258e-5$	$0.808 \pm 3.126e-4$	$1 \pm 6.397e-5$
No_sum_zero_shot zero_shot +TFM lstm	$0.799 \pm 2.811e-4$	$1 \pm 2.29e-4$	$0.802 \pm 8.083e-5$	$0.999 \pm 6.351e-5$	$0.807 \pm 1.231e-3$	$0.999 \pm 6.614e-4$
No_sum_zero_shot zero_shot +TFM mlp	$0.8 \pm 1.34e-3$	$1 \pm 6.005e-4$	$0.802 \pm 2.732e-4$	$1 \pm 3.02e-4$	$0.81 \pm 4.343e-4$	$1 \pm 4.859e-4$
No_sum_zero_shot zero_shot +TFM patchtsmixer	$0.803 \pm 1.455e-3$	$1.016 \pm 4.599e-3$	$0.806 \pm 1.482e-3$	$1.011 \pm 1.815e-3$	$0.82 \pm 3.048e-3$	$1.016 \pm 3.578e-3$
No_sum_zero_shot zero_shot +TFM timemixer	$0.8 \pm 6.369e-4$	$1 \pm 1.5e-4$	$0.802 \pm 1.35e-4$	$1 \pm 2.082e-5$	$0.809 \pm 2.229e-3$	$1.001 \pm 7.192e-4$
No_sum_zero_shot lstm	$0.801 \pm 1.9e-3$	$1.001 \pm 5.583e-4$	$0.802 \pm 7.188e-5$	$1 \pm 4.193e-5$	$0.809 \pm 4.696e-4$	$1 \pm 1.339e-4$
No_sum_zero_shot mlp	$0.801 \pm 1.75e-3$	$1.001 \pm 8.046e-4$	$0.802 \pm 1.159e-4$	$1 \pm 3.862e-5$	$0.811 \pm 1.089e-3$	$1.001 \pm 3.03e-4$
No_sum_zero_shot patchtsmixer	$0.802 \pm 2.277e-3$	$1.005 \pm 2.572e-3$	$0.804 \pm 2.032e-3$	$1.004 \pm 1.011e-3$	$0.813 \pm 4.082e-3$	$1.005 \pm 2.859e-3$
No_sum_zero_shot timemixer	$0.8 \pm 6.189e-4$	$1 \pm 1.24e-4$	$0.802 \pm 1.409e-4$	$1 \pm 1.893e-5$	$0.81 \pm 5.082e-4$	$1.001 \pm 1.623e-4$
right lstm	$0.795 \pm 7.442e-4$	$0.994 \pm 7.231e-4$	$0.793 \pm 2.285e-4$	$0.977 \pm 1.96e-4$	$0.754 \pm 8.797e-4$	$0.934 \pm 1.571e-4$
right mlp	$0.795 \pm 7.886e-4$	$0.993 \pm 2.05e-4$	$0.793 \pm 1.916e-4$	$0.977 \pm 2.916e-4$	$0.754 \pm 2.142e-3$	$0.933 \pm 9.933e-4$
right patchtsmixer	$0.803 \pm 3.212e-3$	$1.015 \pm 3.019e-3$	$0.798 \pm 3.152e-3$	$0.988 \pm 3.915e-3$	$0.768 \pm 4.237e-3$	$0.947 \pm 3.971e-3$
right timemixer	$0.795 \pm 1.039e-3$	$0.995 \pm 1.193e-4$	$0.793 \pm 8.347e-5$	$0.978 \pm 2.062e-5$	$0.753 \pm 1.141e-3$	$0.935 \pm 2.627e-4$

Table 19: In-distribution results - Age. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	mae	mse	mae	mse
Llama-3.1 CoT +TFM lstm	0.808 ± 1.305e − 3	1 ± 4.309e − 4	0.806 ± 1.069e − 4	1 ± 6.028e − 5
Llama-3.1 CoT +TFM mlp	0.809 ± 7.101e − 4	1.001 ± 6.509e − 4	0.806 ± 1.457e − 4	1 ± 3.215e − 5
Llama-3.1 CoT +TFM patchtsmixer	0.82 ± 2.642e − 3	1.019 ± 4.362e − 3	0.81 ± 2.535e − 3	1.01 ± 6.56e − 3
Llama-3.1 CoT +TFM timemixer	0.809 ± 1.014e − 3	1 ± 2.955e − 4	0.806 ± 1.986e − 4	1 ± 1e − 5
Llama-3.1 CoT lstm	0.809 ± 2.399e − 4	1 ± 9.274e − 5	0.806 ± 1.097e − 4	1 ± 2.944e − 5
Llama-3.1 CoT mlp	0.811 ± 9.011e − 4	1.001 ± 2.553e − 4	0.806 ± 2.191e − 4	1 ± 4.349e − 5
Llama-3.1 CoT patchtsmixer	0.813 ± 3.711e − 3	1.008 ± 3.914e − 3	0.809 ± 3.074e − 3	1.004 ± 3.167e − 3
Llama-3.1 CoT timemixer	0.809 ± 9.832e − 5	1 ± 2.363e − 5	0.806 ± 9.743e − 5	1 ± 5.774e − 6
Llama-3.1 ICD +TFM lstm	0.808 ± 1.906e − 3	1.001 ± 6.766e − 4	0.806 ± 1.652e − 4	1 ± 1.193e − 4
Llama-3.1 ICD +TFM mlp	0.81 ± 7.51e − 4	1.001 ± 6.311e − 4	0.806 ± 1.825e − 4	1 ± 3.786e − 5
Llama-3.1 ICD +TFM patchtsmixer	0.815 ± 2.4e − 3	1.016 ± 5.238e − 3	0.811 ± 3.936e − 3	1.01 ± 3.038e − 3
Llama-3.1 ICD +TFM timemixer	0.809 ± 4.583e − 4	1 ± 8.544e − 5	0.806 ± 2.517e − 5	1 ± 0e0
Llama-3.1 ICD lstm	0.809 ± 8.811e − 4	1 ± 3.636e − 4	0.806 ± 2.156e − 4	1 ± 1.258e − 5
Llama-3.1 ICD mlp	0.811 ± 2.589e − 4	1.001 ± 4.671e − 4	0.806 ± 2.696e − 4	1 ± 9.287e − 5
Llama-3.1 ICD patchtsmixer	0.812 ± 3.624e − 3	1.007 ± 1.829e − 3	0.806 ± 2.141e − 3	1.003 ± 5.351e − 4
Llama-3.1 ICD timemixer	0.808 ± 1.322e − 3	1 ± 1.954e − 4	0.806 ± 1.735e − 4	1 ± 1.155e − 5
Llama-3.1 Trend +TFM lstm	0.808 ± 7.317e − 4	1 ± 1.686e − 4	0.806 ± 1.852e − 4	1 ± 2.122e − 4
Llama-3.1 Trend +TFM mlp	0.81 ± 5.956e − 4	1.002 ± 7.123e − 4	0.806 ± 2.358e − 4	1 ± 1.332e − 4
Llama-3.1 Trend +TFM patchtsmixer	0.817 ± 1.749e − 3	1.015 ± 1.827e − 3	0.81 ± 1.337e − 3	1.008 ± 2.62e − 3
Llama-3.1 Trend +TFM timemixer	0.809 ± 1.489e − 3	1 ± 3.053e − 4	0.806 ± 1.429e − 4	1 ± 5.774e − 6
Llama-3.1 Trend lstm	0.809 ± 6.381e − 4	1 ± 1.916e − 4	0.806 ± 6.055e − 5	1 ± 1.204e − 4
Llama-3.1 Trend mlp	0.81 ± 1.442e − 3	1.001 ± 6.115e − 4	0.806 ± 2.001e − 4	1 ± 5.56e − 5
Llama-3.1 Trend patchtsmixer	0.814 ± 5.117e − 3	1.006 ± 2.978e − 3	0.808 ± 2.131e − 3	1.003 ± 8.315e − 4
Llama-3.1 Trend timemixer	0.809 ± 1.561e − 3	1.001 ± 5.321e − 4	0.806 ± 7.544e − 5	1 ± 5.774e − 6
Llama-3.1 zero_shot +TFM lstm	0.808 ± 1.195e − 3	1 ± 2.488e − 4	0.806 ± 2.303e − 4	1 ± 6.245e − 5
Llama-3.1 zero_shot +TFM mlp	0.809 ± 1.027e − 3	1.001 ± 5.934e − 4	0.806 ± 6.986e − 4	1 ± 4.151e − 4
Llama-3.1 zero_shot +TFM patchtsmixer	0.818 ± 3.996e − 3	1.017 ± 4.479e − 3	0.81 ± 3.067e − 3	1.011 ± 4.664e − 3
Llama-3.1 zero_shot +TFM timemixer	0.809 ± 2.103e − 4	1 ± 5.568e − 5	0.806 ± 5.774e − 5	1 ± 5.774e − 6
Llama-3.1 zero_shot lstm	0.809 ± 6.559e − 4	1 ± 1.919e − 4	0.806 ± 1.746e − 4	1 ± 3.5e − 5
Llama-3.1 zero_shot mlp	0.811 ± 6.466e − 4	1.001 ± 3.994e − 4	0.806 ± 2.325e − 4	1 ± 1.041e − 4
Llama-3.1 zero_shot patchtsmixer	0.812 ± 2.08e − 3	1.006 ± 1.313e − 3	0.81 ± 2.757e − 3	1.006 ± 2.814e − 3
Llama-3.1 zero_shot timemixer	0.809 ± 8.806e − 4	1 ± 2.117e − 4	0.806 ± 2.986e − 5	1 ± 0e0
TFM lstm	0.752 ± 8.624e − 4	0.935 ± 1.439e − 4	0.75 ± 1.715e − 4	0.935 ± 2.179e − 4
TFM mlp	0.753 ± 1.029e − 3	0.935 ± 1.105e − 4	0.75 ± 9.912e − 5	0.935 ± 2.134e − 4
TFM patchtsmixer	0.761 ± 3.671e − 3	0.947 ± 1.981e − 3	0.753 ± 1.024e − 3	0.943 ± 1.584e − 3
TFM timemixer	0.753 ± 8.952e − 4	0.934 ± 2.106e − 4	0.75 ± 2.007e − 4	0.934 ± 4.435e − 5
TSDE lstm	0.751 ± 4.323e − 3	0.936 ± 9.137e − 4	0.751 ± 3.034e − 4	0.935 ± 7.839e − 4
TSDE mlp	0.748 ± 1.451e − 3	0.939 ± 5.069e − 3	0.752 ± 8.398e − 3	0.936 ± 5.379e − 3
TSDE patchtsmixer	0.762 ± 2.922e − 3	0.951 ± 3.481e − 3	0.755 ± 2.716e − 3	0.944 ± 6.36e − 3
TSDE timemixer	0.753 ± 8.294e − 4	0.934 ± 1.949e − 4	0.75 ± 1.949e − 4	0.934 ± 4.646e − 5
gemini-2.0-flash CoT +TFM lstm	0.807 ± 1.032e − 3	1 ± 6.643e − 4	0.806 ± 1.493e − 4	0.999 ± 3.544e − 4
gemini-2.0-flash CoT +TFM mlp	0.81 ± 1.06e − 3	1.001 ± 5.029e − 4	0.806 ± 2.875e − 4	1 ± 9.074e − 5
gemini-2.0-flash CoT +TFM patchtsmixer	0.814 ± 3.898e − 3	1.016 ± 6.893e − 3	0.809 ± 3.754e − 3	1.011 ± 5.063e − 3
gemini-2.0-flash CoT +TFM timemixer	0.809 ± 1.33e − 3	1 ± 3.384e − 4	0.806 ± 1.955e − 4	1 ± 1.528e − 5
gemini-2.0-flash CoT lstm	0.809 ± 6.393e − 4	1 ± 2.156e − 4	0.806 ± 3.094e − 4	1 ± 4.5e − 5
gemini-2.0-flash CoT mlp	0.81 ± 5.832e − 4	1.001 ± 1.909e − 4	0.806 ± 9.032e − 5	1 ± 1.152e − 4
gemini-2.0-flash CoT patchtsmixer	0.813 ± 7.459e − 3	1.007 ± 5.039e − 3	0.808 ± 1.732e − 3	1.003 ± 7.908e − 4
gemini-2.0-flash CoT timemixer	0.808 ± 3.555e − 4	1 ± 6.397e − 5	0.806 ± 1.782e − 4	1 ± 5.774e − 6
gemini-2.0-flash ICD +TFM lstm	0.808 ± 4.688e − 4	1 ± 7.38e − 4	0.806 ± 1.361e − 4	1 ± 1.32e − 4
gemini-2.0-flash ICD +TFM mlp	0.81 ± 1.535e − 3	1.001 ± 4.917e − 4	0.806 ± 2.272e − 4	1 ± 4.293e − 4
gemini-2.0-flash ICD +TFM patchtsmixer	0.818 ± 8.371e − 3	1.02 ± 1.025e − 2	0.809 ± 1.245e − 3	1.009 ± 2.844e − 3
gemini-2.0-flash ICD +TFM timemixer	0.809 ± 4.362e − 4	1 ± 9.539e − 5	0.806 ± 1.127e − 4	1 ± 1.155e − 5
gemini-2.0-flash ICD lstm	0.809 ± 1.269e − 3	1.001 ± 3.017e − 4	0.806 ± 1.464e − 4	1 ± 8.098e − 5
gemini-2.0-flash ICD mlp	0.811 ± 8.003e − 4	1.001 ± 2.453e − 4	0.806 ± 2.757e − 4	1 ± 4.031e − 5
gemini-2.0-flash ICD patchtsmixer	0.812 ± 2.618e − 3	1.006 ± 1.148e − 3	0.809 ± 5.485e − 3	1.005 ± 1.02e − 3
gemini-2.0-flash ICD timemixer	0.809 ± 8.427e − 4	1 ± 2.123e − 4	0.806 ± 4.717e − 5	1 ± 8.165e − 6
gemini-2.0-flash Trend +TFM lstm	0.807 ± 6.799e − 4	1 ± 4.355e − 4	0.806 ± 2.787e − 4	1 ± 1.25e − 4
gemini-2.0-flash Trend +TFM mlp	0.81 ± 8.972e − 4	1.001 ± 3.132e − 4	0.806 ± 4.257e − 4	1 ± 3.781e − 4
gemini-2.0-flash Trend +TFM patchtsmixer	0.821 ± 1.446e − 3	1.02 ± 3.373e − 3	0.809 ± 1.505e − 3	1.012 ± 4.289e − 3
gemini-2.0-flash Trend +TFM timemixer	0.81 ± 6.006e − 4	1.001 ± 2.248e − 4	0.806 ± 1.589e − 4	1 ± 1.155e − 5
gemini-2.0-flash Trend lstm	0.809 ± 1.841e − 3	1.001 ± 5.033e − 4	0.806 ± 2.012e − 4	1 ± 1.688e − 4
gemini-2.0-flash Trend mlp	0.81 ± 7.437e − 4	1.001 ± 5.1e − 4	0.806 ± 2.729e − 4	1 ± 4.573e − 5
gemini-2.0-flash Trend patchtsmixer	0.814 ± 5.695e − 3	1.009 ± 6.999e − 3	0.807 ± 4.126e − 3	1.005 ± 2.361e − 3
gemini-2.0-flash Trend timemixer	0.809 ± 1.008e − 3	1 ± 2.575e − 4	0.806 ± 1.406e − 4	1 ± 5e − 6
gemini-2.0-flash zero_shot +TFM lstm	0.808 ± 4.114e − 4	1 ± 3.051e − 4	0.806 ± 4.583e − 5	1 ± 1.825e − 4
gemini-2.0-flash zero_shot +TFM mlp	0.81 ± 1.831e − 3	1.002 ± 5.443e − 4	0.806 ± 2.179e − 4	1 ± 2.193e − 4
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.816 ± 2.724e − 3	1.015 ± 4.367e − 3	0.809 ± 4.541e − 3	1.009 ± 5.198e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.809 ± 9.551e − 4	1 ± 2.359e − 4	0.806 ± 4.966e − 4	1 ± 5.774e − 6
gemini-2.0-flash zero_shot lstm	0.808 ± 9.849e − 4	1 ± 2.015e − 4	0.806 ± 1.696e − 4	1 ± 4.435e − 5
gemini-2.0-flash zero_shot mlp	0.81 ± 6.791e − 4	1.001 ± 3.608e − 4	0.806 ± 9.592e − 5	1 ± 7.5e − 5
gemini-2.0-flash zero_shot patchtsmixer	0.814 ± 2.721e − 3	1.007 ± 3.521e − 3	0.809 ± 5.301e − 3	1.005 ± 2.145e − 3
gemini-2.0-flash zero_shot timemixer	0.809 ± 6.028e − 4	1 ± 1.916e − 4	0.806 ± 2.674e − 4	1 ± 8.165e − 6
interp lstm	0.751 ± 8.926e − 4	0.933 ± 1.031e − 3	0.751 ± 1.276e − 3	0.935 ± 7.468e − 4
interp mlp	0.754 ± 6.643e − 4	0.939 ± 3.273e − 3	0.752 ± 1.327e − 3	0.938 ± 4.643e − 3
interp patchtsmixer	0.785 ± 2.196e − 2	0.998 ± 2.299e − 2	0.778 ± 6.66e − 3	0.976 ± 8.186e − 3
interp timemixer	0.753 ± 7.487e − 4	0.934 ± 1.144e − 4	0.75 ± 7.165e − 5	0.935 ± 2.449e − 5
mean lstm	0.752 ± 1.66e − 3	0.934 ± 3.463e − 4	0.75 ± 1.994e − 3	0.934 ± 8.96e − 4
mean mlp	0.752 ± 1.087e − 3	0.933 ± 1.557e − 3	0.751 ± 1.89e − 3	0.936 ± 1.947e − 3
mean patchtsmixer	0.805 ± 7.924e − 3	1.022 ± 1.104e − 2	0.786 ± 1.582e − 2	0.991 ± 2.368e − 2
mean timemixer	0.753 ± 7.514e − 4	0.934 ± 1.1e − 4	0.75 ± 1.05e − 4	0.934 ± 2.363e − 5
No_sum_CoT CoT +TFM lstm	0.809 ± 2.112e − 3	1 ± 6.208e − 4	0.805 ± 7.478e − 4	1 ± 4.276e − 4
No_sum_CoT CoT +TFM mlp	0.809 ± 4.597e − 4	1.001 ± 6.295e − 4	0.806 ± 5.831e − 4	1 ± 4.072e − 4
No_sum_CoT CoT +TFM patchtsmixer	0.813 ± 2.364e − 4	1.012 ± 1.163e − 3	0.815 ± 5.372e − 3	1.013 ± 5.639e − 3
No_sum_CoT CoT +TFM timemixer	0.809 ± 1.329e − 3	1 ± 4.188e − 4	0.806 ± 1.617e − 4	1 ± 1e − 5
No_sum_CoT lstm	0.809 ± 7.066e − 4	1 ± 1.576e − 4	0.806 ± 1.992e − 4	1 ± 1.435e − 4
No_sum_CoT mlp	0.811 ± 7.703e − 4	1.001 ± 8.169e − 4	0.806 ± 2.83e − 4	1 ± 7.5e − 5
No_sum_CoT patchtsmixer	0.811 ± 6.678e − 3	1.007 ± 3.376e − 3	0.809 ± 4.199e − 3	1.003 ± 2.354e − 3
No_sum_CoT timemixer	0.808 ± 7.487e − 4	1 ± 1.182e − 4	0.806 ± 1.556e − 4	1 ± 9.574e − 6
No_sum_ICD ICD +TFM lstm	0.807 ± 8.242e − 4	1 ± 6.817e − 4	0.806 ± 1.559e − 4	0.999 ± 1.57e − 4
No_sum_ICD ICD +TFM mlp	0.81 ± 1.585e − 3	1 ± 1.733e − 3	0.806 ± 4.96e − 4	1.001 ± 7.355e − 4
No_sum_ICD ICD +TFM patchtsmixer	0.82 ± 5.856e − 4	1.019 ± 2.224e − 3	0.808 ± 1.143e − 3	1.006 ± 3.275e − 3
No_sum_ICD ICD +TFM timemixer	0.809 ± 1.065e − 3	1 ± 3.055e − 4	0.806 ± 7.506e − 5	1 ± 5.774e − 6
No_sum_ICD lstm	0.809 ± 1.448e − 3	1.001 ± 4.847e − 4	0.806 ± 4.021e − 4	1 ± 1.452e − 4
No_sum_ICD mlp	0.81 ± 7.661e − 4	1.001 ± 1.434e − 4	0.806 ± 1.575e − 4	1 ± 8.421e − 5
No_sum_ICD patchtsmixer	0.81 ± 3.309e − 3	1.006 ± 2.53e − 3	0.807 ± 2.762e − 3	1.004 ± 2.054e − 3
No_sum_ICD timemixer	0.809 ± 2.861e − 4	1 ± 6.481e − 5	0.806 ± 7.394e − 5	1 ± 5.774e − 6
No_sum_Trend Trend +TFM lstm	0.807 ± 4.004e − 4	0.999 ± 1.106e − 3	0.806 ± 5.069e − 4	0.999 ± 8.869e − 4
No_sum_Trend Trend +TFM mlp	0.81 ± 8.718e − 5	1.001 ± 1e − 3	0.806 ± 3.279e − 4	1 ± 3.315e − 4
No_sum_Trend Trend +TFM patchtsmixer	0.818 ± 1.8e − 3	1.02 ± 6.814e − 3	0.808 ± 5.538e − 4	1.009 ± 2.116e − 3
No_sum_Trend Trend +TFM timemixer	0.809 ± 1.429e − 3	1 ± 4.236e − 4	0.806 ± 1.501e − 4	1 ± 1.155e − 5

Table 20: Cross-site transfer results - Age. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	mae	mse	mae	mse
No_sum_Trend lstm	$0.808 \pm 1.044e-3$	$1 \pm 3.807e-4$	$0.806 \pm 1.065e-4$	$1 \pm 1.88e-4$
No_sum_Trend mlp	$0.81 \pm 1.264e-3$	$1 \pm 7.27e-4$	$0.806 \pm 3.889e-4$	$1 \pm 2.887e-5$
No_sum_Trend patchtsmixer	$0.814 \pm 3.215e-3$	$1.007 \pm 3.595e-3$	$0.808 \pm 4.422e-3$	$1.005 \pm 3.252e-3$
No_sum_Trend timemixer	$0.809 \pm 5.934e-4$	$1 \pm 1.409e-4$	$0.806 \pm 7.5e-5$	$1 \pm 5e-6$
medgemma CoT +TFM lstm	$0.807 \pm 6.886e-4$	$1 \pm 3.897e-4$	$0.806 \pm 2.281e-4$	$1 \pm 1.756e-4$
medgemma CoT +TFM mlp	$0.81 \pm 9.819e-4$	$1.002 \pm 4.46e-4$	$0.806 \pm 7.937e-5$	$1 \pm 3.215e-5$
medgemma CoT +TFM patchtsmixer	$0.819 \pm 4.652e-3$	$1.026 \pm 1.106e-2$	$0.809 \pm 1.123e-3$	$1.009 \pm 1.528e-3$
medgemma CoT +TFM timemixer	$0.81 \pm 1.138e-3$	$1.001 \pm 3.758e-4$	$0.806 \pm 1.301e-4$	$1 \pm 5.774e-6$
medgemma CoT lstm	$0.808 \pm 1.51e-3$	$1 \pm 4.267e-4$	$0.806 \pm 5.745e-5$	$1 \pm 4.573e-5$
medgemma CoT mlp	$0.811 \pm 1.383e-3$	$1.001 \pm 6.851e-4$	$0.806 \pm 1.718e-4$	$1 \pm 9.5e-5$
medgemma CoT patchtsmixer	$0.812 \pm 2.312e-3$	$1.006 \pm 1.899e-3$	$0.811 \pm 2.944e-3$	$1.006 \pm 2.257e-3$
medgemma CoT timemixer	$0.809 \pm 4.816e-4$	$1 \pm 1.329e-4$	$0.806 \pm 2.278e-4$	$1 \pm 9.574e-6$
medgemma ICD +TFM lstm	$0.809 \pm 5.368e-4$	$1 \pm 2.483e-4$	$0.806 \pm 1.159e-4$	$1 \pm 2.397e-4$
medgemma ICD +TFM mlp	$0.811 \pm 1.223e-3$	$1.002 \pm 1.109e-3$	$0.806 \pm 1.345e-4$	$1 \pm 4.107e-4$
medgemma ICD +TFM patchtsmixer	$0.819 \pm 2.332e-3$	$1.018 \pm 3.501e-3$	$0.81 \pm 2.333e-3$	$1.009 \pm 1.351e-3$
medgemma ICD +TFM timemixer	$0.809 \pm 3.502e-4$	$1 \pm 7.506e-5$	$0.806 \pm 7.572e-5$	$1 \pm 1.155e-5$
medgemma ICD lstm	$0.808 \pm 8.344e-4$	$1 \pm 1.992e-4$	$0.806 \pm 1.976e-4$	$1 \pm 4.349e-5$
medgemma ICD mlp	$0.811 \pm 5.262e-4$	$1.001 \pm 5.917e-4$	$0.806 \pm 3.189e-4$	$1 \pm 1.192e-4$
medgemma ICD patchtsmixer	$0.813 \pm 4.584e-4$	$1.007 \pm 3.122e-3$	$0.809 \pm 2.417e-3$	$1.004 \pm 1.357e-3$
medgemma ICD timemixer	$0.808 \pm 5.545e-4$	$1 \pm 1.08e-4$	$0.806 \pm 8.655e-5$	$1 \pm 0e0$
medgemma Trend +TFM lstm	$0.807 \pm 1.015e-3$	$1 \pm 1.852e-4$	$0.806 \pm 1.069e-4$	$1 \pm 1.701e-4$
medgemma Trend +TFM mlp	$0.81 \pm 1.065e-3$	$1.001 \pm 9.487e-4$	$0.806 \pm 1.931e-4$	$1 \pm 1.222e-4$
medgemma Trend +TFM patchtsmixer	$0.818 \pm 1.52e-3$	$1.017 \pm 3.255e-3$	$0.809 \pm 3.577e-3$	$1.011 \pm 5.994e-3$
medgemma Trend +TFM timemixer	$0.809 \pm 1.036e-3$	$1 \pm 2.227e-4$	$0.806 \pm 1.501e-4$	$1 \pm 1e-5$
medgemma Trend lstm	$0.808 \pm 4.689e-4$	$1 \pm 1.658e-4$	$0.806 \pm 1.389e-4$	$1 \pm 4.031e-5$
medgemma Trend mlp	$0.81 \pm 9.645e-4$	$1.001 \pm 5.056e-4$	$0.806 \pm 1.05e-4$	$1 \pm 1.323e-4$
medgemma Trend patchtsmixer	$0.811 \pm 2.372e-3$	$1.006 \pm 1.387e-3$	$0.807 \pm 1.916e-3$	$1.004 \pm 4.877e-4$
medgemma Trend timemixer	$0.809 \pm 7.393e-4$	$1 \pm 1.797e-4$	$0.806 \pm 1.977e-4$	$1 \pm 9.574e-6$
medgemma zero_shot +TFM lstm	$0.808 \pm 1.404e-3$	$1 \pm 8.8e-4$	$0.806 \pm 1.114e-4$	$1 \pm 1.629e-4$
medgemma zero_shot +TFM mlp	$0.81 \pm 1.093e-3$	$1.001 \pm 5.905e-4$	$0.806 \pm 1.206e-4$	$1 \pm 1.277e-4$
medgemma zero_shot +TFM patchtsmixer	$0.819 \pm 2.865e-3$	$1.019 \pm 5.564e-3$	$0.809 \pm 2.487e-3$	$1.008 \pm 1.207e-3$
medgemma zero_shot +TFM timemixer	$0.809 \pm 1.033e-3$	$1 \pm 2.088e-4$	$0.806 \pm 3.606e-5$	$1 \pm 5.774e-6$
medgemma zero_shot lstm	$0.808 \pm 1.186e-3$	$1 \pm 3.106e-4$	$0.806 \pm 2.29e-4$	$1 \pm 8.347e-5$
medgemma zero_shot mlp	$0.809 \pm 7.985e-4$	$1 \pm 4.812e-4$	$0.806 \pm 2.216e-4$	$1 \pm 4.272e-5$
medgemma zero_shot patchtsmixer	$0.813 \pm 2.217e-3$	$1.007 \pm 3.561e-3$	$0.805 \pm 4.02e-3$	$1.005 \pm 1.791e-3$
medgemma zero_shot timemixer	$0.808 \pm 6.186e-4$	$1 \pm 1.103e-4$	$0.806 \pm 1.676e-4$	$1 \pm 1.291e-5$
No_sum_zero_shot zero_shot +TFM lstm	$0.809 \pm 3.194e-4$	$1.001 \pm 4.528e-4$	$0.806 \pm 1.311e-3$	$1 \pm 1.171e-3$
No_sum_zero_shot zero_shot +TFM mlp	$0.81 \pm 2.697e-3$	$1 \pm 3.251e-3$	$0.805 \pm 9.241e-4$	$1 \pm 3.913e-4$
No_sum_zero_shot zero_shot +TFM patchtsmixer	$0.816 \pm 1.487e-3$	$1.016 \pm 5.008e-3$	$0.81 \pm 1.121e-3$	$1.01 \pm 1.845e-3$
No_sum_zero_shot zero_shot +TFM timemixer	$0.809 \pm 5.515e-4$	$1 \pm 1.752e-4$	$0.806 \pm 2.05e-4$	$1 \pm 1.528e-5$
No_sum_zero_shot lstm	$0.81 \pm 1.741e-3$	$1.001 \pm 4.948e-4$	$0.806 \pm 1.261e-4$	$1 \pm 5.123e-5$
No_sum_zero_shot mlp	$0.81 \pm 1.638e-3$	$1.001 \pm 9.81e-4$	$0.806 \pm 1.682e-4$	$1 \pm 1.037e-4$
No_sum_zero_shot patchtsmixer	$0.812 \pm 1.574e-3$	$1.007 \pm 2.336e-3$	$0.807 \pm 4.249e-3$	$1.004 \pm 9.251e-4$
No_sum_zero_shot timemixer	$0.809 \pm 5.216e-4$	$1 \pm 1.253e-4$	$0.806 \pm 2.181e-4$	$1 \pm 1.732e-5$
right lstm	$0.751 \pm 9.323e-4$	$0.933 \pm 1.35e-3$	$0.751 \pm 1.816e-3$	$0.935 \pm 5.847e-4$
right mlp	$0.753 \pm 1.278e-3$	$0.933 \pm 1.511e-3$	$0.75 \pm 6.719e-4$	$0.934 \pm 5.928e-4$
right patchtsmixer	$0.792 \pm 1.487e-2$	$0.994 \pm 1.356e-2$	$0.772 \pm 2.593e-3$	$0.966 \pm 5.168e-3$
right timemixer	$0.753 \pm 7.362e-4$	$0.934 \pm 1.1e-4$	$0.75 \pm 1.352e-4$	$0.935 \pm 4.655e-5$

Table 21: Cross-site transfer results - Age. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	auprc	auroc	auprc	auroc	auprc	auroc
Llama-3.1 CoT +TFM lstm	0.682 ± 8.894e − 3	0.555 ± 1.246e − 2	0.655 ± 1.33e − 2	0.585 ± 1.425e − 2	0.693 ± 7.3e − 3	0.573 ± 1.212e − 2
Llama-3.1 CoT +TFM mlp	0.72 ± 7.983e − 3	0.6 ± 1.015e − 2	0.705 ± 1.309e − 2	0.64 ± 1.339e − 2	0.725 ± 1.677e − 2	0.608 ± 2.07e − 2
Llama-3.1 CoT +TFM patchtsmixer	0.704 ± 3.667e − 4	0.58 ± 1.341e − 3	0.702 ± 8.468e − 4	0.63 ± 2e − 4	0.714 ± 5.538e − 3	0.595 ± 6.329e − 3
Llama-3.1 CoT +TFM timemixer	0.707 ± 2.153e − 3	0.588 ± 2.084e − 3	0.696 ± 1.456e − 2	0.631 ± 1.694e − 2	0.7 ± 6.089e − 3	0.584 ± 5.492e − 3
Llama-3.1 CoT lstm	0.702 ± 3.827e − 3	0.579 ± 5.909e − 3	0.685 ± 2.532e − 2	0.612 ± 2.763e − 2	0.696 ± 6.709e − 3	0.573 ± 1.042e − 2
Llama-3.1 CoT mlp	0.704 ± 3.167e − 3	0.581 ± 4.62e − 3	0.702 ± 8.377e − 3	0.634 ± 9.491e − 3	0.707 ± 1.567e − 2	0.586 ± 2.046e − 2
Llama-3.1 CoT patchtsmixer	0.697 ± 3.109e − 3	0.572 ± 3.425e − 3	0.698 ± 1.118e − 2	0.62 ± 1.096e − 2	0.713 ± 1.649e − 2	0.593 ± 2.004e − 2
Llama-3.1 CoT timemixer	0.7 ± 8.593e − 4	0.574 ± 9.507e − 4	0.682 ± 6.902e − 3	0.608 ± 6.689e − 3	0.698 ± 1.743e − 3	0.573 ± 2.178e − 3
Llama-3.1 ICD +TFM lstm	0.681 ± 1.215e − 2	0.555 ± 1.428e − 2	0.624 ± 1.094e − 2	0.554 ± 9.214e − 3	0.692 ± 4.155e − 3	0.567 ± 5.781e − 3
Llama-3.1 ICD +TFM mlp	0.721 ± 1.447e − 2	0.603 ± 1.76e − 2	0.658 ± 1.66e − 2	0.591 ± 1.936e − 2	0.717 ± 8.973e − 3	0.596 ± 1.237e − 2
Llama-3.1 ICD +TFM patchtsmixer	0.708 ± 5.577e − 4	0.587 ± 3.686e − 4	0.657 ± 8.486e − 3	0.587 ± 8.952e − 3	0.719 ± 7.324e − 3	0.598 ± 8.555e − 3
Llama-3.1 ICD +TFM timemixer	0.705 ± 1.578e − 3	0.586 ± 3.109e − 3	0.662 ± 4.638e − 3	0.597 ± 5.18e − 3	0.695 ± 8.775e − 3	0.571 ± 8.912e − 3
Llama-3.1 ICD lstm	0.69 ± 3.973e − 3	0.565 ± 6.009e − 3	0.643 ± 5.652e − 3	0.571 ± 5.622e − 3	0.69 ± 5.427e − 3	0.565 ± 7.829e − 3
Llama-3.1 ICD mlp	0.698 ± 7.389e − 3	0.575 ± 9.757e − 3	0.645 ± 3.347e − 3	0.575 ± 4.081e − 3	0.699 ± 3.943e − 3	0.574 ± 4.436e − 3
Llama-3.1 ICD patchtsmixer	0.692 ± 1.261e − 3	0.569 ± 2.164e − 3	0.648 ± 9.119e − 3	0.577 ± 8.647e − 3	0.689 ± 2.35e − 3	0.561 ± 3.104e − 3
Llama-3.1 ICD timemixer	0.695 ± 1.662e − 3	0.571 ± 1.684e − 3	0.65 ± 2.146e − 3	0.577 ± 2.153e − 3	0.688 ± 1.409e − 3	0.558 ± 2.071e − 3
Llama-3.1 Trend +TFM lstm	0.683 ± 8.034e − 3	0.556 ± 9.091e − 3	0.62 ± 2.583e − 3	0.553 ± 9.563e − 4	0.697 ± 3.774e − 3	0.575 ± 4.552e − 3
Llama-3.1 Trend +TFM mlp	0.713 ± 1.134e − 2	0.593 ± 1.302e − 2	0.669 ± 1.752e − 2	0.599 ± 2.069e − 2	0.711 ± 5.604e − 4	0.587 ± 7.629e − 4
Llama-3.1 Trend +TFM patchtsmixer	0.708 ± 3.699e − 3	0.585 ± 5.518e − 3	0.672 ± 6.595e − 3	0.602 ± 7.253e − 3	0.713 ± 7.321e − 3	0.591 ± 9.864e − 3
Llama-3.1 Trend +TFM timemixer	0.707 ± 8.358e − 3	0.589 ± 9.674e − 3	0.661 ± 7.25e − 3	0.591 ± 7.592e − 3	0.713 ± 3.603e − 3	0.596 ± 4.388e − 3
Llama-3.1 Trend lstm	0.7 ± 4.246e − 3	0.576 ± 6.409e − 3	0.649 ± 4.433e − 3	0.575 ± 4.399e − 3	0.701 ± 5.662e − 3	0.579 ± 7.719e − 3
Llama-3.1 Trend mlp	0.699 ± 5.085e − 3	0.574 ± 8.231e − 3	0.663 ± 8.269e − 3	0.592 ± 9.672e − 3	0.7 ± 5.528e − 3	0.576 ± 6.935e − 3
Llama-3.1 Trend patchtsmixer	0.7 ± 9.98e − 4	0.575 ± 1.986e − 3	0.667 ± 1.93e − 2	0.592 ± 1.845e − 2	0.701 ± 3.676e − 3	0.577 ± 4.355e − 3
Llama-3.1 Trend timemixer	0.701 ± 5.141e − 4	0.575 ± 8.269e − 4	0.664 ± 9.245e − 3	0.589 ± 1.015e − 2	0.7 ± 1.113e − 3	0.574 ± 1.374e − 3
Llama-3.1 zero_shot +TFM lstm	0.687 ± 6.535e − 3	0.56 ± 1.161e − 2	0.64 ± 9.101e − 3	0.573 ± 9.565e − 3	0.687 ± 1.161e − 2	0.564 ± 1.657e − 2
Llama-3.1 zero_shot +TFM mlp	0.719 ± 1.746e − 2	0.602 ± 2.091e − 2	0.729 ± 8.446e − 3	0.675 ± 9.738e − 3	0.716 ± 1.302e − 2	0.595 ± 1.367e − 2
Llama-3.1 zero_shot +TFM patchtsmixer	0.704 ± 3.635e − 4	0.581 ± 3.557e − 4	0.702 ± 2.06e − 2	0.64 ± 2.288e − 2	0.716 ± 5.97e − 3	0.595 ± 6.605e − 3
Llama-3.1 zero_shot +TFM timemixer	0.702 ± 2.551e − 3	0.583 ± 5.376e − 3	0.681 ± 1.312e − 2	0.621 ± 1.569e − 2	0.703 ± 3.164e − 3	0.577 ± 5.926e − 3
Llama-3.1 zero_shot lstm	0.697 ± 3.063e − 3	0.575 ± 5.793e − 3	0.67 ± 1.631e − 2	0.605 ± 1.946e − 2	0.7 ± 8.366e − 3	0.578 ± 1.064e − 2
Llama-3.1 zero_shot mlp	0.7 ± 2.574e − 3	0.577 ± 2.975e − 3	0.693 ± 9.806e − 3	0.635 ± 1.078e − 2	0.698 ± 5.621e − 3	0.573 ± 5.56e − 3
Llama-3.1 zero_shot patchtsmixer	0.698 ± 1.777e − 3	0.575 ± 3.099e − 3	0.709 ± 4.749e − 3	0.651 ± 4.87e − 3	0.699 ± 6.637e − 3	0.574 ± 6.645e − 3
Llama-3.1 zero_shot timemixer	0.702 ± 1.709e − 3	0.579 ± 2.678e − 3	0.699 ± 9.912e − 3	0.639 ± 1.057e − 2	0.696 ± 8.549e − 3	0.569 ± 8.527e − 3
TFM lstm	0.67 ± 5.001e − 3	0.541 ± 4.894e − 3	0.594 ± 4.994e − 3	0.529 ± 3.431e − 3	0.654 ± 5.393e − 3	0.517 ± 3.594e − 3
TFM mlp	0.671 ± 4.555e − 3	0.545 ± 5.417e − 3	0.586 ± 2.113e − 3	0.532 ± 1.159e − 3	0.661 ± 1.712e − 3	0.528 ± 1.28e − 3
TFM patchtsmixer	0.673 ± 1.693e − 3	0.548 ± 8.308e − 4	0.589 ± 1.442e − 3	0.532 ± 5.452e − 4	0.664 ± 6.923e − 4	0.53 ± 1.105e − 3
TFM timemixer	0.671 ± 3.346e − 3	0.549 ± 3.121e − 3	0.601 ± 3.253e − 3	0.536 ± 1.391e − 3	0.661 ± 2.082e − 3	0.526 ± 1.573e − 3
TSDE lstm	0.681 ± 6.848e − 3	0.554 ± 1.058e − 2	0.819 ± 2.7e − 3	0.773 ± 2.974e − 3	0.87 ± 5.957e − 3	0.795 ± 8.274e − 3
TSDE mlp	0.791 ± 3.345e − 3	0.681 ± 3.648e − 3	0.85 ± 2.178e − 3	0.811 ± 2.348e − 3	0.898 ± 1.83e − 3	0.832 ± 3.731e − 3
TSDE patchtsmixer	0.799 ± 7.006e − 4	0.688 ± 2.859e − 3	0.851 ± 1.679e − 3	0.812 ± 2.108e − 3	0.901 ± 2.816e − 3	0.836 ± 4.382e − 3
TSDE timemixer	0.767 ± 1.683e − 2	0.65 ± 2.076e − 2	0.846 ± 3.735e − 3	0.804 ± 4.207e − 3	0.885 ± 1.801e − 3	0.815 ± 1.597e − 3
gemini-2.0-flash CoT +TFM lstm	0.688 ± 5.348e − 3	0.564 ± 5.267e − 3	0.669 ± 1.464e − 2	0.605 ± 1.303e − 2	0.717 ± 1.128e − 2	0.601 ± 1.495e − 2
gemini-2.0-flash CoT +TFM mlp	0.718 ± 1.129e − 3	0.597 ± 2.09e − 3	0.712 ± 1.346e − 2	0.653 ± 1.543e − 2	0.75 ± 1.31e − 2	0.637 ± 1.75e − 2
gemini-2.0-flash CoT +TFM patchtsmixer	0.72 ± 1.202e − 3	0.6 ± 1.908e − 4	0.695 ± 1.63e − 3	0.631 ± 1.855e − 3	0.747 ± 7.968e − 4	0.632 ± 1.418e − 3
gemini-2.0-flash CoT +TFM timemixer	0.714 ± 2.798e − 3	0.597 ± 1.736e − 3	0.688 ± 5.022e − 3	0.629 ± 5.829e − 3	0.723 ± 8.138e − 3	0.605 ± 1.228e − 2
gemini-2.0-flash CoT lstm	0.706 ± 6.648e − 3	0.585 ± 8.732e − 3	0.691 ± 1.918e − 2	0.628 ± 2.157e − 2	0.721 ± 4.388e − 3	0.604 ± 6.022e − 3
gemini-2.0-flash CoT mlp	0.71 ± 4.685e − 3	0.588 ± 5.742e − 3	0.685 ± 8.121e − 3	0.624 ± 8.335e − 3	0.723 ± 5.409e − 3	0.603 ± 6.741e − 3
gemini-2.0-flash CoT patchtsmixer	0.707 ± 1.058e − 3	0.585 ± 1.192e − 3	0.692 ± 7.343e − 3	0.627 ± 7.429e − 3	0.723 ± 1.102e − 2	0.601 ± 1.311e − 2
gemini-2.0-flash CoT timemixer	0.707 ± 2.137e − 3	0.584 ± 1.325e − 3	0.683 ± 6.923e − 3	0.617 ± 7.006e − 3	0.724 ± 8.265e − 3	0.601 ± 9.491e − 3
gemini-2.0-flash ICD +TFM lstm	0.688 ± 8.135e − 3	0.566 ± 1.118e − 2	0.667 ± 1.216e − 2	0.607 ± 1.217e − 2	0.704 ± 1.131e − 2	0.589 ± 1.698e − 2
gemini-2.0-flash ICD +TFM mlp	0.725 ± 1.092e − 2	0.605 ± 1.293e − 2	0.709 ± 3.413e − 3	0.653 ± 2.715e − 3	0.732 ± 1.242e − 2	0.615 ± 1.693e − 2
gemini-2.0-flash ICD +TFM patchtsmixer	0.727 ± 1.79e − 2	0.606 ± 2.454e − 2	0.703 ± 1.196e − 2	0.642 ± 1.193e − 2	0.741 ± 5.058e − 3	0.627 ± 6.805e − 3
gemini-2.0-flash ICD +TFM timemixer	0.714 ± 8.728e − 4	0.596 ± 2.768e − 3	0.696 ± 4.552e − 3	0.641 ± 5.327e − 3	0.711 ± 1.125e − 2	0.596 ± 1.01e − 2
gemini-2.0-flash ICD lstm	0.706 ± 4.078e − 3	0.584 ± 5.169e − 3	0.681 ± 1.47e − 2	0.621 ± 1.588e − 2	0.719 ± 5.841e − 3	0.601 ± 9.043e − 3
gemini-2.0-flash ICD mlp	0.717 ± 8.876e − 3	0.597 ± 1.284e − 2	0.693 ± 2.445e − 3	0.635 ± 2.166e − 3	0.721 ± 6.321e − 3	0.601 ± 8.13e − 3
gemini-2.0-flash ICD patchtsmixer	0.707 ± 1.419e − 3	0.583 ± 1.455e − 3	0.696 ± 1.089e − 2	0.637 ± 1.187e − 2	0.718 ± 8.059e − 3	0.597 ± 9.69e − 3
gemini-2.0-flash ICD timemixer	0.707 ± 5.886e − 4	0.581 ± 1.143e − 3	0.688 ± 7.978e − 3	0.628 ± 8.641e − 3	0.714 ± 4.16e − 3	0.592 ± 4.244e − 3
gemini-2.0-flash Trend +TFM lstm	0.677 ± 4.555e − 3	0.546 ± 8.996e − 3	0.659 ± 7.126e − 3	0.592 ± 7.249e − 3	0.716 ± 5.095e − 3	0.605 ± 6.79e − 3
gemini-2.0-flash Trend +TFM mlp	0.719 ± 1.302e − 3	0.596 ± 1.14e − 3	0.701 ± 1.563e − 3	0.638 ± 1.693e − 3	0.719 ± 1.442e − 3	0.599 ± 1.209e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.707 ± 7.753e − 4	0.581 ± 1.084e − 3	0.692 ± 1.293e − 3	0.625 ± 1.285e − 3	0.721 ± 7.448e − 3	0.603 ± 9.94e − 3
gemini-2.0-flash Trend +TFM timemixer	0.707 ± 6.438e − 4	0.588 ± 3.874e − 4	0.685 ± 1.121e − 2	0.623 ± 1.354e − 2	0.715 ± 1.523e − 3	0.601 ± 3.991e − 3
gemini-2.0-flash Trend lstm	0.698 ± 1.22e − 4	0.574 ± 1.025e − 3	0.671 ± 7.364e − 3	0.6 ± 7.048e − 3	0.712 ± 2.628e − 3	0.595 ± 4.96e − 3
gemini-2.0-flash Trend mlp	0.7 ± 2.748e − 3	0.576 ± 2.57e − 3	0.68 ± 9.396e − 3	0.612 ± 1.043e − 2	0.709 ± 5.213e − 3	0.587 ± 6.987e − 3
gemini-2.0-flash Trend patchtsmixer	0.699 ± 1.321e − 3	0.575 ± 1.197e − 3	0.686 ± 7.111e − 3	0.614 ± 7.194e − 3	0.708 ± 2.187e − 3	0.585 ± 3.249e − 3
gemini-2.0-flash Trend timemixer	0.7 ± 1.712e − 3	0.576 ± 1.076e − 3	0.678 ± 3.487e − 3	0.608 ± 3.882e − 3	0.707 ± 3.588e − 3	0.583 ± 4.458e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.7 ± 1.392e − 2	0.579 ± 2e − 2	0.662 ± 1.173e − 2	0.597 ± 1.356e − 2	0.714 ± 6.285e − 3	0.6 ± 8.373e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.742 ± 1.127e − 2	0.627 ± 1.357e − 2	0.694 ± 2.292e − 2	0.634 ± 2.392e − 2	0.749 ± 1.657e − 3	0.637 ± 2.262e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.748 ± 5.132e − 3	0.635 ± 6.363e − 3	0.702 ± 1.048e − 2	0.64 ± 1.122e − 2	0.742 ± 1.007e − 2	0.628 ± 1.227e − 2
gemini-2.0-flash zero_shot +TFM timemixer	0.73 ± 1.432e − 2	0.611 ± 1.471e − 2	0.675 ± 2.83e − 3	0.616 ± 3.266e − 3	0.718 ± 9.66e − 3	0.601 ± 1.524e − 2
gemini-2.0-flash zero_shot lstm	0.715 ± 9.405e − 3	0.597 ± 1.684e − 2	0.678 ± 7.654e − 3	0.613 ± 8.143e − 3	0.72 ± 7.176e − 3	0.603 ± 1.028e − 2
gemini-2.0-flash zero_shot mlp	0.738 ± 9.791e − 3	0.621 ± 1.109e − 2	0.684 ± 5.914e − 3	0.623 ± 7.038e − 3	0.723 ± 6.817e − 3	0.604 ± 8.556e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.749 ± 2.068e − 3	0.63 ± 2.351e − 3	0.687 ± 2.964e − 3	0.624 ± 2.399e − 3	0.725 ± 1.792e − 3	0.606 ± 2.215e − 3
gemini-2.0-flash zero_shot timemixer	0.733 ± 1.782e − 2	0.612 ± 2.063e − 2	0.673 ± 7.553e − 3	0.608 ± 7.632e − 3	0.721 ± 6.148e − 3	0.601 ± 7.37e − 3
interp lstm	0.664 ± 4.33e − 2	0.528 ± 5.538e − 2	0.586 ± 3.228e − 2	0.52 ± 3.973e − 2	0.658 ± 3.933e − 2	0.526 ± 5.13e − 2
interp mlp	0.785 ± 3.045e − 3	0.684 ± 6.478e − 3	0.748 ± 2.634e − 3	0.695 ± 3.103e − 3	0.821 ± 3.047e − 3	0.729 ± 3.135e − 3
interp patchtsmixer	0.868 ± 4.851e − 4	0.786 ± 1.611e − 3	0.782 ± 1.957e − 3	0.73 ± 2.016e − 3	0.863 ± 2.761e − 3	0.78 ± 2.455e − 3
interp timemixer	0.739 ± 3.808e − 3	0.612 ± 5.098e − 3	0.664 ± 4.762e − 4	0.6 ± 3.306e − 3	0.732 ± 5.991e − 3	0.612 ± 1.174e − 2
mean lstm	0.684 ± 8.276e − 2	0.555 ± 1.094e − 1	0.61 ± 7.946e − 2	0.545 ± 8.924e − 2	0.639 ± 0e0	0.5 ± 0e0
mean mlp	0.837 ± 1.092e − 3	0.751 ± 1.993e − 3	0.731 ± 2.236e − 3	0.684 ± 2.358e − 3	0.854 ± 1.325e − 3	0.763 ± 2.016e − 3
mean patchtsmixer	0.879 ± 2.175e − 3	0.81 ± 4.419e − 3	0.79 ± 6.211e − 3	0.741 ± 5.469e − 3	0.847 ± 1.344e − 3	0.756 ± 4.7e − 3
mean timemixer	0.872 ± 1.497e − 3	0.793 ± 1.085e − 3	0.808 ± 1.141e − 3	0.761 ± 2.156e − 3	0.836 ± 1.65e − 3	0.743 ± 2.735e − 3
No_sum_CoT CoT +TFM lstm	0.723 ± 1.544e − 3	0.619 ± 1.426e − 3	0.75 ± 1.476e − 2	0.695 ± 1.678e − 2	0.777 ± 8.079e − 3	0.688 ± 1.272e − 2
No_sum_CoT CoT +TFM mlp	0.765 ± 1.529e − 3	0.654 ± 2.325e − 3	0.78 ± 8.862e − 4	0.729 ± 2.401e − 4	0.8 ± 1.956e − 2	0.703 ± 2.468e − 2
No_sum_CoT CoT +TFM patchtsmixer	0.762 ± 3.793e − 3	0.649 ± 5.79e − 3	0.774 ± 3.033e − 2	0.711 ± 3.255e − 2	0.814 ± 1.059e − 2	0.716 ± 1.263e − 2
No_sum_CoT CoT +TFM timemixer	0.755 ± 1.472e − 2	0.641 ± 2.036e − 2	0.766 ± 2.158e − 2	0.712 ± 2.287e − 2	0.764 ± 2.202e − 2	0.659 ± 2.382e − 2
No_sum_CoT lstm	0.735 ± 3.67e − 3	0.625 ± 3.715e − 3	0.716 ± 1.338e − 2	0.656 ± 1.607e − 2	0.75 ± 8.467e − 3	0.64 ± 1.349e − 2
No_sum_CoT mlp	0.755 ± 9.858e − 3	0.643 ± 1.27e − 2	0.745 ± 3.568e − 3	0.689 ± 4.358e − 3	0.788 ± 3.896e − 3	0.688 ± 5.356e − 3
No_sum_CoT patchtsmixer	0.755 ± 8.893e − 3	0.64 ± 1.08e − 2	0.746 ± 1.058e − 2	0.686 ± 1.145e − 2	0.782 ± 3.733e − 3	0.673 ± 5.333e − 3
No_sum_CoT timemixer	0.751 ± 2.922e − 3	0.633 ± 3.603e − 3	0.727 ± 1.61e − 2	0.666 ± 1.753e − 2	0.779 ± 1.016e − 2	0.672 ± 1.441e − 2
No_sum_ICD ICD +TFM lstm	0.714 ± 4.662e − 3	0.604 ± 7.298e − 3	0.755 ± 1.394e − 3	0.702 ± 2.384e − 3	0.771 ± 3.647e − 3	0.677 ± 5.379e − 3
No_sum_ICD ICD +TFM mlp	0.771 ± 6.736e − 3	0.661 ± 8.792e − 3	0.787 ± 7.887e − 3	0.737 ± 9.344e − 3	0.797 ± 1.163e − 2	0.698 ± 1.555e − 2
No_sum_ICD ICD +TFM patchtsmixer	0.762 ± 4.834e − 3	0.649 ± 6.052e − 3	0.798 ± 9.244e − 3	0.74 ± 9.79e − 3	0.814 ± 1.707e − 2	0.714 ± 2.154e − 2
No_sum_ICD ICD +TFM timemixer	0.765 ± 6.781e − 3	0.655 ± 9.045e − 3	0.789 ± 1.89e − 3	0.739 ± 3.565e − 3	0.778 ± 3.037e − 3	0.673 ± 2.51e − 3
No_sum_ICD lstm	0.733 ± 1.321e − 3	0.621 ± 2.176e − 3	0.735 ± 7.334e − 3	0.681 ± 9.124e − 3	0.75 ± 3.354e − 3	0.641 ± 4.25e − 3
No_sum_ICD mlp	0.751 ± 5.511e − 3	0.636 ± 6.843e − 3	0.751 ± 1.153e − 2	0.696 ± 1.24e − 2	0.781 ± 5.254e − 3	0.675 ± 6.613e − 3
No_sum_ICD patchtsmixer	0.761 ± 5.675e − 3	0.645 ± 7.142e − 3	0.747 ± 3.78e − 3	0.689 ± 4.212e − 3	0.79 ± 9.069e − 3	0.683 ± 1.25e − 2
No_sum_ICD timemixer	0.751 ± 1.169e − 3	0.633 ± 1.206e − 3	0.742 ± 1.722e − 2	0.685 ± 1.817e − 2	0.775 ± 3.997e − 3	0.666 ± 5.715e − 3
No_sum_Trend Trend +TFM lstm	0.722 ± 4.245e − 3	0.614 ± 5.841e − 3	0.758 ± 2.191e − 2	0.703 ± 2.598e − 2	0.787 ± 2.879e − 2	0.702 ± 4.131e − 2
No_sum_Trend Trend +TFM mlp	0.756 ± 1.644e − 2	0.643 ± 2.253e − 2	0.784 ± 1.677e − 2	0.732 ± 1.867e − 2	0.79 ± 1.126e − 2	0.69 ± 1.437e − 2
No_sum_Trend Trend +TFM patchtsmixer	0.754 ± 4.161e − 3	0.639 ± 5.573e − 3	0.797 ± 1.046e − 2	0.735 ± 1.128e − 2	0.808 ± 1.192e − 2	0.707 ± 1.543e − 2
No_sum_Trend Trend +TFM timemixer	0.75 ± 1.1e − 2	0.639 ± 8.155e − 3	0.761 ± 3.013e − 3	0.706 ± 3.665e − 3	0.762 ± 2.516e − 2	0.656 ± 2.332e − 2

Table 22: In-distribution results - Gender. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	auprc	auroc	auprc	auroc	auprc	auroc
No_sum_Trend lstm	0.74 ± 2.23e − 3	0.63 ± 3.426e − 3	0.722 ± 1.427e − 2	0.663 ± 1.613e − 2	0.778 ± 2.76e − 3	0.681 ± 3.846e − 3
No_sum_Trend mlp	0.755 ± 3.027e − 3	0.641 ± 3.93e − 3	0.753 ± 1.008e − 2	0.698 ± 1.139e − 2	0.787 ± 9.173e − 3	0.686 ± 1.275e − 2
No_sum_Trend patchtsmixer	0.749 ± 4.704e − 3	0.631 ± 5.715e − 3	0.754 ± 6.242e − 3	0.695 ± 7.373e − 3	0.79 ± 1.108e − 2	0.684 ± 1.554e − 2
No_sum_Trend timemixer	0.751 ± 2.802e − 3	0.632 ± 3.398e − 3	0.736 ± 1.755e − 2	0.675 ± 1.889e − 2	0.78 ± 2.308e − 3	0.674 ± 2.988e − 3
medgemma CoT +TFM lstm	0.701 ± 9.768e − 3	0.582 ± 1.444e − 2	0.659 ± 4.775e − 3	0.598 ± 4.392e − 3	0.724 ± 1.238e − 2	0.615 ± 2.059e − 2
medgemma CoT +TFM mlp	0.73 ± 1.005e − 2	0.614 ± 1.175e − 2	0.716 ± 1.392e − 2	0.657 ± 1.404e − 2	0.762 ± 1.148e − 3	0.657 ± 2.568e − 3
medgemma CoT +TFM patchtsmixer	0.74 ± 8.523e − 3	0.626 ± 9.423e − 3	0.714 ± 2.428e − 3	0.651 ± 1.848e − 3	0.741 ± 8.583e − 3	0.627 ± 1.134e − 2
medgemma CoT +TFM timemixer	0.718 ± 9.627e − 3	0.598 ± 9.804e − 3	0.71 ± 8.869e − 3	0.653 ± 9.106e − 3	0.726 ± 3.822e − 3	0.612 ± 3.668e − 3
medgemma CoT lstm	0.716 ± 5.031e − 4	0.603 ± 1.864e − 3	0.694 ± 8.079e − 3	0.632 ± 8.779e − 3	0.734 ± 2.271e − 2	0.628 ± 3.664e − 2
medgemma CoT mlp	0.717 ± 3.543e − 3	0.598 ± 4.867e − 3	0.719 ± 3.884e − 3	0.66 ± 4.348e − 3	0.753 ± 1.132e − 2	0.647 ± 1.391e − 2
medgemma CoT patchtsmixer	0.721 ± 2.206e − 3	0.602 ± 2.562e − 3	0.715 ± 9.317e − 3	0.652 ± 9.335e − 3	0.73 ± 2.966e − 2	0.613 ± 3.604e − 2
medgemma CoT timemixer	0.716 ± 3.997e − 3	0.595 ± 4.219e − 3	0.705 ± 5.001e − 3	0.642 ± 5.563e − 3	0.711 ± 2.667e − 3	0.59 ± 3.162e − 3
medgemma ICD +TFM lstm	0.704 ± 1.716e − 2	0.586 ± 2.449e − 2	0.674 ± 5.636e − 3	0.61 ± 6.298e − 3	0.688 ± 7.053e − 3	0.564 ± 8.507e − 3
medgemma ICD +TFM mlp	0.733 ± 1.209e − 2	0.615 ± 1.689e − 2	0.693 ± 1.062e − 2	0.633 ± 1.192e − 2	0.724 ± 1.222e − 2	0.606 ± 1.465e − 2
medgemma ICD +TFM patchtsmixer	0.736 ± 5.126e − 3	0.615 ± 6.441e − 3	0.691 ± 9.462e − 3	0.628 ± 9.006e − 3	0.721 ± 8.924e − 3	0.604 ± 1.059e − 2
medgemma ICD +TFM timemixer	0.716 ± 9.304e − 3	0.597 ± 5.295e − 3	0.69 ± 2.087e − 3	0.631 ± 2.762e − 3	0.718 ± 6.332e − 3	0.602 ± 4.055e − 3
medgemma ICD lstm	0.711 ± 4.333e − 3	0.588 ± 6.317e − 3	0.673 ± 1.307e − 2	0.61 ± 1.482e − 2	0.704 ± 6.042e − 3	0.584 ± 7.67e − 3
medgemma ICD mlp	0.717 ± 1.656e − 3	0.593 ± 2.615e − 3	0.693 ± 9.13e − 3	0.634 ± 9.341e − 3	0.713 ± 4.461e − 3	0.595 ± 4.885e − 3
medgemma ICD patchtsmixer	0.719 ± 1.953e − 3	0.595 ± 2.597e − 3	0.687 ± 8.396e − 3	0.625 ± 8.509e − 3	0.71 ± 7.246e − 3	0.589 ± 8.342e − 3
medgemma ICD timemixer	0.717 ± 3.023e − 3	0.59 ± 3.521e − 3	0.676 ± 3.96e − 3	0.614 ± 3.539e − 3	0.705 ± 4.061e − 3	0.582 ± 4.864e − 3
medgemma Trend +TFM lstm	0.69 ± 1.662e − 3	0.568 ± 5.877e − 4	0.666 ± 4.095e − 2	0.599 ± 4.256e − 2	0.69 ± 1.934e − 2	0.567 ± 2.605e − 2
medgemma Trend +TFM mlp	0.723 ± 9.772e − 3	0.602 ± 1.185e − 2	0.711 ± 1.28e − 2	0.651 ± 1.555e − 2	0.719 ± 3.211e − 3	0.596 ± 4.299e − 3
medgemma Trend +TFM patchtsmixer	0.709 ± 7.731e − 4	0.585 ± 4.571e − 4	0.715 ± 1.91e − 3	0.649 ± 1.747e − 3	0.715 ± 6.824e − 4	0.592 ± 5.277e − 4
medgemma Trend +TFM timemixer	0.709 ± 7.321e − 4	0.59 ± 3.491e − 3	0.689 ± 3.296e − 3	0.627 ± 3.766e − 3	0.713 ± 3.195e − 4	0.598 ± 1.229e − 3
medgemma Trend lstm	0.7 ± 6.957e − 3	0.576 ± 7.229e − 3	0.686 ± 7.68e − 3	0.619 ± 8.363e − 3	0.701 ± 9.49e − 3	0.574 ± 1.131e − 2
medgemma Trend mlp	0.707 ± 4.45e − 3	0.583 ± 5.281e − 3	0.7 ± 6.592e − 3	0.635 ± 7.411e − 3	0.705 ± 4.399e − 3	0.578 ± 5.072e − 3
medgemma Trend patchtsmixer	0.705 ± 2.01e − 3	0.581 ± 2.071e − 3	0.706 ± 9.576e − 3	0.635 ± 9.128e − 3	0.704 ± 1.071e − 3	0.576 ± 1.102e − 3
medgemma Trend timemixer	0.707 ± 1.524e − 3	0.581 ± 1.171e − 3	0.688 ± 1.065e − 2	0.618 ± 1.111e − 2	0.702 ± 9.072e − 4	0.573 ± 1.085e − 3
medgemma zero_shot +TFM lstm	0.696 ± 2.587e − 2	0.578 ± 3.51e − 2	0.675 ± 1.045e − 2	0.612 ± 1.099e − 2	0.731 ± 1.36e − 2	0.624 ± 2.043e − 2
medgemma zero_shot +TFM mlp	0.737 ± 1.751e − 2	0.62 ± 2.358e − 2	0.716 ± 1.13e − 2	0.658 ± 1.286e − 2	0.752 ± 7.858e − 3	0.642 ± 1.103e − 2
medgemma zero_shot +TFM patchtsmixer	0.732 ± 2.435e − 3	0.612 ± 4.336e − 3	0.716 ± 2.947e − 2	0.654 ± 3.129e − 2	0.741 ± 7.613e − 3	0.627 ± 8.152e − 3
medgemma zero_shot +TFM timemixer	0.728 ± 7.906e − 3	0.611 ± 1.199e − 2	0.692 ± 5.17e − 3	0.633 ± 6.136e − 3	0.722 ± 9.043e − 3	0.609 ± 9.301e − 3
medgemma zero_shot lstm	0.717 ± 4.841e − 3	0.603 ± 9.247e − 3	0.67 ± 9.442e − 3	0.606 ± 1.063e − 2	0.729 ± 1.197e − 2	0.621 ± 2.029e − 2
medgemma zero_shot mlp	0.723 ± 7.89e − 3	0.602 ± 1.147e − 2	0.696 ± 1.046e − 2	0.636 ± 1.195e − 2	0.733 ± 2.28e − 3	0.621 ± 2.99e − 3
medgemma zero_shot patchtsmixer	0.722 ± 7.805e − 3	0.602 ± 1.014e − 2	0.709 ± 8.92e − 3	0.647 ± 9.963e − 3	0.735 ± 9.767e − 3	0.621 ± 1.114e − 2
medgemma zero_shot timemixer	0.723 ± 2.405e − 3	0.599 ± 3.898e − 3	0.687 ± 1.206e − 2	0.622 ± 1.365e − 2	0.736 ± 3.054e − 3	0.621 ± 3.758e − 3
No_sum_zero_shot zero_shot +TFM lstm	0.722 ± 5.1e − 3	0.616 ± 1.094e − 2	0.742 ± 4.867e − 3	0.684 ± 5.837e − 3	0.76 ± 2.029e − 2	0.664 ± 2.82e − 2
No_sum_zero_shot zero_shot +TFM mlp	0.751 ± 9.506e − 3	0.634 ± 1.456e − 2	0.752 ± 9.467e − 3	0.696 ± 1.189e − 2	0.806 ± 6.152e − 3	0.71 ± 8.249e − 3
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.753 ± 5.825e − 3	0.639 ± 7.477e − 3	0.773 ± 1.104e − 2	0.71 ± 1.255e − 2	0.81 ± 1.939e − 2	0.709 ± 2.312e − 2
No_sum_zero_shot zero_shot +TFM timemixer	0.749 ± 5.594e − 3	0.632 ± 7.303e − 3	0.759 ± 1.388e − 2	0.706 ± 1.614e − 2	0.761 ± 1.982e − 2	0.655 ± 2.072e − 2
No_sum_zero_shot lstm	0.735 ± 1.049e − 3	0.625 ± 3.211e − 4	0.727 ± 1.031e − 2	0.669 ± 1.276e − 2	0.734 ± 9.081e − 3	0.618 ± 1.173e − 2
No_sum_zero_shot mlp	0.751 ± 6.221e − 3	0.637 ± 8.498e − 3	0.75 ± 8.165e − 3	0.695 ± 9.902e − 3	0.782 ± 6.794e − 3	0.678 ± 9.684e − 3
No_sum_zero_shot patchtsmixer	0.749 ± 3.79e − 3	0.633 ± 4.857e − 3	0.742 ± 8.508e − 3	0.682 ± 8.934e − 3	0.784 ± 6.863e − 3	0.675 ± 9.66e − 3
No_sum_zero_shot timemixer	0.749 ± 1.463e − 3	0.631 ± 1.507e − 3	0.727 ± 7.659e − 3	0.666 ± 8.418e − 3	0.742 ± 2.761e − 2	0.626 ± 3.416e − 2
right lstm	0.662 ± 3.943e − 2	0.527 ± 5.419e − 2	0.57 ± 0e0	0.5 ± 0e0	0.639 ± 0e0	0.5 ± 5e − 6
right mlp	0.842 ± 1.711e − 3	0.763 ± 4.486e − 3	0.786 ± 8.476e − 4	0.736 ± 1.542e − 3	0.876 ± 1.087e − 3	0.797 ± 1.703e − 3
right patchtsmixer	0.875 ± 8.592e − 4	0.802 ± 1.787e − 3	0.809 ± 2.09e − 3	0.764 ± 2.146e − 3	0.876 ± 2.212e − 3	0.795 ± 3.708e − 3
right timemixer	0.755 ± 8.163e − 3	0.633 ± 1.476e − 2	0.727 ± 2.988e − 3	0.681 ± 1.945e − 3	0.768 ± 3.868e − 3	0.66 ± 4.337e − 3

Table 23: In-distribution results - Gender. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	auprc	auroc	auprc	auroc
Llama-3.1 CoT +TFM lstm	0.664 ± 6.426e − 3	0.528 ± 5.865e − 3	0.673 ± 2.82e − 3	0.539 ± 4.609e − 3
Llama-3.1 CoT +TFM mlp	0.673 ± 3.117e − 3	0.535 ± 3.658e − 3	0.674 ± 2.093e − 3	0.54 ± 4.741e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.68 ± 1.007e − 3	0.548 ± 1.27e − 3	0.678 ± 4.151e − 4	0.546 ± 1.174e − 3
Llama-3.1 CoT +TFM timemixer	0.677 ± 3.143e − 3	0.544 ± 6.236e − 3	0.675 ± 1.276e − 3	0.542 ± 1.627e − 3
Llama-3.1 CoT lstm	0.674 ± 5.913e − 4	0.542 ± 2.421e − 3	0.675 ± 3.617e − 4	0.543 ± 8.516e − 4
Llama-3.1 CoT mlp	0.676 ± 8.942e − 4	0.543 ± 1.65e − 3	0.677 ± 1.266e − 3	0.543 ± 2.263e − 3
Llama-3.1 CoT patchtsmixer	0.677 ± 1.087e − 3	0.544 ± 1.92e − 3	0.676 ± 1.235e − 3	0.545 ± 1.534e − 3
Llama-3.1 CoT timemixer	0.677 ± 1.531e − 3	0.545 ± 6.271e − 4	0.674 ± 1.844e − 3	0.543 ± 1.893e − 3
Llama-3.1 ICD +TFM lstm	0.657 ± 2.57e − 3	0.52 ± 1.982e − 3	0.664 ± 2.207e − 3	0.53 ± 3.553e − 3
Llama-3.1 ICD +TFM mlp	0.664 ± 3.384e − 3	0.526 ± 3.965e − 3	0.667 ± 1.799e − 3	0.535 ± 3.067e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.672 ± 6.484e − 4	0.535 ± 1.28e − 3	0.671 ± 7.286e − 4	0.54 ± 1.842e − 3
Llama-3.1 ICD +TFM timemixer	0.669 ± 3.524e − 3	0.533 ± 3.805e − 3	0.67 ± 7.071e − 4	0.538 ± 3.141e − 3
Llama-3.1 ICD lstm	0.667 ± 1.733e − 3	0.531 ± 1.912e − 3	0.665 ± 1.754e − 3	0.534 ± 1.14e − 3
Llama-3.1 ICD mlp	0.667 ± 1.79e − 3	0.532 ± 2.489e − 3	0.668 ± 1.333e − 3	0.536 ± 1.124e − 4
Llama-3.1 ICD patchtsmixer	0.668 ± 2.909e − 4	0.533 ± 1.122e − 3	0.667 ± 1.141e − 3	0.536 ± 6.438e − 4
Llama-3.1 ICD timemixer	0.669 ± 7.352e − 4	0.535 ± 7.313e − 4	0.667 ± 1.223e − 3	0.536 ± 4.909e − 4
Llama-3.1 Trend +TFM lstm	0.662 ± 5.545e − 3	0.527 ± 4.806e − 3	0.663 ± 3.815e − 3	0.528 ± 3.898e − 3
Llama-3.1 Trend +TFM mlp	0.668 ± 2.891e − 3	0.533 ± 6.51e − 3	0.667 ± 4.11e − 3	0.53 ± 3.45e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.677 ± 1.747e − 3	0.541 ± 3.033e − 3	0.672 ± 1.755e − 3	0.537 ± 2.944e − 3
Llama-3.1 Trend +TFM timemixer	0.674 ± 1.827e − 3	0.542 ± 3.232e − 3	0.672 ± 1.816e − 3	0.539 ± 1.319e − 3
Llama-3.1 Trend lstm	0.669 ± 9.636e − 4	0.534 ± 2.643e − 3	0.667 ± 8.415e − 4	0.534 ± 9.089e − 4
Llama-3.1 Trend mlp	0.67 ± 2.436e − 3	0.537 ± 4.101e − 3	0.667 ± 3.737e − 4	0.534 ± 8.304e − 4
Llama-3.1 Trend patchtsmixer	0.672 ± 7.463e − 4	0.537 ± 8.551e − 4	0.667 ± 8.675e − 4	0.533 ± 1.001e − 3
Llama-3.1 Trend timemixer	0.673 ± 8.969e − 4	0.54 ± 6.96e − 4	0.667 ± 9.77e − 4	0.533 ± 1.519e − 3
Llama-3.1 zero_shot +TFM lstm	0.656 ± 2.526e − 3	0.518 ± 3.192e − 3	0.671 ± 3.138e − 3	0.539 ± 1.484e − 3
Llama-3.1 zero_shot +TFM mlp	0.662 ± 2.076e − 3	0.521 ± 2.602e − 3	0.672 ± 1.873e − 3	0.539 ± 2.119e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.664 ± 1.153e − 3	0.527 ± 1.649e − 3	0.676 ± 1.949e − 3	0.543 ± 2.982e − 3
Llama-3.1 zero_shot +TFM timemixer	0.667 ± 3.793e − 3	0.534 ± 4.487e − 3	0.673 ± 9.395e − 4	0.542 ± 2.297e − 3
Llama-3.1 zero_shot lstm	0.661 ± 1.774e − 3	0.522 ± 1.272e − 3	0.672 ± 2.568e − 3	0.539 ± 2.97e − 3
Llama-3.1 zero_shot mlp	0.662 ± 2.248e − 3	0.523 ± 2.711e − 3	0.675 ± 1.282e − 3	0.545 ± 2.477e − 3
Llama-3.1 zero_shot patchtsmixer	0.662 ± 1.64e − 4	0.524 ± 7.276e − 4	0.676 ± 8.784e − 4	0.546 ± 7.578e − 4
Llama-3.1 zero_shot timemixer	0.662 ± 1.446e − 3	0.525 ± 1.079e − 3	0.672 ± 6.774e − 4	0.541 ± 1.045e − 3
TFM lstm	0.647 ± 1.332e − 3	0.514 ± 4.081e − 3	0.647 ± 3.477e − 3	0.513 ± 4.214e − 3
TFM mlp	0.656 ± 4.364e − 3	0.521 ± 3.65e − 3	0.652 ± 2.906e − 3	0.521 ± 1.976e − 3
TFM patchtsmixer	0.661 ± 1.895e − 3	0.527 ± 1.407e − 3	0.651 ± 1.066e − 3	0.522 ± 1.018e − 3
TFM timemixer	0.664 ± 2.007e − 3	0.529 ± 2.858e − 3	0.658 ± 1.02e − 3	0.526 ± 1.029e − 3
TSDE lstm	0.638 ± 9.617e − 4	0.496 ± 4.173e − 3	0.64 ± 6.798e − 3	0.502 ± 9.967e − 3
TSDE mlp	0.653 ± 2.68e − 2	0.516 ± 3.931e − 2	0.691 ± 1.726e − 2	0.571 ± 2.475e − 2
TSDE patchtsmixer	0.65 ± 1.516e − 2	0.521 ± 2.29e − 2	0.695 ± 1.768e − 2	0.574 ± 1.8e − 2
TSDE timemixer	0.64 ± 2.737e − 3	0.498 ± 3.735e − 3	0.66 ± 3.474e − 3	0.528 ± 4.731e − 3
gemini-2.0-flash CoT +TFM lstm	0.667 ± 3.387e − 3	0.534 ± 6.672e − 3	0.684 ± 2.266e − 3	0.553 ± 2.386e − 3
gemini-2.0-flash CoT +TFM mlp	0.681 ± 3.592e − 3	0.549 ± 4.444e − 3	0.688 ± 2.068e − 3	0.559 ± 3.566e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.687 ± 1.273e − 3	0.558 ± 1.278e − 3	0.695 ± 2.92e − 3	0.567 ± 2.923e − 3
gemini-2.0-flash CoT +TFM timemixer	0.681 ± 2.181e − 3	0.555 ± 1.612e − 3	0.694 ± 1.233e − 3	0.565 ± 1.432e − 3
gemini-2.0-flash CoT lstm	0.68 ± 6.536e − 3	0.552 ± 9.249e − 3	0.685 ± 3.19e − 3	0.558 ± 3.331e − 3
gemini-2.0-flash CoT mlp	0.683 ± 4.006e − 3	0.556 ± 4.798e − 3	0.688 ± 1.267e − 3	0.561 ± 1.209e − 3
gemini-2.0-flash CoT patchtsmixer	0.681 ± 1.202e − 3	0.553 ± 1.498e − 3	0.69 ± 1.952e − 3	0.563 ± 1.933e − 3
gemini-2.0-flash CoT timemixer	0.679 ± 1.521e − 3	0.552 ± 1.252e − 3	0.687 ± 2.086e − 3	0.561 ± 1.86e − 3
gemini-2.0-flash ICD +TFM lstm	0.671 ± 6.3e − 3	0.54 ± 1.002e − 2	0.688 ± 1.499e − 3	0.56 ± 3.216e − 3
gemini-2.0-flash ICD +TFM mlp	0.681 ± 1.129e − 3	0.549 ± 4.585e − 3	0.691 ± 1.102e − 3	0.564 ± 3.598e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.688 ± 2.907e − 3	0.56 ± 1.108e − 3	0.692 ± 2.075e − 3	0.569 ± 2.007e − 3
gemini-2.0-flash ICD +TFM timemixer	0.686 ± 2.3e − 3	0.561 ± 1.33e − 3	0.69 ± 3.373e − 3	0.565 ± 2.886e − 3
gemini-2.0-flash ICD lstm	0.679 ± 3.447e − 3	0.552 ± 5.017e − 3	0.689 ± 1.697e − 3	0.568 ± 1.536e − 3
gemini-2.0-flash ICD mlp	0.686 ± 1.743e − 3	0.559 ± 1.456e − 3	0.688 ± 6.494e − 4	0.567 ± 7.23e − 4
gemini-2.0-flash ICD patchtsmixer	0.683 ± 6.85e − 4	0.556 ± 6.835e − 4	0.688 ± 2.879e − 3	0.568 ± 3.261e − 3
gemini-2.0-flash ICD timemixer	0.678 ± 1.054e − 3	0.552 ± 1.234e − 3	0.685 ± 8.961e − 4	0.565 ± 1.171e − 3
gemini-2.0-flash Trend +TFM lstm	0.664 ± 3.821e − 3	0.529 ± 4.386e − 3	0.677 ± 7.19e − 4	0.543 ± 5.579e − 4
gemini-2.0-flash Trend +TFM mlp	0.678 ± 6.497e − 3	0.547 ± 6.906e − 3	0.682 ± 4.529e − 3	0.551 ± 5.453e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.681 ± 2.555e − 3	0.552 ± 3.869e − 3	0.685 ± 2.19e − 3	0.554 ± 3.018e − 3
gemini-2.0-flash Trend +TFM timemixer	0.682 ± 6.809e − 4	0.558 ± 2.641e − 4	0.685 ± 5.95e − 4	0.554 ± 5.81e − 4
gemini-2.0-flash Trend lstm	0.677 ± 1.297e − 3	0.547 ± 2.634e − 3	0.682 ± 2.711e − 3	0.554 ± 2.045e − 3
gemini-2.0-flash Trend mlp	0.681 ± 1.482e − 3	0.551 ± 2.168e − 3	0.681 ± 1.688e − 3	0.553 ± 1.432e − 3
gemini-2.0-flash Trend patchtsmixer	0.68 ± 1.276e − 3	0.551 ± 1.95e − 3	0.681 ± 9.38e − 4	0.554 ± 1.776e − 3
gemini-2.0-flash Trend timemixer	0.68 ± 1.174e − 3	0.551 ± 7.688e − 4	0.682 ± 6.751e − 4	0.555 ± 1.328e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.673 ± 3.99e − 3	0.541 ± 5.391e − 3	0.684 ± 2.167e − 3	0.556 ± 1.379e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.688 ± 4.124e − 3	0.558 ± 5.348e − 3	0.691 ± 1.859e − 3	0.564 ± 3.858e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.691 ± 1.151e − 3	0.559 ± 6.519e − 4	0.693 ± 2.076e − 3	0.566 ± 2.471e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.689 ± 2.983e − 3	0.56 ± 1.665e − 3	0.692 ± 1.139e − 3	0.566 ± 1.297e − 3
gemini-2.0-flash zero_shot lstm	0.682 ± 4.034e − 3	0.553 ± 3.975e − 3	0.692 ± 7.208e − 4	0.565 ± 1.489e − 3
gemini-2.0-flash zero_shot mlp	0.686 ± 2.353e − 3	0.558 ± 2.876e − 3	0.691 ± 1.573e − 3	0.565 ± 1.388e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.685 ± 2.944e − 3	0.557 ± 3.471e − 3	0.693 ± 1.073e − 3	0.566 ± 1.022e − 3
gemini-2.0-flash zero_shot timemixer	0.687 ± 5.444e − 3	0.558 ± 5.147e − 3	0.692 ± 1.375e − 3	0.566 ± 1.655e − 3
interp lstm	0.644 ± 1.164e − 2	0.508 ± 1.574e − 2	0.641 ± 3.775e − 3	0.504 ± 7.525e − 3
interp mlp	0.657 ± 1.27e − 3	0.52 ± 9.396e − 4	0.637 ± 4.554e − 4	0.499 ± 6.287e − 4
interp patchtsmixer	0.681 ± 3.759e − 3	0.546 ± 5.672e − 3	0.655 ± 2.889e − 3	0.534 ± 4.291e − 3
interp timemixer	0.659 ± 4.302e − 3	0.524 ± 7.454e − 3	0.662 ± 7.809e − 4	0.526 ± 1.762e − 3
mean lstm	0.638 ± 1.495e − 3	0.499 ± 2.835e − 3	0.643 ± 9.4e − 3	0.507 ± 1.477e − 2
mean mlp	0.66 ± 3.843e − 3	0.528 ± 3.351e − 3	0.667 ± 5.15e − 3	0.549 ± 7.58e − 3
mean patchtsmixer	0.694 ± 6.732e − 3	0.567 ± 3.063e − 3	0.663 ± 1.742e − 2	0.542 ± 3.163e − 2
mean timemixer	0.622 ± 2.893e − 3	0.468 ± 1.83e − 3	0.644 ± 1.102e − 2	0.497 ± 1.143e − 2
No_sum_CoT CoT +TFM lstm	0.689 ± 1.166e − 3	0.56 ± 2.305e − 3	0.711 ± 1.29e − 3	0.583 ± 2.859e − 3
No_sum_CoT CoT +TFM mlp	0.701 ± 1.357e − 3	0.568 ± 1.455e − 3	0.701 ± 1.873e − 3	0.574 ± 3.119e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.701 ± 9.133e − 4	0.568 ± 8.514e − 4	0.711 ± 5.239e − 3	0.586 ± 6.898e − 3
No_sum_CoT CoT +TFM timemixer	0.702 ± 1.437e − 3	0.572 ± 3.253e − 3	0.709 ± 3.145e − 3	0.582 ± 2.542e − 3
No_sum_CoT lstm	0.691 ± 1.583e − 3	0.565 ± 1.564e − 3	0.708 ± 2.701e − 3	0.586 ± 3.573e − 3
No_sum_CoT mlp	0.694 ± 3.843e − 3	0.572 ± 6.092e − 3	0.697 ± 1.963e − 3	0.576 ± 1.669e − 3
No_sum_CoT patchtsmixer	0.695 ± 2.291e − 3	0.571 ± 3.742e − 3	0.705 ± 1.04e − 3	0.582 ± 2.895e − 3
No_sum_CoT timemixer	0.696 ± 1.185e − 3	0.57 ± 1.322e − 3	0.697 ± 2.566e − 3	0.577 ± 3.58e − 3
No_sum_ICD ICD +TFM lstm	0.687 ± 7.433e − 4	0.559 ± 9.199e − 4	0.711 ± 2.782e − 3	0.587 ± 3.324e − 3
No_sum_ICD ICD +TFM mlp	0.703 ± 5.818e − 3	0.572 ± 7.284e − 3	0.712 ± 5.727e − 3	0.586 ± 6.684e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.703 ± 1.164e − 3	0.573 ± 2.192e − 3	0.714 ± 2.505e − 3	0.589 ± 1.042e − 3
No_sum_ICD ICD +TFM timemixer	0.702 ± 1.975e − 3	0.57 ± 4.537e − 4	0.709 ± 7.882e − 3	0.583 ± 8.063e − 3
No_sum_ICD lstm	0.695 ± 1.836e − 3	0.573 ± 2.081e − 3	0.707 ± 1.467e − 3	0.584 ± 1.121e − 3
No_sum_ICD mlp	0.697 ± 2.346e − 3	0.576 ± 3.113e − 3	0.706 ± 6.114e − 3	0.585 ± 6.285e − 3
No_sum_ICD patchtsmixer	0.699 ± 2.085e − 3	0.578 ± 1.83e − 3	0.711 ± 1.838e − 3	0.589 ± 1.995e − 3
No_sum_ICD timemixer	0.703 ± 1.428e − 3	0.58 ± 1.191e − 3	0.705 ± 2.969e − 3	0.584 ± 2.066e − 3
No_sum_Trend Trend +TFM lstm	0.687 ± 2.651e − 3	0.556 ± 3.955e − 3	0.715 ± 3.465e − 3	0.589 ± 4.91e − 3
No_sum_Trend Trend +TFM mlp	0.698 ± 5.702e − 3	0.568 ± 5.788e − 3	0.699 ± 8.263e − 3	0.572 ± 1.13e − 2
No_sum_Trend Trend +TFM patchtsmixer	0.7 ± 1.456e − 3	0.568 ± 3.095e − 3	0.708 ± 4.263e − 3	0.586 ± 5.725e − 3
No_sum_Trend Trend +TFM timemixer	0.701 ± 1.005e − 3	0.569 ± 3.575e − 3	0.714 ± 1.818e − 3	0.588 ± 1.287e − 3

Table 24: Cross-site transfer results - Gender. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	auprc	auroc	auprc	auroc
No_sum_Trend lstm	0.686 ± 2.715e − 3	0.561 ± 3.173e − 3	0.712 ± 2.536e − 3	0.588 ± 2.901e − 3
No_sum_Trend mlp	0.687 ± 1.638e − 3	0.563 ± 1.828e − 3	0.707 ± 3.987e − 3	0.588 ± 3.721e − 3
No_sum_Trend patchtsmixer	0.691 ± 1.504e − 3	0.566 ± 1.569e − 3	0.702 ± 2.505e − 3	0.584 ± 2.975e − 3
No_sum_Trend timemixer	0.694 ± 1.373e − 3	0.568 ± 1.335e − 3	0.702 ± 1.888e − 3	0.582 ± 3.587e − 3
medgemma CoT +TFM lstm	0.674 ± 3.333e − 3	0.542 ± 4.159e − 3	0.68 ± 9.923e − 4	0.55 ± 1.624e − 3
medgemma CoT +TFM mlp	0.687 ± 2.608e − 4	0.553 ± 5.26e − 4	0.686 ± 1.44e − 3	0.558 ± 2.216e − 3
medgemma CoT +TFM patchtsmixer	0.69 ± 8.329e − 4	0.556 ± 8.536e − 4	0.687 ± 1.499e − 3	0.56 ± 1.612e − 3
medgemma CoT +TFM timemixer	0.685 ± 6.618e − 3	0.555 ± 5.886e − 3	0.685 ± 8.743e − 4	0.559 ± 1.573e − 3
medgemma CoT lstm	0.683 ± 6.904e − 4	0.551 ± 1.304e − 3	0.683 ± 1.401e − 3	0.558 ± 1.623e − 3
medgemma CoT mlp	0.685 ± 1.403e − 3	0.555 ± 2.095e − 3	0.685 ± 9.946e − 5	0.561 ± 6.614e − 4
medgemma CoT patchtsmixer	0.687 ± 7.489e − 4	0.558 ± 8.431e − 4	0.686 ± 6.106e − 4	0.561 ± 3.438e − 4
medgemma CoT timemixer	0.686 ± 1.536e − 3	0.556 ± 2.005e − 3	0.681 ± 1.419e − 3	0.556 ± 1.152e − 3
medgemma ICD +TFM lstm	0.67 ± 2.968e − 3	0.539 ± 6.595e − 3	0.675 ± 1.207e − 3	0.543 ± 1.559e − 3
medgemma ICD +TFM mlp	0.677 ± 1.649e − 3	0.548 ± 2.954e − 3	0.676 ± 3.333e − 3	0.547 ± 3.011e − 3
medgemma ICD +TFM patchtsmixer	0.682 ± 9.678e − 4	0.551 ± 1.749e − 3	0.679 ± 7.088e − 4	0.55 ± 1.454e − 3
medgemma ICD +TFM timemixer	0.681 ± 7.295e − 4	0.552 ± 1.295e − 3	0.68 ± 1.527e − 3	0.55 ± 9.28e − 4
medgemma ICD lstm	0.673 ± 2.193e − 3	0.546 ± 1.734e − 3	0.676 ± 1.544e − 3	0.55 ± 9.854e − 4
medgemma ICD mlp	0.676 ± 9.677e − 4	0.551 ± 6.531e − 4	0.673 ± 1.987e − 3	0.546 ± 2.311e − 3
medgemma ICD patchtsmixer	0.679 ± 2.547e − 4	0.554 ± 2.541e − 4	0.674 ± 1.957e − 3	0.548 ± 2.653e − 3
medgemma ICD timemixer	0.68 ± 5.372e − 4	0.554 ± 1.056e − 3	0.675 ± 6.08e − 4	0.549 ± 6.74e − 4
medgemma Trend +TFM lstm	0.67 ± 5.701e − 4	0.538 ± 3.443e − 4	0.676 ± 3.496e − 3	0.544 ± 4.863e − 3
medgemma Trend +TFM mlp	0.682 ± 3.816e − 3	0.546 ± 4.199e − 3	0.683 ± 3.324e − 3	0.554 ± 5.511e − 3
medgemma Trend +TFM patchtsmixer	0.685 ± 1.427e − 3	0.552 ± 1.289e − 3	0.683 ± 2.212e − 3	0.554 ± 1.831e − 3
medgemma Trend +TFM timemixer	0.683 ± 3.498e − 3	0.553 ± 4.794e − 3	0.684 ± 1.091e − 3	0.555 ± 1.287e − 3
medgemma Trend lstm	0.68 ± 3.709e − 3	0.547 ± 4.37e − 3	0.683 ± 2.223e − 3	0.557 ± 1.936e − 3
medgemma Trend mlp	0.683 ± 2.096e − 3	0.551 ± 2.115e − 3	0.685 ± 8.504e − 4	0.56 ± 1.115e − 3
medgemma Trend patchtsmixer	0.683 ± 1.228e − 3	0.551 ± 8.638e − 4	0.685 ± 9.167e − 4	0.56 ± 1.063e − 3
medgemma Trend timemixer	0.684 ± 1.472e − 3	0.553 ± 6.251e − 4	0.68 ± 1.174e − 3	0.557 ± 1.081e − 3
medgemma zero_shot +TFM lstm	0.671 ± 6.166e − 3	0.538 ± 5.835e − 3	0.681 ± 2.502e − 3	0.552 ± 2.85e − 3
medgemma zero_shot +TFM mlp	0.679 ± 6.633e − 3	0.548 ± 9.38e − 3	0.683 ± 3.667e − 3	0.555 ± 3.361e − 3
medgemma zero_shot +TFM patchtsmixer	0.684 ± 2.966e − 3	0.553 ± 4.421e − 3	0.688 ± 4.088e − 3	0.56 ± 5.639e − 3
medgemma zero_shot +TFM timemixer	0.681 ± 1.818e − 3	0.55 ± 2.888e − 3	0.684 ± 7.047e − 4	0.558 ± 9.539e − 4
medgemma zero_shot lstm	0.68 ± 1.026e − 3	0.553 ± 1.347e − 3	0.682 ± 8.385e − 4	0.558 ± 1.123e − 3
medgemma zero_shot mlp	0.68 ± 1.632e − 3	0.555 ± 2.142e − 3	0.685 ± 1.068e − 3	0.56 ± 2.674e − 3
medgemma zero_shot patchtsmixer	0.681 ± 1.601e − 3	0.557 ± 1.468e − 3	0.683 ± 1.665e − 3	0.557 ± 2.348e − 3
medgemma zero_shot timemixer	0.682 ± 7.89e − 4	0.556 ± 5.766e − 4	0.681 ± 1.748e − 3	0.556 ± 3.155e − 3
No_sum_zero_shot zero_shot +TFM lstm	0.687 ± 2.507e − 3	0.555 ± 2.461e − 3	0.709 ± 1.499e − 3	0.582 ± 2.53e − 3
No_sum_zero_shot zero_shot +TFM mlp	0.699 ± 1.001e − 3	0.567 ± 4.949e − 4	0.701 ± 6.731e − 3	0.573 ± 8.181e − 3
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.7 ± 5.422e − 3	0.567 ± 6.611e − 3	0.709 ± 7.127e − 4	0.58 ± 1.602e − 3
No_sum_zero_shot zero_shot +TFM timemixer	0.698 ± 1.78e − 3	0.57 ± 2.616e − 3	0.709 ± 8.243e − 3	0.584 ± 6.554e − 3
No_sum_zero_shot lstm	0.689 ± 8.802e − 4	0.562 ± 8.527e − 4	0.699 ± 2.181e − 3	0.573 ± 3.481e − 3
No_sum_zero_shot mlp	0.688 ± 1.266e − 3	0.564 ± 2.873e − 3	0.701 ± 1.758e − 3	0.579 ± 2.061e − 3
No_sum_zero_shot patchtsmixer	0.689 ± 1.275e − 3	0.564 ± 1.727e − 3	0.698 ± 1.747e − 3	0.575 ± 1.635e − 3
No_sum_zero_shot timemixer	0.688 ± 1.633e − 3	0.563 ± 1.083e − 3	0.691 ± 2.098e − 3	0.568 ± 2.346e − 3
right lstm	0.642 ± 6.967e − 3	0.506 ± 1.217e − 2	0.639 ± 0e0	0.5 ± 0e0
right mlp	0.65 ± 1.583e − 3	0.511 ± 2.671e − 3	0.64 ± 1.61e − 3	0.504 ± 2.429e − 3
right patchtsmixer	0.682 ± 2.058e − 3	0.55 ± 1.6e − 3	0.649 ± 5.373e − 3	0.519 ± 2.978e − 3
right timemixer	0.662 ± 6.162e − 3	0.528 ± 6.041e − 3	0.67 ± 2.405e − 3	0.534 ± 7.208e − 3

Table 25: Cross-site transfer results - Gender. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro	f1_micro	recall_micro
Llama-3.1 CoT +TFM lstm	0.759 ± 1.504e − 3	0.848 ± 5.911e − 3	0.76 ± 4.839e − 3	0.836 ± 2.494e − 3	0.808 ± 2.843e − 3	0.917 ± 5.773e − 3
Llama-3.1 CoT +TFM mlp	0.771 ± 2.913e − 3	0.847 ± 3.027e − 3	0.77 ± 7.323e − 3	0.831 ± 4.135e − 3	0.814 ± 3.384e − 3	0.915 ± 2.589e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.736 ± 1.973e − 2	0.859 ± 1.498e − 2	0.753 ± 7.457e − 3	0.82 ± 6.964e − 3	0.798 ± 8.902e − 3	0.921 ± 4.925e − 3
Llama-3.1 CoT +TFM timemixer	0.747 ± 6.439e − 3	0.854 ± 1.833e − 3	0.737 ± 1.381e − 2	0.811 ± 1.056e − 2	0.805 ± 8.986e − 3	0.925 ± 2.113e − 3
Llama-3.1 CoT lstm	0.733 ± 9.055e − 3	0.855 ± 3.436e − 3	0.692 ± 6.113e − 3	0.815 ± 4.863e − 3	0.793 ± 3.488e − 3	0.935 ± 4.021e − 3
Llama-3.1 CoT mlp	0.738 ± 8.59e − 3	0.854 ± 2.156e − 3	0.69 ± 4.948e − 3	0.811 ± 4.759e − 3	0.793 ± 2.72e − 3	0.935 ± 7.473e − 3
Llama-3.1 CoT patchtsmixer	0.725 ± 6.203e − 3	0.855 ± 1.603e − 2	0.683 ± 3.666e − 3	0.813 ± 2.29e − 3	0.787 ± 2.354e − 3	0.937 ± 5.183e − 3
Llama-3.1 CoT timemixer	0.728 ± 4.499e − 3	0.855 ± 1.809e − 2	0.684 ± 3.584e − 3	0.813 ± 3.468e − 3	0.79 ± 2.229e − 3	0.939 ± 2.984e − 3
Llama-3.1 ICD +TFM lstm	0.782 ± 2.866e − 3	0.86 ± 4.17e − 3	0.784 ± 6.085e − 3	0.858 ± 7.348e − 3	0.82 ± 3.275e − 3	0.919 ± 2.594e − 3
Llama-3.1 ICD +TFM mlp	0.791 ± 6.305e − 3	0.86 ± 3.877e − 3	0.786 ± 7.182e − 3	0.845 ± 6.57e − 3	0.826 ± 5.987e − 3	0.91 ± 3.833e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.77 ± 1.35e − 2	0.854 ± 1.041e − 2	0.767 ± 4.433e − 3	0.84 ± 4.443e − 3	0.805 ± 9.498e − 3	0.917 ± 5.297e − 3
Llama-3.1 ICD +TFM timemixer	0.759 ± 9.084e − 3	0.856 ± 3.575e − 3	0.768 ± 5.454e − 3	0.846 ± 6.041e − 3	0.816 ± 2.132e − 3	0.922 ± 4.818e − 3
Llama-3.1 ICD lstm	0.752 ± 8.791e − 3	0.857 ± 5.166e − 3	0.728 ± 4.776e − 3	0.834 ± 6.282e − 3	0.804 ± 2.747e − 3	0.932 ± 7.23e − 3
Llama-3.1 ICD mlp	0.756 ± 7.923e − 3	0.86 ± 2.418e − 3	0.733 ± 5.609e − 3	0.827 ± 2.671e − 3	0.803 ± 4.286e − 3	0.93 ± 6.317e − 3
Llama-3.1 ICD patchtsmixer	0.744 ± 9.307e − 3	0.853 ± 4.546e − 3	0.719 ± 8.979e − 3	0.821 ± 2.673e − 3	0.795 ± 2.433e − 3	0.934 ± 6.717e − 3
Llama-3.1 ICD timemixer	0.748 ± 6.901e − 3	0.864 ± 3.019e − 3	0.723 ± 7.084e − 3	0.826 ± 1.007e − 3	0.803 ± 3.418e − 3	0.935 ± 8.746e − 3
Llama-3.1 Trend +TFM lstm	0.809 ± 5.031e − 3	0.861 ± 1.306e − 3	0.792 ± 5.698e − 3	0.85 ± 4.228e − 3	0.853 ± 3.27e − 3	0.905 ± 1.584e − 3
Llama-3.1 Trend +TFM mlp	0.817 ± 2.262e − 3	0.858 ± 5.886e − 3	0.797 ± 8.779e − 3	0.833 ± 4.905e − 3	0.856 ± 6.346e − 3	0.906 ± 5.637e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.789 ± 1.251e − 2	0.847 ± 1.06e − 2	0.781 ± 6.412e − 3	0.828 ± 5.052e − 3	0.831 ± 1.194e − 2	0.896 ± 6.258e − 3
Llama-3.1 Trend +TFM timemixer	0.788 ± 8.871e − 3	0.852 ± 6.465e − 4	0.776 ± 9.525e − 4	0.832 ± 8.903e − 4	0.827 ± 1.403e − 2	0.898 ± 5.407e − 3
Llama-3.1 Trend lstm	0.777 ± 1.378e − 2	0.837 ± 1.26e − 2	0.744 ± 5.898e − 3	0.816 ± 4.685e − 3	0.835 ± 6.602e − 3	0.901 ± 4.228e − 3
Llama-3.1 Trend mlp	0.781 ± 1.431e − 2	0.847 ± 6.816e − 3	0.747 ± 7.288e − 3	0.815 ± 1.982e − 3	0.836 ± 7.237e − 3	0.903 ± 2.543e − 3
Llama-3.1 Trend patchtsmixer	0.766 ± 1.143e − 2	0.845 ± 3.502e − 3	0.733 ± 6.448e − 3	0.807 ± 2.011e − 3	0.824 ± 5.114e − 3	0.899 ± 3.857e − 3
Llama-3.1 Trend timemixer	0.771 ± 1.132e − 2	0.844 ± 3.25e − 3	0.738 ± 8.357e − 3	0.814 ± 2.792e − 3	0.828 ± 5.605e − 3	0.898 ± 4.965e − 3
Llama-3.1 zero_shot +TFM lstm	0.809 ± 2.576e − 3	0.872 ± 7.616e − 3	0.778 ± 3.649e − 3	0.856 ± 4.224e − 3	0.822 ± 5.651e − 3	0.917 ± 1.668e − 3
Llama-3.1 zero_shot +TFM mlp	0.819 ± 2.948e − 3	0.872 ± 5.474e − 3	0.782 ± 5.126e − 3	0.844 ± 3.205e − 3	0.824 ± 3.438e − 3	0.914 ± 1.721e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.791 ± 5.223e − 3	0.865 ± 3.37e − 3	0.76 ± 5.657e − 3	0.836 ± 3.153e − 3	0.804 ± 3.756e − 3	0.915 ± 4.702e − 3
Llama-3.1 zero_shot +TFM timemixer	0.793 ± 8.257e − 3	0.865 ± 4.343e − 3	0.77 ± 5.763e − 3	0.845 ± 9.868e − 4	0.815 ± 3.645e − 3	0.923 ± 4.415e − 3
Llama-3.1 zero_shot lstm	0.778 ± 9.139e − 3	0.858 ± 7.707e − 3	0.72 ± 9.989e − 3	0.825 ± 5.409e − 3	0.805 ± 1.598e − 3	0.931 ± 2.449e − 3
Llama-3.1 zero_shot mlp	0.785 ± 1.097e − 2	0.862 ± 7.484e − 3	0.72 ± 8.082e − 3	0.82 ± 4.275e − 3	0.804 ± 3.139e − 3	0.93 ± 4.183e − 3
Llama-3.1 zero_shot patchtsmixer	0.772 ± 1.008e − 2	0.861 ± 5.025e − 3	0.71 ± 6.547e − 3	0.824 ± 1.434e − 3	0.799 ± 1.577e − 3	0.934 ± 8.589e − 4
Llama-3.1 zero_shot timemixer	0.773 ± 1.038e − 2	0.853 ± 1.372e − 2	0.708 ± 5.895e − 3	0.823 ± 5.838e − 3	0.8 ± 4.755e − 4	0.934 ± 1.143e − 3
TFM lstm	0.741 ± 9.034e − 4	0.848 ± 4.087e − 3	0.722 ± 1.278e − 3	0.817 ± 5.07e − 3	0.801 ± 1.61e − 3	0.924 ± 2.687e − 3
TFM mlp	0.742 ± 4.86e − 4	0.85 ± 7.832e − 3	0.726 ± 1.487e − 3	0.813 ± 3.211e − 3	0.801 ± 1.695e − 3	0.922 ± 2.964e − 3
TFM patchtsmixer	0.721 ± 7.536e − 3	0.85 ± 1.464e − 2	0.713 ± 1.868e − 3	0.806 ± 2.089e − 3	0.794 ± 1.514e − 3	0.923 ± 4.105e − 3
TFM timemixer	0.728 ± 3.021e − 3	0.844 ± 3.174e − 3	0.719 ± 1.011e − 3	0.805 ± 8.749e − 3	0.797 ± 1.061e − 3	0.927 ± 6.955e − 3
TSDE mlp	0.882 ± 5.15e − 3	0.9 ± 6.275e − 3	0.858 ± 3.274e − 3	0.889 ± 2.551e − 3	0.914 ± 3.144e − 3	0.897 ± 4.171e − 3
TSDE lstm	0.885 ± 4.684e − 3	0.902 ± 5.807e − 3	0.862 ± 2.948e − 3	0.89 ± 2.344e − 3	0.915 ± 2.896e − 3	0.899 ± 3.746e − 3
TSDE patchtsmixer	0.882 ± 2.559e − 3	0.901 ± 3.322e − 3	0.864 ± 1.931e − 3	0.888 ± 1.861e − 3	0.917 ± 5.714e − 4	0.899 ± 1.961e − 3
TSDE timemixer	0.887 ± 4.508e − 3	0.904 ± 5.48e − 3	0.866 ± 2.983e − 3	0.892 ± 2.44e − 3	0.914 ± 4.716e − 3	0.902 ± 3.798e − 3
gemini-2.0-flash CoT +TFM lstm	0.823 ± 7.215e − 3	0.879 ± 2.338e − 3	0.803 ± 5.517e − 3	0.866 ± 3.885e − 3	0.858 ± 3.412e − 3	0.911 ± 2.789e − 3
gemini-2.0-flash CoT +TFM mlp	0.826 ± 6.77e − 3	0.874 ± 2.381e − 3	0.801 ± 5.121e − 3	0.854 ± 2.831e − 3	0.863 ± 8.082e − 3	0.907 ± 1.713e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.799 ± 6.955e − 3	0.872 ± 2.665e − 4	0.774 ± 3.663e − 2	0.837 ± 2.663e − 2	0.846 ± 9.188e − 4	0.9 ± 3.033e − 3
gemini-2.0-flash CoT +TFM timemixer	0.816 ± 5.945e − 3	0.875 ± 8.525e − 3	0.79 ± 3.963e − 3	0.855 ± 1.119e − 3	0.839 ± 2.197e − 3	0.897 ± 3.462e − 3
gemini-2.0-flash CoT lstm	0.8 ± 8.11e − 3	0.869 ± 7.762e − 3	0.759 ± 8.907e − 3	0.823 ± 6.927e − 3	0.842 ± 7.014e − 3	0.909 ± 5.895e − 3
gemini-2.0-flash CoT mlp	0.805 ± 9.615e − 3	0.868 ± 4.96e − 3	0.762 ± 8.676e − 3	0.827 ± 9.066e − 3	0.845 ± 7.499e − 3	0.909 ± 4.946e − 3
gemini-2.0-flash CoT patchtsmixer	0.794 ± 8.547e − 3	0.867 ± 1.346e − 3	0.753 ± 8.305e − 3	0.82 ± 9.924e − 3	0.833 ± 5.914e − 3	0.907 ± 2.657e − 3
gemini-2.0-flash CoT timemixer	0.794 ± 8.417e − 3	0.868 ± 2.328e − 3	0.752 ± 6.761e − 3	0.821 ± 7.768e − 3	0.839 ± 5.666e − 3	0.914 ± 4.218e − 3
gemini-2.0-flash ICD +TFM lstm	0.819 ± 5.572e − 3	0.885 ± 5.662e − 3	0.783 ± 6.991e − 3	0.85 ± 5.63e − 3	0.825 ± 3.245e − 3	0.905 ± 2.993e − 3
gemini-2.0-flash ICD +TFM mlp	0.821 ± 3.211e − 3	0.877 ± 1.819e − 3	0.782 ± 6.43e − 3	0.838 ± 5.876e − 3	0.83 ± 5.002e − 4	0.905 ± 5.133e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.806 ± 2.056e − 3	0.87 ± 5.716e − 3	0.767 ± 6.095e − 3	0.832 ± 4.4e − 3	0.81 ± 8.236e − 3	0.907 ± 2.886e − 3
gemini-2.0-flash ICD +TFM timemixer	0.804 ± 7.385e − 3	0.878 ± 1.637e − 3	0.771 ± 7.967e − 3	0.841 ± 4.318e − 3	0.813 ± 9.3e − 3	0.91 ± 3.109e − 3
gemini-2.0-flash ICD lstm	0.799 ± 7.573e − 3	0.873 ± 8.271e − 3	0.73 ± 6.661e − 3	0.829 ± 4.84e − 3	0.809 ± 4.763e − 3	0.913 ± 5.972e − 3
gemini-2.0-flash ICD mlp	0.803 ± 9.185e − 3	0.875 ± 6.48e − 3	0.733 ± 6.468e − 3	0.822 ± 3.005e − 3	0.809 ± 5.461e − 3	0.912 ± 6.076e − 3
gemini-2.0-flash ICD patchtsmixer	0.794 ± 1.271e − 2	0.875 ± 5.899e − 3	0.717 ± 1.344e − 2	0.819 ± 5.034e − 3	0.802 ± 4.614e − 3	0.916 ± 1.032e − 2
gemini-2.0-flash ICD timemixer	0.789 ± 4.491e − 3	0.873 ± 3.424e − 3	0.719 ± 5.836e − 3	0.822 ± 2.4e − 3	0.806 ± 2.704e − 3	0.918 ± 4.693e − 3
gemini-2.0-flash Trend +TFM lstm	0.805 ± 2.186e − 3	0.869 ± 3.724e − 3	0.804 ± 7.659e − 3	0.857 ± 7.629e − 3	0.851 ± 7.153e − 3	0.906 ± 5.343e − 4
gemini-2.0-flash Trend +TFM mlp	0.814 ± 1.586e − 3	0.864 ± 4.928e − 3	0.811 ± 3.376e − 3	0.85 ± 6.824e − 4	0.851 ± 3.532e − 3	0.908 ± 1.825e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.79 ± 1.242e − 2	0.851 ± 7.54e − 3	0.789 ± 5.27e − 3	0.84 ± 4.126e − 3	0.837 ± 7.07e − 3	0.896 ± 2.298e − 3
gemini-2.0-flash Trend +TFM timemixer	0.795 ± 4.414e − 3	0.864 ± 1.924e − 3	0.788 ± 3.605e − 3	0.843 ± 4.348e − 3	0.83 ± 1.029e − 2	0.904 ± 2.206e − 3
gemini-2.0-flash Trend lstm	0.786 ± 1.001e − 2	0.854 ± 5.272e − 3	0.76 ± 9.8e − 3	0.824 ± 2.86e − 3	0.838 ± 7e − 3	0.91 ± 5.622e − 3
gemini-2.0-flash Trend mlp	0.79 ± 8.753e − 3	0.852 ± 6.831e − 3	0.762 ± 9.526e − 3	0.821 ± 4.046e − 3	0.838 ± 7.23e − 3	0.912 ± 7.303e − 3
gemini-2.0-flash Trend patchtsmixer	0.775 ± 6.171e − 3	0.853 ± 1.358e − 2	0.757 ± 1.251e − 2	0.816 ± 8.842e − 3	0.827 ± 3.24e − 3	0.916 ± 5.829e − 3
gemini-2.0-flash Trend timemixer	0.776 ± 1.131e − 2	0.849 ± 1.048e − 2	0.746 ± 4.479e − 3	0.817 ± 7.589e − 3	0.825 ± 4.43e − 3	0.914 ± 7.294e − 4
gemini-2.0-flash zero_shot +TFM lstm	0.863 ± 4.987e − 3	0.917 ± 5.998e − 3	0.864 ± 5.265e − 3	0.913 ± 5.542e − 3	0.896 ± 1.43e − 3	0.92 ± 8.127e − 4
gemini-2.0-flash zero_shot +TFM mlp	0.862 ± 2.848e − 3	0.902 ± 2.268e − 3	0.864 ± 3.762e − 3	0.903 ± 2.7e − 3	0.901 ± 6.656e − 3	0.924 ± 2.546e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.816 ± 2.215e − 2	0.877 ± 1.103e − 2	0.847 ± 3.211e − 3	0.897 ± 2.887e − 3	0.886 ± 4.837e − 3	0.911 ± 2.735e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.849 ± 2.128e − 3	0.895 ± 3.301e − 3	0.848 ± 4.601e − 3	0.902 ± 1.879e − 3	0.876 ± 1.335e − 2	0.904 ± 6.197e − 3
gemini-2.0-flash zero_shot lstm	0.843 ± 9.137e − 3	0.897 ± 5.75e − 3	0.852 ± 8.835e − 3	0.903 ± 3.167e − 3	0.889 ± 9.931e − 3	0.918 ± 7.836e − 3
gemini-2.0-flash zero_shot mlp	0.846 ± 7.775e − 3	0.894 ± 6.53e − 3	0.851 ± 9.353e − 3	0.9 ± 1.722e − 3	0.889 ± 1.015e − 2	0.914 ± 5.727e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.832 ± 1.033e − 2	0.89 ± 2.778e − 3	0.838 ± 7.206e − 3	0.894 ± 2.895e − 3	0.869 ± 8.608e − 3	0.907 ± 2.16e − 3
gemini-2.0-flash zero_shot timemixer	0.832 ± 6.102e − 3	0.89 ± 4.892e − 3	0.842 ± 7.26e − 3	0.895 ± 5.414e − 3	0.876 ± 7.177e − 3	0.91 ± 3.559e − 3
interp lstm	0.899 ± 9.323e − 4	0.886 ± 3.813e − 3	0.882 ± 1.06e − 3	0.875 ± 5.809e − 3	0.923 ± 1.772e − 3	0.882 ± 5.66e − 3
interp mlp	0.898 ± 9.976e − 4	0.889 ± 2.19e − 3	0.869 ± 2.169e − 3	0.84 ± 4.341e − 3	0.925 ± 6.745e − 4	0.887 ± 4.192e − 3
interp patchtsmixer	0.887 ± 1.387e − 3	0.879 ± 6.754e − 3	0.848 ± 2.734e − 3	0.844 ± 1.791e − 2	0.848 ± 4.845e − 3	0.837 ± 1.682e − 2
interp timemixer	0.889 ± 3.617e − 3	0.89 ± 6.172e − 3	0.857 ± 2.195e − 3	0.843 ± 7.52e − 3	0.851 ± 3.304e − 3	0.846 ± 1.36e − 2
mean lstm	0.905 ± 1.05e − 3	0.892 ± 3.051e − 3	0.888 ± 2.364e − 3	0.884 ± 4.763e − 3	0.924 ± 7.926e − 4	0.881 ± 2.693e − 3
mean mlp	0.901 ± 6.388e − 4	0.887 ± 2.212e − 3	0.875 ± 1.632e − 3	0.855 ± 4.365e − 3	0.926 ± 2.396e − 4	0.887 ± 1.678e − 3
mean patchtsmixer	0.883 ± 3.368e − 3	0.878 ± 1.275e − 2	0.844 ± 5.126e − 3	0.838 ± 1.595e − 2	0.844 ± 2.806e − 3	0.834 ± 1.359e − 2
mean timemixer	0.892 ± 2.522e − 3	0.898 ± 5.943e − 3	0.861 ± 2.061e − 3	0.854 ± 4.307e − 3	0.863 ± 9.474e − 4	0.859 ± 4.567e − 3
No_sum_CoT CoT +TFM lstm	0.9 ± 4.496e − 3	0.915 ± 1.419e − 3	0.892 ± 9.205e − 4	0.907 ± 2.807e − 3	0.931 ± 3.132e − 3	0.915 ± 2.932e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.879 ± 0e0	0.905 ± 0e0	0.875 ± 0e0	0.903 ± 0e0	0.92 ± 0e0	0.913 ± 0e0
No_sum_CoT lstm	0.879 ± 3.153e − 3	0.911 ± 1.34e − 3	0.876 ± 2.63e − 3	0.903 ± 2.297e − 3	0.92 ± 4.208e − 3	0.911 ± 2.396e − 3
No_sum_CoT mlp	0.88 ± 5.433e − 3	0.911 ± 3.895e − 3	0.875 ± 2.234e − 3	0.901 ± 8.699e − 4	0.922 ± 4.788e − 3	0.913 ± 4.862e − 3
No_sum_CoT patchtsmixer	0.868 ± 1.84e − 3	0.91 ± 2.465e − 3	0.867 ± 1.265e − 3	0.9 ± 2.293e − 3	0.915 ± 6.256e − 3	0.914 ± 5.206e − 3
No_sum_CoT timemixer	0.866 ± 4.418e − 3	0.912 ± 3.306e − 3	0.864 ± 2.326e − 3	0.902 ± 2.705e − 3	0.91 ± 1.786e − 3	0.913 ± 1.181e − 3
No_sum_ICD ICD +TFM lstm	0.9 ± 1.977e − 3	0.915 ± 1.642e − 3	0.895 ± 2.344e − 3	0.909 ± 1.955e − 3	0.932 ± 2.082e − 3	0.916 ± 2.257e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.883 ± 0e0	0.906 ± 0e0	0.867 ± 0e0	0.897 ± 0e0	0.908 ± 0e0	0.906 ± 0e0
No_sum_ICD lstm	0.876 ± 2.698e − 3	0.91 ± 2.129e − 3	0.873 ± 1.294e − 3	0.902 ± 1.303e − 3	0.92 ± 3.666e − 3	0.91 ± 1.341e − 3
No_sum_ICD mlp	0.873 ± 2.943e − 3	0.908 ± 2.162e − 3	0.873 ± 4.679e − 3	0.901 ± 1.18e − 3	0.922 ± 3.821e − 3	0.911 ± 3.056e − 3
No_sum_ICD patchtsmixer	0.868 ± 2.209e − 3	0.911 ± 1.463e − 3	0.867 ± 1.446e − 3	0.9 ± 2.08e − 3	0.911 ± 4.231e − 3	0.91 ± 2.351e − 3
No_sum_ICD timemixer	0.861 ± 3.385e − 3	0.911 ± 1.886e − 3	0.861 ± 3.087e − 3	0.901 ± 2.361e − 3	0.909 ± 5.276e − 3	0.91 ± 3.126e − 3
No_sum_Trend Trend +TFM lstm	0.899 ± 3.564e − 3	0.913 ± 1.287e − 3	0.893 ± 4.497e − 3	0.908 ± 1.725e − 3	0.933 ± 2.694e − 3	0.917 ± 2.875e − 4
No_sum_Trend Trend +TFM patchtsmixer	0.891 ± 0e0	0.91 ± 0e0	0.88 ± 0e0	0.904 ± 0e0	0.927 ± 0e0	0.921 ± 0e0
No_sum_Trend lstm	0.879 ± 6.174e − 3	0.91 ± 3.441e − 3	0.877 ± 3.142e − 3	0.901 ± 1.489e − 3	0.919 ± 5.606e − 3	0.91 ± 2.737e − 3
No_sum_Trend mlp	0.882 ± 3.976e − 3	0.91 ± 2.915e − 3	0.877 ± 3.551e − 3	0.901 ± 7.657e − 4	0.922 ± 5.061e − 3	0.914 ± 3.682e − 3
No_sum_Trend patchtsmixer	0.869 ± 5.659e − 3	0.909 ± 2.656e − 3	0.869 ± 2.656e − 3	0.9 ± 1.744e − 3	0.914 ± 2.523e − 3	0.911 ± 3.485e − 3
No_sum_Trend timemixer	0.862 ± 3.996e − 3	0.912 ± 3.809e − 3	0.869 ± 1.694e − 3	0.901 ± 1.501e − 3	0.91 ± 3.811e − 3	0.911 ± 1.347e − 3
medgemma CoT +TFM lstm	0.845 ± 3.854e − 3	0.893 ± 3.063e − 3	0.822 ± 7.75e − 3	0.886 ± 5.495e − 3	0.853 ± 2.784e − 3	0.914 ± 9.8e − 4
medgemma CoT +TFM mlp	0.845 ± 7.988e − 3	0.893 ± 5.966e − 3	0.828 ± 8.741e − 4	0.874 ± 2.578e − 3	0.859 ± 5.718e − 3	0.914 ± 2.687e − 3

Table 26: In-distribution results - Drug. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro	f1_micro	recall_micro
medgemma CoT +TFM patchtsmixer	$0.816 \pm 1.606e-2$	$0.876 \pm 6.866e-3$	$0.811 \pm 6.325e-3$	$0.867 \pm 3.918e-3$	$0.833 \pm 1.357e-2$	$0.912 \pm 5.332e-3$
medgemma CoT +TFM timemixer	$0.824 \pm 4.624e-3$	$0.887 \pm 2.053e-3$	$0.805 \pm 6.915e-3$	$0.873 \pm 6.623e-3$	$0.841 \pm 8.596e-3$	$0.904 \pm 7.508e-3$
medgemma CoT lstm	$0.821 \pm 8.278e-3$	$0.88 \pm 5.981e-3$	$0.787 \pm 1.078e-2$	$0.855 \pm 2.705e-3$	$0.84 \pm 7.288e-3$	$0.919 \pm 4.913e-3$
medgemma CoT mlp	$0.824 \pm 7.467e-3$	$0.884 \pm 2.531e-3$	$0.788 \pm 1.038e-2$	$0.849 \pm 4.009e-3$	$0.841 \pm 6.244e-3$	$0.918 \pm 3.274e-3$
medgemma CoT patchtsmixer	$0.816 \pm 6.976e-3$	$0.882 \pm 1.833e-3$	$0.773 \pm 8.922e-3$	$0.848 \pm 5.822e-3$	$0.828 \pm 4.332e-3$	$0.919 \pm 3.551e-3$
medgemma CoT timemixer	$0.81 \pm 4.429e-3$	$0.88 \pm 1.775e-3$	$0.775 \pm 5.807e-3$	$0.847 \pm 2.43e-3$	$0.833 \pm 5.616e-3$	$0.922 \pm 3.817e-3$
medgemma ICD +TFM lstm	$0.833 \pm 5.142e-3$	$0.893 \pm 6.274e-3$	$0.832 \pm 5.928e-3$	$0.891 \pm 4.639e-3$	$0.863 \pm 3.365e-3$	$0.916 \pm 3.334e-3$
medgemma ICD +TFM mlp	$0.835 \pm 4.035e-3$	$0.89 \pm 1.505e-3$	$0.831 \pm 3.681e-3$	$0.877 \pm 2.38e-3$	$0.869 \pm 3.292e-3$	$0.914 \pm 2.275e-3$
medgemma ICD +TFM patchtsmixer	$0.821 \pm 5.796e-3$	$0.882 \pm 2.384e-3$	$0.816 \pm 1.204e-2$	$0.87 \pm 4.382e-3$	$0.847 \pm 1.173e-2$	$0.907 \pm 3.869e-3$
medgemma ICD +TFM timemixer	$0.824 \pm 9.086e-3$	$0.891 \pm 2.021e-3$	$0.82 \pm 2.987e-3$	$0.876 \pm 5.201e-3$	$0.846 \pm 2.419e-3$	$0.912 \pm 2.627e-3$
medgemma ICD lstm	$0.814 \pm 8.569e-3$	$0.888 \pm 3.051e-3$	$0.79 \pm 6.621e-3$	$0.869 \pm 2.355e-3$	$0.854 \pm 9.239e-3$	$0.916 \pm 3.989e-3$
medgemma ICD mlp	$0.812 \pm 7.204e-3$	$0.882 \pm 2.433e-3$	$0.797 \pm 9.16e-3$	$0.863 \pm 8.231e-4$	$0.856 \pm 7.955e-3$	$0.914 \pm 4.12e-3$
medgemma ICD patchtsmixer	$0.801 \pm 5.345e-3$	$0.887 \pm 5.366e-3$	$0.782 \pm 5.648e-3$	$0.861 \pm 5.315e-3$	$0.843 \pm 7.95e-3$	$0.916 \pm 2.581e-3$
medgemma ICD timemixer	$0.804 \pm 7.716e-3$	$0.883 \pm 2.69e-3$	$0.784 \pm 6.648e-3$	$0.86 \pm 2.863e-3$	$0.848 \pm 6.34e-3$	$0.916 \pm 1.932e-3$
medgemma Trend +TFM lstm	$0.807 \pm 5.339e-3$	$0.878 \pm 4.003e-3$	$0.791 \pm 5.387e-3$	$0.856 \pm 4.259e-3$	$0.826 \pm 9.506e-4$	$0.915 \pm 5.121e-3$
medgemma Trend +TFM mlp	$0.825 \pm 5.077e-3$	$0.875 \pm 3.013e-3$	$0.792 \pm 4.857e-3$	$0.842 \pm 1.411e-3$	$0.83 \pm 7.246e-3$	$0.911 \pm 2.66e-3$
medgemma Trend +TFM patchtsmixer	$0.791 \pm 3.926e-3$	$0.871 \pm 3.617e-3$	$0.769 \pm 6.276e-3$	$0.829 \pm 6.85e-3$	$0.812 \pm 1.385e-2$	$0.912 \pm 6.464e-3$
medgemma Trend +TFM timemixer	$0.797 \pm 1.197e-2$	$0.872 \pm 3.333e-3$	$0.776 \pm 3.095e-3$	$0.838 \pm 2.689e-3$	$0.818 \pm 9.05e-3$	$0.915 \pm 3.635e-3$
medgemma Trend lstm	$0.789 \pm 9.638e-3$	$0.872 \pm 4.757e-3$	$0.737 \pm 8.961e-3$	$0.818 \pm 5.73e-3$	$0.813 \pm 4.827e-3$	$0.924 \pm 1.156e-2$
medgemma Trend mlp	$0.794 \pm 9.509e-3$	$0.868 \pm 5.495e-3$	$0.74 \pm 9.047e-3$	$0.82 \pm 4.334e-3$	$0.812 \pm 3.613e-3$	$0.923 \pm 8.403e-3$
medgemma Trend patchtsmixer	$0.781 \pm 6.313e-3$	$0.87 \pm 7.499e-3$	$0.73 \pm 6.975e-3$	$0.821 \pm 5.311e-3$	$0.809 \pm 2.999e-3$	$0.926 \pm 9.635e-3$
medgemma Trend timemixer	$0.781 \pm 5.326e-3$	$0.87 \pm 9.221e-3$	$0.728 \pm 6.475e-3$	$0.822 \pm 2.376e-3$	$0.81 \pm 3.251e-3$	$0.927 \pm 1.016e-2$
medgemma zero_shot +TFM lstm	$0.853 \pm 3.201e-3$	$0.908 \pm 2.792e-3$	$0.863 \pm 7.753e-4$	$0.914 \pm 6.976e-4$	$0.89 \pm 5.206e-3$	$0.919 \pm 3.282e-3$
medgemma zero_shot +TFM mlp	$0.871 \pm 5.77e-3$	$0.908 \pm 2.679e-3$	$0.855 \pm 4.225e-3$	$0.899 \pm 1.999e-3$	$0.891 \pm 5.443e-3$	$0.917 \pm 3.266e-3$
medgemma zero_shot +TFM patchtsmixer	$0.84 \pm 7.043e-3$	$0.897 \pm 7.143e-4$	$0.844 \pm 7.993e-3$	$0.893 \pm 6.471e-3$	$0.87 \pm 9.705e-3$	$0.904 \pm 5.752e-3$
medgemma zero_shot +TFM timemixer	$0.837 \pm 5.705e-3$	$0.892 \pm 4.526e-3$	$0.842 \pm 2.596e-3$	$0.896 \pm 3.402e-3$	$0.869 \pm 7.885e-3$	$0.907 \pm 6.656e-3$
medgemma zero_shot lstm	$0.838 \pm 7.46e-3$	$0.901 \pm 3.792e-3$	$0.838 \pm 9.671e-3$	$0.895 \pm 9.065e-3$	$0.874 \pm 9.389e-3$	$0.915 \pm 3.963e-3$
medgemma zero_shot mlp	$0.84 \pm 8.177e-3$	$0.899 \pm 1.083e-2$	$0.839 \pm 9.327e-3$	$0.893 \pm 6.112e-3$	$0.878 \pm 8.874e-3$	$0.91 \pm 3.853e-3$
medgemma zero_shot patchtsmixer	$0.825 \pm 2.188e-3$	$0.887 \pm 3.195e-3$	$0.828 \pm 1.111e-2$	$0.883 \pm 6.946e-3$	$0.866 \pm 8.997e-3$	$0.907 \pm 4.078e-3$
medgemma zero_shot timemixer	$0.828 \pm 9.289e-3$	$0.895 \pm 5.052e-3$	$0.83 \pm 8.241e-3$	$0.888 \pm 4.706e-3$	$0.868 \pm 8.289e-3$	$0.911 \pm 3.518e-3$
No_sum_zero_shot zero_shot +TFM lstm	$0.901 \pm 6.503e-3$	$0.912 \pm 1.125e-3$	$0.891 \pm 4.215e-3$	$0.906 \pm 1.323e-3$	$0.932 \pm 3.949e-3$	$0.916 \pm 3.882e-3$
No_sum_zero_shot zero_shot +TFM patchtsmixer	$0.875 \pm 0e0$	$0.903 \pm 0e0$	$0.883 \pm 0e0$	$0.905 \pm 0e0$	$0.921 \pm 0e0$	$0.914 \pm 0e0$
No_sum_zero_shot lstm	$0.884 \pm 3.829e-3$	$0.908 \pm 2.785e-3$	$0.878 \pm 5.327e-3$	$0.902 \pm 2.01e-3$	$0.921 \pm 6.224e-3$	$0.911 \pm 2.664e-3$
No_sum_zero_shot mlp	$0.884 \pm 4.159e-3$	$0.911 \pm 3.471e-3$	$0.875 \pm 5.187e-3$	$0.901 \pm 6.915e-4$	$0.923 \pm 5.792e-3$	$0.913 \pm 4.059e-3$
No_sum_zero_shot patchtsmixer	$0.872 \pm 4.123e-3$	$0.907 \pm 3.57e-3$	$0.867 \pm 1.975e-3$	$0.903 \pm 3.036e-3$	$0.913 \pm 3.249e-3$	$0.914 \pm 1.405e-3$
No_sum_zero_shot timemixer	$0.864 \pm 1.586e-3$	$0.912 \pm 1.225e-3$	$0.867 \pm 1.081e-3$	$0.902 \pm 2.729e-3$	$0.908 \pm 3.663e-3$	$0.915 \pm 3.573e-3$
right lstm	$0.903 \pm 9.254e-4$	$0.891 \pm 2.123e-3$	$0.886 \pm 2.159e-3$	$0.882 \pm 5.877e-3$	$0.924 \pm 1.166e-3$	$0.883 \pm 3.702e-3$
right mlp	$0.901 \pm 2.061e-4$	$0.89 \pm 1.767e-3$	$0.88 \pm 8.09e-4$	$0.86 \pm 3.343e-3$	$0.927 \pm 6.405e-4$	$0.892 \pm 1.716e-3$
right patchtsmixer	$0.888 \pm 2.581e-3$	$0.884 \pm 6.439e-3$	$0.849 \pm 3.944e-3$	$0.841 \pm 2.423e-2$	$0.848 \pm 6.675e-3$	$0.831 \pm 2.645e-2$
right timemixer	$0.9 \pm 1.44e-3$	$0.895 \pm 4.593e-3$	$0.863 \pm 3.713e-4$	$0.85 \pm 2.895e-3$	$0.859 \pm 4.092e-3$	$0.846 \pm 1.355e-2$

Table 27: In-distribution results - Drug. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro
Llama-3.1 CoT +TFM lstm	0.769 ± 1.565e − 3	0.931 ± 1.886e − 3	0.777 ± 2.04e − 3	0.931 ± 1.593e − 3
Llama-3.1 CoT +TFM mlp	0.763 ± 1.471e − 3	0.933 ± 1.939e − 3	0.774 ± 1.628e − 3	0.938 ± 4.576e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.75 ± 2.781e − 3	0.945 ± 4.113e − 3	0.767 ± 3.535e − 3	0.932 ± 1.907e − 3
Llama-3.1 CoT +TFM timemixer	0.764 ± 3.39e − 3	0.937 ± 5.739e − 3	0.772 ± 1.664e − 3	0.936 ± 6.964e − 4
Llama-3.1 CoT lstm	0.761 ± 3.771e − 3	0.944 ± 3.279e − 3	0.766 ± 4.793e − 4	0.95 ± 7.115e − 3
Llama-3.1 CoT mlp	0.762 ± 2.565e − 3	0.944 ± 7.042e − 3	0.765 ± 1.757e − 3	0.95 ± 6.242e − 3
Llama-3.1 CoT patchtsmixer	0.757 ± 2.87e − 3	0.953 ± 8.317e − 3	0.764 ± 1.703e − 3	0.949 ± 6.546e − 3
Llama-3.1 CoT timemixer	0.76 ± 3.443e − 3	0.951 ± 6.63e − 3	0.764 ± 9.235e − 4	0.952 ± 8.599e − 3
Llama-3.1 ICD +TFM lstm	0.769 ± 1.131e − 3	0.927 ± 4.855e − 3	0.78 ± 2.845e − 4	0.934 ± 3.966e − 3
Llama-3.1 ICD +TFM mlp	0.772 ± 3.155e − 3	0.937 ± 1.884e − 3	0.78 ± 1.354e − 3	0.937 ± 3.736e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.754 ± 1.392e − 3	0.935 ± 4.183e − 3	0.774 ± 2.245e − 3	0.928 ± 2.228e − 3
Llama-3.1 ICD +TFM timemixer	0.76 ± 6.022e − 3	0.934 ± 4.85e − 3	0.779 ± 1.468e − 3	0.937 ± 5.118e − 3
Llama-3.1 ICD lstm	0.773 ± 1.46e − 3	0.947 ± 3.975e − 3	0.776 ± 1.883e − 3	0.948 ± 8.712e − 3
Llama-3.1 ICD mlp	0.774 ± 1.227e − 3	0.942 ± 6.249e − 3	0.775 ± 2.145e − 3	0.947 ± 3.594e − 3
Llama-3.1 ICD patchtsmixer	0.771 ± 2.857e − 3	0.941 ± 4.237e − 3	0.774 ± 2.882e − 3	0.952 ± 3.825e − 3
Llama-3.1 ICD timemixer	0.774 ± 2.123e − 3	0.946 ± 2.746e − 3	0.776 ± 2.973e − 3	0.95 ± 2.51e − 3
Llama-3.1 Trend +TFM lstm	0.789 ± 1.766e − 3	0.925 ± 2.241e − 3	0.789 ± 1.162e − 3	0.922 ± 3.262e − 3
Llama-3.1 Trend +TFM mlp	0.786 ± 3.692e − 3	0.931 ± 1.383e − 3	0.79 ± 2.015e − 3	0.93 ± 3.038e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.771 ± 2.673e − 3	0.923 ± 3.235e − 3	0.784 ± 1.017e − 3	0.926 ± 6.444e − 3
Llama-3.1 Trend +TFM timemixer	0.785 ± 2.486e − 3	0.932 ± 2.821e − 3	0.788 ± 1.255e − 3	0.921 ± 2.751e − 3
Llama-3.1 Trend lstm	0.789 ± 1.779e − 3	0.935 ± 3.835e − 3	0.785 ± 1.248e − 3	0.925 ± 1.794e − 3
Llama-3.1 Trend mlp	0.789 ± 1.819e − 3	0.937 ± 2.334e − 3	0.784 ± 2.716e − 3	0.929 ± 5.426e − 3
Llama-3.1 Trend patchtsmixer	0.786 ± 2.478e − 3	0.939 ± 8.783e − 4	0.779 ± 4.816e − 3	0.934 ± 4.188e − 3
Llama-3.1 Trend timemixer	0.789 ± 2.811e − 3	0.939 ± 3.333e − 3	0.782 ± 2.986e − 3	0.932 ± 2.18e − 3
Llama-3.1 zero_shot +TFM lstm	0.77 ± 2.102e − 3	0.94 ± 2.732e − 3	0.779 ± 2.321e − 3	0.933 ± 2.519e − 3
Llama-3.1 zero_shot +TFM mlp	0.767 ± 2.487e − 3	0.942 ± 1.719e − 3	0.778 ± 1.441e − 3	0.94 ± 4.035e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.747 ± 1.155e − 3	0.941 ± 3.095e − 3	0.771 ± 2.038e − 3	0.928 ± 4.455e − 3
Llama-3.1 zero_shot +TFM timemixer	0.763 ± 3.452e − 3	0.941 ± 1.541e − 3	0.773 ± 1.115e − 3	0.934 ± 3.409e − 3
Llama-3.1 zero_shot lstm	0.774 ± 4.573e − 3	0.951 ± 6.489e − 4	0.773 ± 3.396e − 3	0.95 ± 3.991e − 3
Llama-3.1 zero_shot mlp	0.773 ± 3.335e − 3	0.95 ± 4.034e − 3	0.774 ± 2.273e − 3	0.95 ± 2.508e − 3
Llama-3.1 zero_shot patchtsmixer	0.772 ± 4.054e − 3	0.952 ± 2.209e − 3	0.774 ± 2.935e − 3	0.949 ± 4.178e − 3
Llama-3.1 zero_shot timemixer	0.775 ± 3.184e − 3	0.952 ± 2.072e − 3	0.775 ± 2.595e − 3	0.953 ± 1.233e − 3
TFM lstm	0.792 ± 6.625e − 4	0.934 ± 3.326e − 3	0.788 ± 2.603e − 4	0.941 ± 2.725e − 3
TFM mlp	0.79 ± 1.096e − 3	0.932 ± 1.591e − 3	0.783 ± 1.592e − 3	0.944 ± 2.947e − 3
TFM patchtsmixer	0.779 ± 1.063e − 3	0.935 ± 2.672e − 3	0.781 ± 1.161e − 3	0.947 ± 2.087e − 3
TFM timemixer	0.785 ± 1.221e − 3	0.943 ± 3.055e − 3	0.787 ± 9.39e − 4	0.941 ± 1.408e − 3
TSDE mlp	0.731 ± 5.055e − 4	0.916 ± 1.303e − 3	0.731 ± 1.478e − 3	0.939 ± 3.459e − 3
TSDE lstm	0.732 ± 5.055e − 4	0.917 ± 1.159e − 3	0.733 ± 1.27e − 3	0.905 ± 2.959e − 3
TSDE patchtsmixer	0.732 ± 5.084e − 4	0.909 ± 9.785e − 4	0.733 ± 3.643e − 4	0.899 ± 2.846e − 4
TSDE timemixer	0.732 ± 5.027e − 4	0.918 ± 1.937e − 3	0.734 ± 1.911e − 3	0.904 ± 3.655e − 3
gemini-2.0-flash CoT +TFM lstm	0.795 ± 2.031e − 4	0.922 ± 4.589e − 3	0.797 ± 1.453e − 3	0.919 ± 4.039e − 3
gemini-2.0-flash CoT +TFM mlp	0.792 ± 9.546e − 4	0.927 ± 4.131e − 3	0.799 ± 2.066e − 3	0.926 ± 1.863e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.77 ± 6.553e − 4	0.912 ± 3.054e − 3	0.778 ± 6.061e − 3	0.916 ± 4.519e − 3
gemini-2.0-flash CoT +TFM timemixer	0.779 ± 2.736e − 3	0.921 ± 2.092e − 3	0.793 ± 1.69e − 3	0.921 ± 2.353e − 3
gemini-2.0-flash CoT lstm	0.79 ± 3.174e − 3	0.93 ± 1.465e − 3	0.798 ± 2.544e − 4	0.934 ± 4.794e − 3
gemini-2.0-flash CoT mlp	0.791 ± 2.826e − 3	0.93 ± 2.745e − 3	0.798 ± 1.862e − 3	0.931 ± 6.755e − 3
gemini-2.0-flash CoT patchtsmixer	0.788 ± 2.483e − 3	0.93 ± 2.14e − 3	0.797 ± 1.708e − 3	0.932 ± 5.13e − 3
gemini-2.0-flash CoT timemixer	0.793 ± 1.82e − 3	0.93 ± 2.094e − 3	0.799 ± 2.242e − 3	0.929 ± 1.58e − 3
gemini-2.0-flash ICD +TFM lstm	0.76 ± 1.129e − 3	0.929 ± 1.224e − 3	0.774 ± 1.572e − 3	0.936 ± 2.818e − 3
gemini-2.0-flash ICD +TFM mlp	0.761 ± 1.293e − 3	0.937 ± 3.918e − 3	0.777 ± 1.703e − 3	0.935 ± 1.565e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.736 ± 2.042e − 3	0.968 ± 2.401e − 3	0.766 ± 7.831e − 4	0.929 ± 1.971e − 3
gemini-2.0-flash ICD +TFM timemixer	0.753 ± 4.107e − 3	0.947 ± 6.716e − 3	0.773 ± 8.622e − 4	0.933 ± 1.654e − 3
gemini-2.0-flash ICD lstm	0.749 ± 2.683e − 3	0.954 ± 5.802e − 3	0.781 ± 2.21e − 3	0.938 ± 7.867e − 4
gemini-2.0-flash ICD mlp	0.749 ± 3.572e − 3	0.956 ± 2.767e − 3	0.781 ± 2.93e − 3	0.939 ± 4.633e − 3
gemini-2.0-flash ICD patchtsmixer	0.749 ± 5.385e − 3	0.959 ± 4.889e − 3	0.779 ± 2.487e − 3	0.94 ± 4.035e − 3
gemini-2.0-flash ICD timemixer	0.757 ± 1.928e − 3	0.949 ± 3.16e − 3	0.78 ± 2.748e − 3	0.94 ± 4.242e − 3
gemini-2.0-flash Trend +TFM lstm	0.781 ± 4.03e − 3	0.928 ± 5.388e − 3	0.783 ± 6.712e − 4	0.93 ± 1.714e − 3
gemini-2.0-flash Trend +TFM mlp	0.777 ± 2.635e − 3	0.931 ± 4.03e − 3	0.784 ± 4.19e − 3	0.933 ± 2.811e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.759 ± 5.442e − 3	0.925 ± 6.722e − 3	0.774 ± 2.541e − 3	0.923 ± 2.473e − 3
gemini-2.0-flash Trend +TFM timemixer	0.774 ± 3.092e − 3	0.934 ± 3.535e − 3	0.783 ± 9.199e − 4	0.928 ± 6.544e − 3
gemini-2.0-flash Trend lstm	0.786 ± 1.8e − 3	0.935 ± 4.027e − 3	0.785 ± 2.572e − 3	0.941 ± 4.679e − 3
gemini-2.0-flash Trend mlp	0.785 ± 2.353e − 3	0.935 ± 4.557e − 3	0.787 ± 1.862e − 3	0.94 ± 3.453e − 3
gemini-2.0-flash Trend patchtsmixer	0.783 ± 2.399e − 3	0.939 ± 5.041e − 3	0.786 ± 1.849e − 3	0.943 ± 3.641e − 3
gemini-2.0-flash Trend timemixer	0.788 ± 1.085e − 3	0.941 ± 2.566e − 3	0.787 ± 2.902e − 3	0.941 ± 1.593e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.805 ± 4.417e − 4	0.934 ± 2.364e − 3	0.803 ± 8.592e − 4	0.931 ± 1.903e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.802 ± 1.462e − 3	0.937 ± 3.03e − 3	0.805 ± 1.826e − 3	0.935 ± 1.732e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.777 ± 4.705e − 3	0.917 ± 4.439e − 3	0.791 ± 2.184e − 3	0.924 ± 1.232e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.792 ± 3.123e − 3	0.927 ± 2.555e − 3	0.801 ± 1.244e − 3	0.929 ± 1.953e − 3
gemini-2.0-flash zero_shot lstm	0.808 ± 1.19e − 3	0.946 ± 1.719e − 3	0.808 ± 1.656e − 3	0.936 ± 2.038e − 3
gemini-2.0-flash zero_shot mlp	0.807 ± 8.551e − 4	0.941 ± 2.503e − 3	0.808 ± 1.284e − 3	0.935 ± 3.185e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.804 ± 1.035e − 3	0.941 ± 3.829e − 3	0.806 ± 1.31e − 3	0.936 ± 3.405e − 3
gemini-2.0-flash zero_shot timemixer	0.808 ± 4.102e − 4	0.947 ± 3.22e − 3	0.808 ± 5.868e − 4	0.938 ± 2.419e − 3
interp lstm	0.742 ± 7.07e − 3	0.87 ± 5.803e − 2	0.516 ± 1.005e − 2	0.4 ± 1.348e − 2
interp mlp	0.723 ± 5.267e − 3	0.834 ± 3.155e − 2	0.621 ± 1.595e − 2	0.554 ± 2.712e − 2
interp patchtsmixer	0.461 ± 3.294e − 2	0.345 ± 3.335e − 2	0.325 ± 3.759e − 2	0.213 ± 3.304e − 2
interp timemixer	0.646 ± 1.65e − 2	0.609 ± 3.489e − 2	0.343 ± 1.279e − 2	0.227 ± 1.028e − 2
mean lstm	0.669 ± 9.804e − 3	0.629 ± 1.499e − 2	0.682 ± 2.045e − 2	0.663 ± 4.92e − 2
mean mlp	0.711 ± 8.323e − 3	0.712 ± 2.228e − 2	0.662 ± 7.924e − 4	0.621 ± 1.92e − 3
mean patchtsmixer	0.511 ± 5.235e − 2	0.422 ± 7.283e − 2	0.424 ± 3.362e − 2	0.303 ± 3.908e − 2
mean timemixer	0.7 ± 4.65e − 3	0.719 ± 1.265e − 2	0.497 ± 7.052e − 3	0.384 ± 1.168e − 2
No_sum_CoT CoT +TFM lstm	0.831 ± 8.552e − 4	0.953 ± 2.666e − 3	0.8 ± 4.439e − 3	0.91 ± 8.198e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.808 ± 0e0	0.93 ± 0e0	0.779 ± 0e0	0.922 ± 0e0
No_sum_CoT lstm	0.839 ± 1.693e − 3	0.961 ± 2.933e − 3	0.819 ± 1.368e − 3	0.946 ± 2.732e − 3
No_sum_CoT mlp	0.837 ± 3.725e − 4	0.957 ± 4.288e − 3	0.819 ± 2.082e − 3	0.946 ± 4.004e − 3
No_sum_CoT patchtsmixer	0.836 ± 1.488e − 3	0.963 ± 1.443e − 3	0.818 ± 6.36e − 4	0.944 ± 3.774e − 3
No_sum_CoT timemixer	0.837 ± 1.145e − 3	0.966 ± 2.013e − 3	0.824 ± 3.503e − 3	0.949 ± 2.709e − 3
No_sum_ICD ICD +TFM lstm	0.828 ± 1.955e − 3	0.947 ± 3.566e − 3	0.801 ± 4.128e − 3	0.916 ± 1.229e − 2
No_sum_ICD ICD +TFM patchtsmixer	0.81 ± 0e0	0.931 ± 0e0	0.781 ± 0e0	0.924 ± 0e0
No_sum_ICD lstm	0.836 ± 5.28e − 4	0.963 ± 4.649e − 3	0.82 ± 1.416e − 3	0.95 ± 2.91e − 3
No_sum_ICD mlp	0.836 ± 9.8e − 4	0.963 ± 2.615e − 3	0.817 ± 1.021e − 3	0.944 ± 2.966e − 3
No_sum_ICD patchtsmixer	0.836 ± 1.277e − 3	0.965 ± 1.212e − 3	0.818 ± 1.81e − 3	0.946 ± 2.933e − 3
No_sum_ICD timemixer	0.836 ± 1.503e − 3	0.971 ± 9.228e − 4	0.825 ± 2.262e − 3	0.954 ± 2.281e − 3
No_sum_Trend Trend +TFM lstm	0.83 ± 3.097e − 3	0.952 ± 5.162e − 3	0.805 ± 6.704e − 3	0.929 ± 2.766e − 3
No_sum_Trend Trend +TFM patchtsmixer	0.806 ± 0e0	0.925 ± 0e0	0.781 ± 0e0	0.925 ± 0e0
No_sum_Trend lstm	0.836 ± 2e − 3	0.961 ± 4.458e − 3	0.823 ± 1.1e − 3	0.948 ± 2.113e − 3
No_sum_Trend mlp	0.833 ± 8.873e − 4	0.953 ± 3.464e − 3	0.821 ± 3.409e − 3	0.945 ± 2.64e − 3
No_sum_Trend patchtsmixer	0.836 ± 1.59e − 3	0.962 ± 2.871e − 3	0.82 ± 1.466e − 3	0.944 ± 3.025e − 3
No_sum_Trend timemixer	0.836 ± 1.247e − 3	0.97 ± 1.808e − 3	0.826 ± 1.335e − 3	0.949 ± 2.029e − 3
medgemma CoT +TFM lstm	0.791 ± 4.491e − 4	0.912 ± 1.675e − 3	0.796 ± 1.672e − 3	0.908 ± 9.96e − 4
medgemma CoT +TFM mlp	0.79 ± 1.493e − 3	0.909 ± 5.324e − 3	0.798 ± 2.988e − 3	0.911 ± 7.002e − 3

Table 28: Cross-site transfer results - Drug. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro
medgemma CoT +TFM patchtsmixer	0.772 ± 2.573e − 3	0.907 ± 3.85e − 3	0.784 ± 9.672e − 4	0.916 ± 9.366e − 4
medgemma CoT +TFM timemixer	0.783 ± 1.245e − 3	0.912 ± 4.037e − 3	0.793 ± 2.516e − 4	0.91 ± 2.613e − 3
medgemma CoT lstm	0.788 ± 2.26e − 3	0.924 ± 1.781e − 3	0.797 ± 6.506e − 4	0.923 ± 5.205e − 3
medgemma CoT mlp	0.79 ± 6.558e − 4	0.924 ± 3.799e − 3	0.795 ± 6.553e − 4	0.926 ± 7.525e − 4
medgemma CoT patchtsmixer	0.791 ± 1.586e − 3	0.922 ± 8.402e − 4	0.793 ± 1.714e − 3	0.924 ± 3.783e − 3
medgemma CoT timemixer	0.796 ± 1.224e − 3	0.926 ± 6.978e − 4	0.795 ± 1.366e − 3	0.927 ± 5.454e − 3
medgemma ICD +TFM lstm	0.803 ± 1.08e − 3	0.936 ± 1.261e − 3	0.805 ± 1.541e − 3	0.927 ± 2.397e − 3
medgemma ICD +TFM mlp	0.803 ± 2.028e − 3	0.94 ± 3.247e − 3	0.809 ± 6.264e − 4	0.937 ± 3.902e − 3
medgemma ICD +TFM patchtsmixer	0.781 ± 6.616e − 4	0.919 ± 1.42e − 3	0.794 ± 2.607e − 3	0.919 ± 5.06e − 3
medgemma ICD +TFM timemixer	0.792 ± 6.278e − 3	0.927 ± 7.214e − 3	0.803 ± 1.484e − 3	0.928 ± 2.035e − 3
medgemma ICD lstm	0.805 ± 1.005e − 3	0.94 ± 1.491e − 3	0.808 ± 8.141e − 4	0.94 ± 1.442e − 3
medgemma ICD mlp	0.805 ± 7.957e − 4	0.941 ± 2.369e − 3	0.808 ± 8.444e − 4	0.938 ± 1.166e − 3
medgemma ICD patchtsmixer	0.803 ± 8.531e − 4	0.939 ± 3.272e − 3	0.805 ± 4.692e − 4	0.94 ± 3.634e − 3
medgemma ICD timemixer	0.806 ± 7.406e − 4	0.943 ± 3.666e − 3	0.808 ± 2.176e − 4	0.94 ± 8.943e − 4
medgemma Trend +TFM lstm	0.772 ± 9.801e − 4	0.94 ± 1.071e − 3	0.779 ± 2.408e − 3	0.928 ± 2.555e − 3
medgemma Trend +TFM mlp	0.768 ± 4.294e − 3	0.945 ± 3.258e − 3	0.781 ± 1.895e − 3	0.934 ± 1.827e − 3
medgemma Trend +TFM patchtsmixer	0.75 ± 6.952e − 4	0.943 ± 4.219e − 3	0.769 ± 1.597e − 3	0.927 ± 8.167e − 3
medgemma Trend +TFM timemixer	0.764 ± 2.796e − 3	0.94 ± 5.22e − 3	0.776 ± 1.255e − 3	0.934 ± 2.515e − 3
medgemma Trend lstm	0.775 ± 2.571e − 3	0.945 ± 4.016e − 3	0.782 ± 1.714e − 3	0.94 ± 3.073e − 3
medgemma Trend mlp	0.777 ± 2.475e − 3	0.944 ± 4.09e − 3	0.782 ± 2.611e − 3	0.941 ± 2.659e − 3
medgemma Trend patchtsmixer	0.774 ± 3.23e − 3	0.945 ± 1.613e − 3	0.78 ± 2.85e − 3	0.94 ± 3.268e − 3
medgemma Trend timemixer	0.777 ± 3.206e − 3	0.948 ± 3.016e − 3	0.782 ± 2.468e − 3	0.942 ± 4.638e − 3
medgemma zero_shot +TFM lstm	0.805 ± 1.895e − 3	0.922 ± 1.735e − 3	0.807 ± 4.661e − 4	0.924 ± 3.651e − 3
medgemma zero_shot +TFM mlp	0.804 ± 9.958e − 4	0.929 ± 1.815e − 3	0.807 ± 6.479e − 4	0.929 ± 3.858e − 3
medgemma zero_shot +TFM patchtsmixer	0.785 ± 1.413e − 3	0.914 ± 4.07e − 3	0.794 ± 2.091e − 3	0.917 ± 3.716e − 3
medgemma zero_shot +TFM timemixer	0.792 ± 8.399e − 3	0.918 ± 5.264e − 3	0.803 ± 2.323e − 3	0.919 ± 1.728e − 3
medgemma zero_shot lstm	0.805 ± 9.658e − 4	0.927 ± 2.208e − 3	0.807 ± 7.651e − 4	0.927 ± 1.444e − 3
medgemma zero_shot mlp	0.807 ± 7.138e − 4	0.931 ± 2.173e − 3	0.807 ± 9.693e − 4	0.926 ± 1.236e − 3
medgemma zero_shot patchtsmixer	0.805 ± 1.884e − 4	0.931 ± 2.154e − 3	0.804 ± 2.671e − 3	0.926 ± 8.539e − 4
medgemma zero_shot timemixer	0.809 ± 7.505e − 4	0.935 ± 1.256e − 3	0.805 ± 1.626e − 3	0.928 ± 3.542e − 3
No_sum_zero_shot zero_shot +TFM lstm	0.83 ± 2.291e − 3	0.953 ± 3.83e − 3	0.789 ± 2.903e − 3	0.909 ± 7.136e − 4
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.812 ± 0e0	0.933 ± 0e0	0.773 ± 0e0	0.918 ± 0e0
No_sum_zero_shot lstm	0.829 ± 1.474e − 3	0.955 ± 2.431e − 3	0.813 ± 2.133e − 3	0.943 ± 3.385e − 3
No_sum_zero_shot mlp	0.829 ± 7.801e − 4	0.952 ± 1.673e − 3	0.814 ± 2.352e − 3	0.943 ± 1.866e − 3
No_sum_zero_shot patchtsmixer	0.832 ± 1.151e − 3	0.958 ± 3.689e − 3	0.816 ± 1.81e − 3	0.945 ± 3.023e − 3
No_sum_zero_shot timemixer	0.835 ± 1.135e − 3	0.966 ± 2.589e − 3	0.821 ± 2.119e − 3	0.947 ± 4.018e − 3
right lstm	0.725 ± 2.011e − 2	0.769 ± 9.362e − 2	0.526 ± 1.089e − 2	0.41 ± 1.39e − 2
right mlp	0.739 ± 8.853e − 4	0.882 ± 9.386e − 3	0.601 ± 6.53e − 3	0.517 ± 8.03e − 3
right patchtsmixer	0.578 ± 2.89e − 2	0.529 ± 5.019e − 2	0.346 ± 2.33e − 2	0.234 ± 2.131e − 2
right timemixer	0.685 ± 9.987e − 3	0.722 ± 5.086e − 2	0.37 ± 5.214e − 2	0.256 ± 4.913e − 2

Table 29: Cross-site transfer results - Drug. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro	f1_micro	recall_micro
Llama-3.1 CoT +TFM lstm	0.872 ± 2.001e − 3	0.92 ± 8.798e − 4	0.895 ± 2.53e − 3	0.936 ± 1.722e − 3	0.847 ± 7.912e − 4	0.917 ± 6.809e − 4
Llama-3.1 CoT +TFM mlp	0.881 ± 2.75e − 3	0.926 ± 8.832e − 4	0.903 ± 1.691e − 3	0.94 ± 4.133e − 4	0.857 ± 1.235e − 3	0.925 ± 6.144e − 4
Llama-3.1 CoT +TFM patchtsmixer	0.857 ± 6.165e − 3	0.919 ± 5.505e − 4	0.892 ± 1.797e − 3	0.938 ± 2.665e − 4	0.839 ± 7.46e − 3	0.922 ± 1.361e − 3
Llama-3.1 CoT +TFM timemixer	0.856 ± 6.02e − 3	0.923 ± 1.165e − 3	0.87 ± 1.593e − 3	0.935 ± 8.472e − 4	0.811 ± 3.494e − 3	0.928 ± 1.361e − 3
Llama-3.1 CoT lstm	0.846 ± 0e0	0.928 ± 0e0	0.881 ± 0e0	0.946 ± 0e0	0.831 ± 0e0	0.931 ± 0e0
Llama-3.1 CoT mlp	0.863 ± 0e0	0.925 ± 0e0	0.89 ± 0e0	0.944 ± 0e0	0.847 ± 0e0	0.924 ± 0e0
Llama-3.1 CoT patchtsmixer	0.845 ± 0e0	0.927 ± 0e0	0.883 ± 0e0	0.947 ± 0e0	0.838 ± 0e0	0.93 ± 0e0
Llama-3.1 CoT timemixer	0.844 ± 0e0	0.927 ± 0e0	0.873 ± 0e0	0.948 ± 0e0	0.826 ± 0e0	0.935 ± 0e0
Llama-3.1 ICD +TFM lstm	0.873 ± 2.458e − 4	0.922 ± 1.169e − 3	0.889 ± 2.848e − 3	0.936 ± 1.269e − 3	0.83 ± 2.191e − 3	0.919 ± 1.804e − 3
Llama-3.1 ICD +TFM mlp	0.877 ± 2.737e − 3	0.926 ± 7.136e − 4	0.899 ± 2.145e − 3	0.94 ± 9.086e − 4	0.848 ± 4.533e − 3	0.923 ± 7.047e − 4
Llama-3.1 ICD +TFM patchtsmixer	0.858 ± 4.406e − 3	0.92 ± 1.552e − 3	0.886 ± 1.498e − 3	0.939 ± 1.021e − 3	0.822 ± 1.137e − 3	0.925 ± 1.746e − 3
Llama-3.1 ICD +TFM timemixer	0.851 ± 4.439e − 3	0.923 ± 9.093e − 4	0.861 ± 2.613e − 3	0.936 ± 1.537e − 4	0.811 ± 6.069e − 3	0.93 ± 7.52e − 4
Llama-3.1 ICD lstm	0.841 ± 0e0	0.931 ± 0e0	0.872 ± 0e0	0.945 ± 0e0	0.811 ± 0e0	0.936 ± 0e0
Llama-3.1 ICD mlp	0.856 ± 0e0	0.924 ± 0e0	0.881 ± 0e0	0.945 ± 0e0	0.826 ± 0e0	0.93 ± 0e0
Llama-3.1 ICD patchtsmixer	0.841 ± 0e0	0.925 ± 0e0	0.875 ± 0e0	0.945 ± 0e0	0.816 ± 0e0	0.936 ± 0e0
Llama-3.1 ICD timemixer	0.837 ± 0e0	0.927 ± 0e0	0.863 ± 0e0	0.943 ± 0e0	0.805 ± 0e0	0.942 ± 0e0
Llama-3.1 Trend +TFM lstm	0.878 ± 3.433e − 3	0.922 ± 1.257e − 3	0.896 ± 6.165e − 4	0.937 ± 6.24e − 4	0.845 ± 2.67e − 3	0.914 ± 8.072e − 4
Llama-3.1 Trend +TFM mlp	0.88 ± 1.597e − 3	0.925 ± 6.38e − 4	0.9 ± 1.567e − 3	0.941 ± 7.65e − 4	0.858 ± 1.94e − 3	0.922 ± 3.614e − 4
Llama-3.1 Trend +TFM patchtsmixer	0.844 ± 2.049e − 2	0.92 ± 2.313e − 3	0.893 ± 2.192e − 3	0.939 ± 1.601e − 4	0.845 ± 6.878e − 3	0.918 ± 6.352e − 4
Llama-3.1 Trend +TFM timemixer	0.848 ± 2.145e − 2	0.929 ± 1.162e − 2	0.875 ± 4.022e − 3	0.937 ± 2.89e − 3	0.819 ± 1.003e − 2	0.926 ± 1.526e − 3
Llama-3.1 Trend lstm	0.844 ± 0e0	0.928 ± 0e0	0.88 ± 0e0	0.943 ± 0e0	0.824 ± 0e0	0.932 ± 0e0
Llama-3.1 Trend mlp	0.861 ± 0e0	0.926 ± 0e0	0.892 ± 0e0	0.941 ± 0e0	0.847 ± 0e0	0.923 ± 0e0
Llama-3.1 Trend patchtsmixer	0.845 ± 0e0	0.924 ± 0e0	0.886 ± 0e0	0.942 ± 0e0	0.838 ± 0e0	0.929 ± 0e0
Llama-3.1 Trend timemixer	0.841 ± 0e0	0.922 ± 0e0	0.875 ± 0e0	0.944 ± 0e0	0.8 ± 0e0	0.94 ± 0e0
Llama-3.1 zero_shot +TFM lstm	0.883 ± 2.982e − 3	0.924 ± 7.996e − 4	0.895 ± 1.667e − 3	0.938 ± 1.127e − 3	0.849 ± 1.869e − 3	0.919 ± 1.553e − 3
Llama-3.1 zero_shot +TFM mlp	0.881 ± 1.58e − 3	0.928 ± 7.543e − 4	0.904 ± 7.637e − 4	0.941 ± 3.35e − 4	0.859 ± 2.73e − 3	0.927 ± 7.239e − 4
Llama-3.1 zero_shot +TFM patchtsmixer	0.861 ± 2.916e − 3	0.922 ± 2.016e − 3	0.891 ± 9.815e − 4	0.94 ± 7.253e − 4	0.842 ± 1.746e − 3	0.921 ± 1.172e − 3
Llama-3.1 zero_shot +TFM timemixer	0.863 ± 1.131e − 3	0.923 ± 7.984e − 4	0.881 ± 2.739e − 3	0.939 ± 7.721e − 4	0.816 ± 8.775e − 3	0.929 ± 1.99e − 3
Llama-3.1 zero_shot lstm	0.854 ± 0e0	0.929 ± 0e0	0.883 ± 0e0	0.944 ± 0e0	0.839 ± 0e0	0.931 ± 0e0
Llama-3.1 zero_shot mlp	0.867 ± 0e0	0.929 ± 0e0	0.891 ± 0e0	0.944 ± 0e0	0.847 ± 0e0	0.931 ± 0e0
Llama-3.1 zero_shot patchtsmixer	0.856 ± 0e0	0.927 ± 0e0	0.884 ± 0e0	0.945 ± 0e0	0.841 ± 0e0	0.933 ± 0e0
Llama-3.1 zero_shot timemixer	0.852 ± 0e0	0.927 ± 0e0	0.875 ± 0e0	0.944 ± 0e0	0.826 ± 0e0	0.936 ± 0e0
TFM lstm	0.85 ± 8.96e − 4	0.924 ± 1.793e − 3	0.877 ± 1.198e − 3	0.939 ± 9.594e − 4	0.83 ± 2.181e − 3	0.923 ± 2.089e − 3
TFM mlp	0.852 ± 1.475e − 3	0.924 ± 1.063e − 3	0.879 ± 8.396e − 4	0.937 ± 1.356e − 3	0.831 ± 1.002e − 3	0.925 ± 7.676e − 4
TFM patchtsmixer	0.842 ± 2.252e − 3	0.925 ± 2.263e − 3	0.869 ± 2.4e − 3	0.94 ± 1.253e − 3	0.82 ± 5.108e − 3	0.925 ± 1.576e − 3
TFM timemixer	0.834 ± 1.109e − 2	0.927 ± 3.234e − 3	0.857 ± 7.105e − 3	0.941 ± 2.644e − 3	0.808 ± 4.268e − 3	0.933 ± 1.69e − 3
TSDE lstm	0.869 ± 1.638e − 3	0.924 ± 2.297e − 3	0.902 ± 4.602e − 4	0.94 ± 3.792e − 4	0.857 ± 2.576e − 3	0.924 ± 2.547e − 3
TSDE mlp	0.875 ± 4.566e − 4	0.928 ± 5.67e − 4	0.908 ± 1.662e − 4	0.941 ± 1.076e − 3	0.874 ± 1.086e − 3	0.93 ± 5.558e − 4
TSDE patchtsmixer	0.741 ± 5e − 6	0.902 ± 6.5e − 5	0.763 ± 1.915e − 3	0.899 ± 2.885e − 3	0.733 ± 5e − 6	0.870 ± 0e0
TSDE timemixer	0.744 ± 3.397e − 3	0.913 ± 1.622e − 3	0.763 ± 3.499e − 4	0.919 ± 1.35e − 4	0.734 ± 7.172e − 4	0.939 ± 3.318e − 3
gemini-2.0-flash CoT +TFM lstm	0.88 ± 2.886e − 3	0.924 ± 8.207e − 4	0.903 ± 1.52e − 3	0.935 ± 1.495e − 3	0.85 ± 2.026e − 3	0.913 ± 1.696e − 3
gemini-2.0-flash CoT +TFM mlp	0.886 ± 2.639e − 3	0.928 ± 5.889e − 4	0.906 ± 9.991e − 4	0.941 ± 3.564e − 4	0.861 ± 1.612e − 3	0.921 ± 7.276e − 4
gemini-2.0-flash CoT +TFM patchtsmixer	0.864 ± 1.437e − 3	0.919 ± 8.718e − 4	0.897 ± 9.968e − 4	0.94 ± 9.819e − 4	0.84 ± 1.2e − 2	0.919 ± 2.696e − 3
gemini-2.0-flash CoT +TFM timemixer	0.859 ± 4.816e − 3	0.924 ± 9.548e − 4	0.878 ± 5.403e − 3	0.936 ± 2.136e − 3	0.814 ± 8.623e − 3	0.929 ± 4.58e − 4
gemini-2.0-flash CoT lstm	0.854 ± 6.322e − 3	0.923 ± 2.85e − 3	0.893 ± 3.543e − 3	0.944 ± 1.556e − 3	0.84 ± 6.074e − 3	0.923 ± 5.77e − 3
gemini-2.0-flash CoT mlp	0.865 ± 0e0	0.924 ± 0e0	0.896 ± 0e0	0.944 ± 0e0	0.854 ± 0e0	0.921 ± 0e0
gemini-2.0-flash CoT patchtsmixer	0.849 ± 0e0	0.929 ± 0e0	0.891 ± 0e0	0.946 ± 0e0	0.845 ± 0e0	0.924 ± 0e0
gemini-2.0-flash CoT timemixer	0.844 ± 0e0	0.923 ± 0e0	0.878 ± 0e0	0.947 ± 0e0	0.828 ± 0e0	0.928 ± 0e0
gemini-2.0-flash ICD +TFM lstm	0.882 ± 4.025e − 3	0.923 ± 1.314e − 3	0.9 ± 5.27e − 3	0.935 ± 1.442e − 3	0.85 ± 1.027e − 3	0.917 ± 1.417e − 3
gemini-2.0-flash ICD +TFM mlp	0.88 ± 2.633e − 3	0.926 ± 1.237e − 3	0.905 ± 2.309e − 3	0.941 ± 6.612e − 4	0.859 ± 1.837e − 3	0.923 ± 1.311e − 4
gemini-2.0-flash ICD +TFM patchtsmixer	0.831 ± 2.852e − 2	0.923 ± 5.748e − 3	0.894 ± 2.291e − 3	0.94 ± 9.271e − 4	0.844 ± 6.047e − 3	0.92 ± 1.276e − 3
gemini-2.0-flash ICD +TFM timemixer	0.864 ± 2.421e − 3	0.922 ± 4.196e − 4	0.876 ± 4.371e − 3	0.936 ± 7.193e − 4	0.824 ± 3.774e − 3	0.926 ± 1.046e − 3
gemini-2.0-flash ICD lstm	0.851 ± 7.425e − 3	0.923 ± 1.202e − 3	0.887 ± 7.99e − 4	0.945 ± 7.212e − 4	0.838 ± 7.34e − 3	0.925 ± 4.172e − 3
gemini-2.0-flash ICD mlp	0.863 ± 0e0	0.924 ± 0e0	0.895 ± 0e0	0.944 ± 0e0	0.853 ± 0e0	0.924 ± 0e0
gemini-2.0-flash ICD patchtsmixer	0.847 ± 0e0	0.925 ± 0e0	0.888 ± 0e0	0.945 ± 0e0	0.841 ± 0e0	0.928 ± 0e0
gemini-2.0-flash ICD timemixer	0.843 ± 0e0	0.927 ± 0e0	0.875 ± 0e0	0.946 ± 0e0	0.83 ± 0e0	0.93 ± 0e0
gemini-2.0-flash Trend +TFM lstm	0.88 ± 1.391e − 3	0.923 ± 1.364e − 3	0.905 ± 2.185e − 3	0.936 ± 9.111e − 4	0.867 ± 3.576e − 4	0.914 ± 1.105e − 3
gemini-2.0-flash Trend +TFM mlp	0.883 ± 1.555e − 3	0.928 ± 7e − 5	0.909 ± 9.385e − 4	0.94 ± 2.974e − 4	0.875 ± 1.33e − 3	0.925 ± 1.106e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.862 ± 5.485e − 3	0.918 ± 1.886e − 3	0.897 ± 1.479e − 3	0.94 ± 2.685e − 4	0.86 ± 2.408e − 3	0.918 ± 3.602e − 4
gemini-2.0-flash Trend +TFM timemixer	0.864 ± 3.814e − 3	0.922 ± 2.445e − 3	0.885 ± 4.067e − 3	0.937 ± 1.064e − 3	0.837 ± 3.347e − 3	0.923 ± 1.102e − 3
gemini-2.0-flash Trend lstm	0.859 ± 9.192e − 3	0.922 ± 2.001e − 3	0.897 ± 3.486e − 3	0.94 ± 1.881e − 3	0.858 ± 7.637e − 3	0.921 ± 1.57e − 3
gemini-2.0-flash Trend mlp	0.869 ± 0e0	0.923 ± 0e0	0.904 ± 0e0	0.94 ± 0e0	0.87 ± 0e0	0.924 ± 0e0
gemini-2.0-flash Trend patchtsmixer	0.854 ± 0e0	0.92 ± 0e0	0.895 ± 0e0	0.942 ± 0e0	0.861 ± 0e0	0.923 ± 0e0
gemini-2.0-flash Trend timemixer	0.847 ± 0e0	0.921 ± 0e0	0.884 ± 0e0	0.941 ± 0e0	0.845 ± 0e0	0.927 ± 0e0
gemini-2.0-flash zero_shot +TFM lstm	0.888 ± 6.233e − 3	0.925 ± 3.482e − 3	0.9 ± 4.405e − 3	0.938 ± 9.188e − 4	0.864 ± 1.14e − 3	0.915 ± 1.201e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.886 ± 2.223e − 3	0.927 ± 6.502e − 4	0.906 ± 2.027e − 3	0.942 ± 6.099e − 4	0.87 ± 5.372e − 4	0.924 ± 1.206e − 4
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.87 ± 5.808e − 3	0.918 ± 3.325e − 3	0.897 ± 1.201e − 3	0.941 ± 5.901e − 4	0.85 ± 3.66e − 3	0.919 ± 9.218e − 4
gemini-2.0-flash zero_shot +TFM timemixer	0.864 ± 1.514e − 3	0.921 ± 2.516e − 3	0.879 ± 1.897e − 3	0.938 ± 2.642e − 3	0.828 ± 2.471e − 3	0.925 ± 6.238e − 4
gemini-2.0-flash zero_shot lstm	0.864 ± 1.317e − 2	0.926 ± 1.541e − 3	0.893 ± 7.425e − 4	0.942 ± 4.455e − 4	0.846 ± 9.744e − 3	0.924 ± 2.984e − 3
gemini-2.0-flash zero_shot mlp	0.874 ± 0e0	0.928 ± 0e0	0.9 ± 0e0	0.942 ± 0e0	0.861 ± 0e0	0.926 ± 0e0
gemini-2.0-flash zero_shot patchtsmixer	0.855 ± 0e0	0.926 ± 0e0	0.894 ± 0e0	0.942 ± 0e0	0.848 ± 0e0	0.927 ± 0e0
gemini-2.0-flash zero_shot timemixer	0.853 ± 0e0	0.923 ± 0e0	0.881 ± 0e0	0.944 ± 0e0	0.834 ± 0e0	0.93 ± 0e0
interp lstm	0.862 ± 4.524e − 4	0.863 ± 7.691e − 3	0.897 ± 1.071e − 3	0.905 ± 6.778e − 3	0.866 ± 1.449e − 3	0.872 ± 7.24e − 3
interp mlp	0.864 ± 3.877e − 4	0.867 ± 2.024e − 3	0.898 ± 4.303e − 4	0.908 ± 1.671e − 3	0.872 ± 6.682e − 4	0.873 ± 2.105e − 3
interp patchtsmixer	0.86 ± 4.162e − 3	0.853 ± 1.984e − 2	0.891 ± 1.864e − 3	0.893 ± 6.767e − 3	0.859 ± 3.309e − 3	0.847 ± 1.608e − 2
interp timemixer	0.864 ± 1.179e − 3	0.868 ± 4.322e − 3	0.895 ± 1.871e − 4	0.907 ± 1.987e − 3	0.865 ± 2.192e − 3	0.864 ± 1.057e − 2
mean lstm	0.857 ± 4.039e − 4	0.859 ± 4.395e − 3	0.889 ± 5.677e − 4	0.893 ± 4.161e − 3	0.84 ± 1.046e − 3	0.833 ± 6.055e − 3
mean mlp	0.852 ± 7.397e − 4	0.853 ± 4.183e − 3	0.888 ± 2.311e − 4	0.894 ± 1.562e − 3	0.846 ± 1.859e − 3	0.844 ± 5.339e − 3
mean patchtsmixer	0.86 ± 1.873e − 3	0.857 ± 1.018e − 2	0.89 ± 3.296e − 3	0.886 ± 9.38e − 3	0.858 ± 3.78e − 3	0.847 ± 1.71e − 2
mean timemixer	0.867 ± 1.337e − 3	0.871 ± 4.503e − 3	0.9 ± 1.135e − 3	0.903 ± 5.124e − 3	0.869 ± 2.332e − 3	0.87 ± 6.14e − 3
No_sum_CoT CoT +TFM lstm	0.899 ± 1.882e − 3	0.931 ± 9.36e − 4	0.913 ± 1.828e − 3	0.938 ± 7.171e − 4	0.887 ± 4.298e − 3	0.926 ± 1.93e − 3
No_sum_CoT CoT +TFM mlp	0.898 ± 1.646e − 4	0.935 ± 6.165e − 4	0.917 ± 1.911e − 3	0.942 ± 7.07e − 4	0.888 ± 3.313e − 3	0.933 ± 6.611e − 4
No_sum_CoT CoT +TFM patchtsmixer	0.88 ± 4.753e − 3	0.923 ± 1.892e − 3	0.907 ± 1.434e − 3	0.942 ± 6.495e − 4	0.871 ± 2.433e − 3	0.924 ± 1.167e − 3
No_sum_CoT CoT +TFM timemixer	0.874 ± 5.09e − 3	0.925 ± 5.211e − 3	0.895 ± 4.736e − 3	0.941 ± 5.895e − 4	0.855 ± 2.125e − 3	0.924 ± 1.743e − 3
No_sum_CoT lstm	0.878 ± 3.551e − 4	0.929 ± 9.75e − 4	0.905 ± 9.539e − 5	0.942 ± 3.9e − 4	0.873 ± 2.108e − 4	0.932 ± 7.855e − 4
No_sum_CoT mlp	0.888 ± 0e0	0.933 ± 0e0	0.914 ± 0e0	0.942 ± 0e0	0.884 ± 0e0	0.933 ± 0e0
No_sum_CoT patchtsmixer	0.877 ± 1.922e − 4	0.931 ± 8.355e − 4	0.906 ± 3.812e − 4	0.944 ± 6.991e − 4	0.876 ± 4.941e − 4	0.934 ± 1.795e − 3
No_sum_CoT timemixer	0.871 ± 1.175e − 3	0.931 ± 1.531e − 4	0.893 ± 1.317e − 3	0.944 ± 1.362e − 3	0.867 ± 5.472e − 4	0.926 ± 1.09e − 3
No_sum_ICD ICD +TFM lstm	0.899 ± 6.612e − 4	0.931 ± 3.923e − 4	0.916 ± 1.685e − 3	0.939 ± 5.954e − 4	0.881 ± 5.792e − 3	0.924 ± 3.529e − 3
No_sum_ICD ICD +TFM mlp	0.899 ± 2.172e − 3	0.935 ± 1.401e − 3	0.916 ± 1.139e − 3	0.943 ± 3.055e − 4	0.885 ± 2.285e − 3	0.932 ± 1.292e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.882 ± 3.476e − 3	0.923 ± 1.524e − 3	0.906 ± 1.865e − 3	0.941 ± 5.905e − 4	0.871 ± 2.455e − 3	0.925 ± 1.528e − 3
No_sum_ICD ICD +TFM timemixer	0.877 ± 2.363e − 3	0.925 ± 3.242e − 3	0.898 ± 2.777e − 3	0.941 ± 6.577e − 4	0.855 ± 2.207e − 3	0.923 ± 6.616e − 4
No_sum_ICD lstm	0.877 ± 4.917e − 4	0.93 ± 7.455e − 4	0.903 ± 4.43e − 4	0.944 ± 6.883e − 4	0.871 ± 2.558e − 4	0.931 ± 1.154e − 3
No_sum_ICD mlp	0.886 ± 0e0	0.934 ± 0e0	0.911 ± 0e0	0.942 ± 0e0	0.88 ± 0e0	0.933 ± 0e0
No_sum_ICD patchtsmixer	0.875 ± 7.904e − 4	0.93 ± 1.542e − 3	0.904 ± 7.81e − 5	0.943 ± 8.528e − 4	0.871 ± 6.478e − 4	0.935 ± 1.466e − 3
No_sum_ICD timemixer	0.869 ± 7.434e − 4	0.932 ± 3.365e − 4	0.9 ± 3.281e − 4	0.943 ± 1.146e − 3	0.862 ± 1.177e − 3	0.928 ± 1.409e − 3
No_sum_Trend Trend +TFM lstm	0.897 ± 2.179e − 3	0.93 ± 8.698e − 4	0.915 ± 9.572e − 4	0.937 ± 1.264e − 3	0.885 ± 1.388e − 3	0.924 ± 1.125e − 3
No_sum_Trend Trend +TFM mlp	0.901 ± 5.023e − 3	0.934 ± 1.867e − 3	0.917 ± 2.971e − 3	0.943 ± 2.888e − 4	0.887 ± 6e − 5	0.933 ± 4.277e − 4
No_sum_Trend Trend +TFM patchtsmixer	0.881 ± 5.503e − 3	0.923 ± 1.202e − 3	0.907 ± 8.208e − 4	0.941 ± 8.731e − 4	0.877 ± 2.42e − 3	0.925 ± 1.415e − 3
No_sum_Trend Trend +TFM timemixer	0.875 ± 6.816e − 3	0.924 ± 2.757e − 3	0.892 ± 3.244e − 3	0.94 ± 8.564e − 4	0.856 ± 1.993e − 3	0.924 ± 1.675e − 3

Table 30: In-distribution results - Lab. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro	f1_micro	recall_micro
No_sum_Trend lstm	0.878 ± 2.658e − 4	0.928 ± 6.463e − 4	0.906 ± 2.312e − 4	0.943 ± 6.671e − 4	0.875 ± 1.3e − 4	0.932 ± 1.242e − 4
No_sum_Trend mlp	0.895 ± 0e0	0.933 ± 0e0	0.914 ± 0e0	0.942 ± 0e0	0.883 ± 0e0	0.933 ± 0e0
No_sum_Trend patchtsmixer	0.879 ± 7.515e − 4	0.929 ± 1.331e − 3	0.905 ± 2.403e − 4	0.942 ± 7.504e − 4	0.875 ± 1.472e − 3	0.936 ± 2.444e − 4
No_sum_Trend timemixer	0.873 ± 3.395e − 4	0.93 ± 4.888e − 4	0.896 ± 8.228e − 4	0.942 ± 6.2e − 4	0.864 ± 3.47e − 4	0.926 ± 1.854e − 3
medgemma CoT +TFM lstm	0.889 ± 5.13e − 3	0.926 ± 4.828e − 4	0.904 ± 2.248e − 3	0.936 ± 1.293e − 3	0.861 ± 2.565e − 3	0.916 ± 1.212e − 3
medgemma CoT +TFM mlp	0.889 ± 2.26e − 3	0.93 ± 4.419e − 4	0.909 ± 1.791e − 3	0.942 ± 3.538e − 4	0.869 ± 7.267e − 4	0.927 ± 7.617e − 4
medgemma CoT +TFM patchtsmixer	0.843 ± 2.567e − 2	0.921 ± 4.432e − 3	0.899 ± 2.595e − 3	0.941 ± 1.213e − 3	0.844 ± 1.187e − 2	0.922 ± 8.976e − 4
medgemma CoT +TFM timemixer	0.87 ± 2.113e − 3	0.923 ± 7.391e − 4	0.883 ± 7.948e − 3	0.938 ± 1.643e − 3	0.833 ± 2.474e − 3	0.926 ± 8.26e − 4
medgemma CoT lstm	0.868 ± 8.252e − 3	0.925 ± 2.256e − 3	0.896 ± 4.759e − 3	0.944 ± 2.836e − 3	0.851 ± 0e0	0.928 ± 0e0
medgemma CoT mlp	0.879 ± 0e0	0.929 ± 0e0	0.903 ± 0e0	0.943 ± 0e0	0.864 ± 0e0	0.928 ± 0e0
medgemma CoT patchtsmixer	0.866 ± 0e0	0.925 ± 0e0	0.895 ± 0e0	0.946 ± 0e0	0.855 ± 0e0	0.932 ± 0e0
medgemma CoT timemixer	0.86 ± 0e0	0.927 ± 0e0	0.888 ± 0e0	0.946 ± 0e0	0.84 ± 0e0	0.927 ± 0e0
medgemma ICD +TFM lstm	0.877 ± 3.05e − 3	0.923 ± 1.422e − 3	0.895 ± 4.337e − 3	0.937 ± 1.362e − 3	0.844 ± 1.449e − 3	0.912 ± 1.579e − 3
medgemma ICD +TFM mlp	0.884 ± 3.329e − 3	0.928 ± 1.335e − 3	0.906 ± 2.99e − 3	0.941 ± 1.008e − 3	0.858 ± 3.321e − 3	0.922 ± 2.843e − 4
medgemma ICD +TFM patchtsmixer	0.863 ± 8.107e − 3	0.918 ± 1.423e − 3	0.894 ± 1.985e − 3	0.942 ± 3.724e − 4	0.832 ± 1.02e − 2	0.923 ± 3.201e − 3
medgemma ICD +TFM timemixer	0.862 ± 2.257e − 3	0.921 ± 8.469e − 4	0.876 ± 8.269e − 3	0.938 ± 2.769e − 3	0.811 ± 1.756e − 3	0.927 ± 8.819e − 4
medgemma ICD lstm	0.856 ± 8.966e − 3	0.924 ± 1.209e − 3	0.885 ± 3.698e − 3	0.946 ± 4.313e − 4	0.832 ± 7.7e − 3	0.926 ± 3.097e − 3
medgemma ICD mlp	0.868 ± 0e0	0.926 ± 0e0	0.895 ± 0e0	0.944 ± 0e0	0.843 ± 0e0	0.925 ± 0e0
medgemma ICD patchtsmixer	0.852 ± 0e0	0.92 ± 0e0	0.888 ± 0e0	0.946 ± 0e0	0.832 ± 0e0	0.931 ± 0e0
medgemma ICD timemixer	0.85 ± 0e0	0.922 ± 0e0	0.876 ± 0e0	0.948 ± 0e0	0.808 ± 0e0	0.936 ± 0e0
medgemma Trend +TFM lstm	0.88 ± 4.772e − 3	0.924 ± 5.742e − 4	0.903 ± 8.193e − 4	0.936 ± 5.575e − 4	0.861 ± 1.843e − 3	0.917 ± 2.003e − 3
medgemma Trend +TFM mlp	0.885 ± 1.961e − 3	0.929 ± 2.15e − 4	0.908 ± 3.774e − 4	0.94 ± 4.464e − 4	0.869 ± 1.008e − 3	0.925 ± 4.424e − 4
medgemma Trend +TFM patchtsmixer	0.864 ± 1.666e − 3	0.919 ± 3.332e − 4	0.896 ± 2.969e − 3	0.939 ± 1.372e − 3	0.849 ± 4.349e − 3	0.919 ± 6.379e − 4
medgemma Trend +TFM timemixer	0.861 ± 4.282e − 3	0.923 ± 1.442e − 3	0.886 ± 6.317e − 4	0.94 ± 1.53e − 3	0.825 ± 1.56e − 2	0.925 ± 3.355e − 3
medgemma Trend lstm	0.861 ± 6.64e − 3	0.925 ± 2.885e − 3	0.896 ± 3.507e − 3	0.94 ± 2.503e − 3	0.85 ± 6.937e − 3	0.924 ± 7.354e − 4
medgemma Trend mlp	0.871 ± 0e0	0.925 ± 0e0	0.9 ± 0e0	0.941 ± 0e0	0.859 ± 0e0	0.926 ± 0e0
medgemma Trend patchtsmixer	0.855 ± 0e0	0.924 ± 0e0	0.894 ± 0e0	0.942 ± 0e0	0.851 ± 0e0	0.925 ± 0e0
medgemma Trend timemixer	0.85 ± 0e0	0.927 ± 0e0	0.88 ± 0e0	0.944 ± 0e0	0.837 ± 0e0	0.928 ± 0e0
medgemma zero_shot +TFM lstm	0.886 ± 1.048e − 3	0.923 ± 6.421e − 4	0.902 ± 2.546e − 3	0.937 ± 7.805e − 4	0.853 ± 5.101e − 3	0.915 ± 1.43e − 3
medgemma zero_shot +TFM mlp	0.891 ± 2.723e − 3	0.929 ± 1.725e − 3	0.909 ± 2.35e − 3	0.942 ± 6.171e − 4	0.865 ± 2.928e − 3	0.924 ± 9.504e − 5
medgemma zero_shot +TFM patchtsmixer	0.866 ± 1.713e − 3	0.917 ± 3.412e − 4	0.896 ± 8.122e − 4	0.942 ± 3.317e − 4	0.843 ± 1.467e − 2	0.921 ± 2.999e − 3
medgemma zero_shot +TFM timemixer	0.847 ± 2.482e − 2	0.926 ± 7.219e − 3	0.88 ± 3.121e − 3	0.939 ± 1.32e − 3	0.828 ± 4.5e − 3	0.925 ± 1.935e − 3
medgemma zero_shot lstm	0.867 ± 1.014e − 2	0.925 ± 3.571e − 3	0.893 ± 4.589e − 3	0.944 ± 2.22e − 3	0.847 ± 8.4e − 3	0.924 ± 4.363e − 3
medgemma zero_shot mlp	0.878 ± 0e0	0.927 ± 0e0	0.9 ± 0e0	0.943 ± 0e0	0.858 ± 0e0	0.926 ± 0e0
medgemma zero_shot patchtsmixer	0.858 ± 0e0	0.923 ± 0e0	0.894 ± 0e0	0.945 ± 0e0	0.852 ± 0e0	0.93 ± 0e0
medgemma zero_shot timemixer	0.852 ± 0e0	0.924 ± 0e0	0.882 ± 0e0	0.947 ± 0e0	0.83 ± 0e0	0.931 ± 0e0
No_sum_zero_shot zero_shot +TFM lstm	0.899 ± 1.292e − 3	0.931 ± 1.177e − 3	0.915 ± 2.725e − 3	0.937 ± 3.164e − 4	0.884 ± 1.782e − 3	0.926 ± 1.312e − 3
No_sum_zero_shot zero_shot +TFM mlp	0.899 ± 3.993e − 4	0.935 ± 9.4e − 4	0.917 ± 1.582e − 3	0.942 ± 3.988e − 4	0.888 ± 2.399e − 3	0.933 ± 8.827e − 4
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.883 ± 1.917e − 3	0.923 ± 1.105e − 3	0.907 ± 7.06e − 4	0.941 ± 4.992e − 4	0.875 ± 2.033e − 3	0.925 ± 1.18e − 3
No_sum_zero_shot zero_shot +TFM timemixer	0.874 ± 3.213e − 3	0.926 ± 4.367e − 3	0.892 ± 3.519e − 3	0.941 ± 7.862e − 4	0.853 ± 4.922e − 3	0.923 ± 1.54e − 3
No_sum_zero_shot lstm	0.877 ± 5.33e − 4	0.931 ± 1.281e − 3	0.905 ± 1.286e − 4	0.942 ± 2.501e − 4	0.873 ± 1.518e − 4	0.933 ± 7.503e − 4
No_sum_zero_shot mlp	0.891 ± 0e0	0.934 ± 0e0	0.912 ± 0e0	0.942 ± 0e0	0.881 ± 0e0	0.933 ± 0e0
No_sum_zero_shot patchtsmixer	0.877 ± 5.975e − 4	0.93 ± 4.858e − 4	0.905 ± 2.001e − 4	0.943 ± 9.239e − 4	0.874 ± 8.505e − 5	0.936 ± 1.778e − 3
No_sum_zero_shot timemixer	0.869 ± 5.951e − 4	0.932 ± 5.065e − 4	0.898 ± 8.837e − 4	0.942 ± 4.535e − 4	0.86 ± 7.794e − 4	0.93 ± 1.172e − 3
right lstm	0.87 ± 8.813e − 4	0.881 ± 3.199e − 3	0.899 ± 4.968e − 4	0.908 ± 3.32e − 3	0.87 ± 1.011e − 3	0.882 ± 3.608e − 3
right mlp	0.873 ± 4.173e − 4	0.88 ± 2.961e − 3	0.9 ± 2.41e − 4	0.91 ± 3.03e − 3	0.876 ± 6.405e − 4	0.879 ± 4.391e − 3
right patchtsmixer	0.863 ± 6.473e − 3	0.858 ± 2.645e − 2	0.89 ± 2.222e − 3	0.884 ± 9.048e − 3	0.858 ± 2.851e − 3	0.846 ± 1.492e − 2
right timemixer	0.874 ± 8.987e − 4	0.884 ± 3.664e − 3	0.899 ± 8.577e − 4	0.911 ± 6.399e − 3	0.869 ± 4.859e − 4	0.868 ± 2.748e − 3

Table 31: In-distribution results - Lab. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro
Llama-3.1 CoT +TFM lstm	0.774 ± 6.907e − 4	0.959 ± 3.078e − 3	0.766 ± 1.933e − 3	0.95 ± 5.956e − 4
Llama-3.1 CoT +TFM mlp	0.78 ± 2.159e − 3	0.956 ± 2.175e − 3	0.761 ± 1.398e − 3	0.956 ± 1.442e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.777 ± 2.362e − 3	0.959 ± 4.087e − 3	0.755 ± 1.652e − 3	0.961 ± 2.694e − 3
Llama-3.1 CoT +TFM timemixer	0.773 ± 1.304e − 3	0.962 ± 2.365e − 3	0.775 ± 3.476e − 3	0.94 ± 5.293e − 3
Llama-3.1 CoT lstm	0.786 ± 0e0	0.954 ± 0e0	0.775 ± 0e0	0.952 ± 0e0
Llama-3.1 CoT mlp	0.781 ± 0e0	0.955 ± 0e0	0.765 ± 0e0	0.951 ± 0e0
Llama-3.1 CoT patchtsmixer	0.785 ± 0e0	0.959 ± 0e0	0.775 ± 0e0	0.953 ± 0e0
Llama-3.1 CoT timemixer	0.786 ± 0e0	0.957 ± 0e0	0.78 ± 0e0	0.949 ± 0e0
Llama-3.1 ICD +TFM lstm	0.769 ± 2.365e − 3	0.967 ± 2.211e − 3	0.77 ± 1.332e − 3	0.944 ± 1.466e − 3
Llama-3.1 ICD +TFM mlp	0.776 ± 2.135e − 3	0.96 ± 2.754e − 3	0.761 ± 1.713e − 3	0.955 ± 1.754e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.775 ± 2.289e − 3	0.959 ± 3.655e − 3	0.76 ± 1.61e − 3	0.956 ± 3.691e − 3
Llama-3.1 ICD +TFM timemixer	0.774 ± 7.144e − 4	0.962 ± 3.057e − 4	0.773 ± 1.563e − 3	0.941 ± 1.077e − 3
Llama-3.1 ICD lstm	0.788 ± 0e0	0.948 ± 0e0	0.775 ± 0e0	0.947 ± 0e0
Llama-3.1 ICD mlp	0.788 ± 0e0	0.947 ± 0e0	0.77 ± 0e0	0.945 ± 0e0
Llama-3.1 ICD patchtsmixer	0.789 ± 0e0	0.954 ± 0e0	0.776 ± 0e0	0.95 ± 0e0
Llama-3.1 ICD timemixer	0.787 ± 0e0	0.951 ± 0e0	0.777 ± 0e0	0.947 ± 0e0
Llama-3.1 Trend +TFM lstm	0.773 ± 4.59e − 3	0.962 ± 6.862e − 3	0.771 ± 2.454e − 3	0.946 ± 1.418e − 3
Llama-3.1 Trend +TFM mlp	0.778 ± 1.86e − 3	0.96 ± 1.412e − 3	0.77 ± 2.548e − 3	0.948 ± 1.384e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.778 ± 1.755e − 3	0.957 ± 7.842e − 4	0.762 ± 2.552e − 3	0.956 ± 3.489e − 3
Llama-3.1 Trend +TFM timemixer	0.779 ± 5.526e − 3	0.959 ± 6.212e − 3	0.771 ± 4.247e − 3	0.945 ± 7.923e − 3
Llama-3.1 Trend lstm	0.79 ± 0e0	0.952 ± 0e0	0.77 ± 0e0	0.948 ± 0e0
Llama-3.1 Trend mlp	0.788 ± 0e0	0.949 ± 0e0	0.768 ± 0e0	0.947 ± 0e0
Llama-3.1 Trend patchtsmixer	0.787 ± 0e0	0.956 ± 0e0	0.769 ± 0e0	0.953 ± 0e0
Llama-3.1 Trend timemixer	0.789 ± 0e0	0.953 ± 0e0	0.769 ± 0e0	0.95 ± 0e0
Llama-3.1 zero_shot +TFM lstm	0.774 ± 1.044e − 4	0.962 ± 1.739e − 4	0.77 ± 4.499e − 3	0.946 ± 1.427e − 3
Llama-3.1 zero_shot +TFM mlp	0.779 ± 5.577e − 4	0.959 ± 2.524e − 4	0.766 ± 4.885e − 3	0.957 ± 3.664e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.78 ± 4.661e − 4	0.954 ± 1.319e − 3	0.765 ± 4.887e − 4	0.954 ± 6.074e − 4
Llama-3.1 zero_shot +TFM timemixer	0.78 ± 2.402e − 3	0.957 ± 3.382e − 3	0.769 ± 2.373e − 3	0.948 ± 2.451e − 3
Llama-3.1 zero_shot lstm	0.791 ± 0e0	0.949 ± 0e0	0.781 ± 0e0	0.945 ± 0e0
Llama-3.1 zero_shot mlp	0.791 ± 0e0	0.946 ± 0e0	0.775 ± 0e0	0.945 ± 0e0
Llama-3.1 zero_shot patchtsmixer	0.793 ± 0e0	0.953 ± 0e0	0.78 ± 0e0	0.948 ± 0e0
Llama-3.1 zero_shot timemixer	0.794 ± 0e0	0.949 ± 0e0	0.781 ± 0e0	0.945 ± 0e0
TFM lstm	0.799 ± 2.24e − 3	0.94 ± 2.723e − 3	0.787 ± 1.781e − 3	0.934 ± 1.136e − 3
TFM mlp	0.798 ± 8.308e − 4	0.942 ± 2.808e − 3	0.786 ± 1.065e − 3	0.934 ± 2.918e − 3
TFM patchtsmixer	0.794 ± 1.727e − 3	0.946 ± 4.452e − 4	0.778 ± 1.301e − 3	0.945 ± 1.541e − 3
TFM timemixer	0.802 ± 2.974e − 3	0.943 ± 2.743e − 3	0.782 ± 2.503e − 3	0.941 ± 6.271e − 3
TSDE lstm	0.746 ± 4.069e − 4	0.898 ± 3.86e − 4	0.738 ± 6.399e − 3	0.929 ± 6.829e − 3
TSDE mlp	0.742 ± 9.998e − 4	0.898 ± 5.68e − 4	0.734 ± 4.081e − 3	0.908 ± 1.485e − 3
TSDE patchtsmixer	0.734 ± 0e0	0.909 ± 2.5e − 5	0.721 ± 2.765e − 3	0.899 ± 1.2e − 4
TSDE timemixer	0.734 ± 4.349e − 5	0.889 ± 0e0	0.72 ± 4.75e − 4	0.879 ± 4.7e − 4
gemini-2.0-flash CoT +TFM lstm	0.779 ± 1.157e − 3	0.952 ± 1.509e − 3	0.77 ± 1.024e − 3	0.941 ± 1.253e − 3
gemini-2.0-flash CoT +TFM mlp	0.781 ± 2.019e − 3	0.955 ± 3.218e − 3	0.765 ± 1.35e − 3	0.951 ± 1.08e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.781 ± 1.446e − 3	0.955 ± 2.37e − 3	0.758 ± 7.211e − 4	0.961 ± 1.573e − 3
gemini-2.0-flash CoT +TFM timemixer	0.778 ± 5.45e − 4	0.958 ± 1.473e − 3	0.771 ± 5.743e − 3	0.945 ± 5.898e − 3
gemini-2.0-flash CoT lstm	0.793 ± 1.584e − 3	0.945 ± 3.267e − 3	0.773 ± 5.346e − 3	0.941 ± 4.391e − 3
gemini-2.0-flash CoT mlp	0.792 ± 0e0	0.946 ± 0e0	0.772 ± 0e0	0.945 ± 0e0
gemini-2.0-flash CoT patchtsmixer	0.791 ± 0e0	0.952 ± 0e0	0.774 ± 0e0	0.948 ± 0e0
gemini-2.0-flash CoT timemixer	0.792 ± 0e0	0.95 ± 0e0	0.777 ± 0e0	0.945 ± 0e0
gemini-2.0-flash ICD +TFM lstm	0.775 ± 3.041e − 3	0.96 ± 4.06e − 3	0.77 ± 2.096e − 3	0.945 ± 3.438e − 3
gemini-2.0-flash ICD +TFM mlp	0.777 ± 1.871e − 3	0.96 ± 1.876e − 3	0.764 ± 1.075e − 3	0.954 ± 2.623e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.771 ± 5.129e − 4	0.965 ± 1.344e − 3	0.759 ± 3.777e − 3	0.962 ± 5.819e − 3
gemini-2.0-flash ICD +TFM timemixer	0.771 ± 2.899e − 3	0.966 ± 5.001e − 3	0.772 ± 7.667e − 3	0.949 ± 8.648e − 3
gemini-2.0-flash ICD lstm	0.772 ± 7.707e − 4	0.974 ± 8.825e − 3	0.772 ± 1.838e − 4	0.948 ± 1.344e − 4
gemini-2.0-flash ICD mlp	0.772 ± 0e0	0.965 ± 0e0	0.772 ± 0e0	0.946 ± 0e0
gemini-2.0-flash ICD patchtsmixer	0.777 ± 0e0	0.969 ± 0e0	0.772 ± 0e0	0.953 ± 0e0
gemini-2.0-flash ICD timemixer	0.772 ± 0e0	0.975 ± 0e0	0.776 ± 0e0	0.949 ± 0e0
gemini-2.0-flash Trend +TFM lstm	0.784 ± 4.197e − 3	0.942 ± 7.127e − 3	0.776 ± 5.278e − 4	0.94 ± 1.476e − 3
gemini-2.0-flash Trend +TFM mlp	0.789 ± 2.121e − 3	0.947 ± 2.406e − 3	0.773 ± 6.574e − 4	0.944 ± 8.593e − 4
gemini-2.0-flash Trend +TFM patchtsmixer	0.789 ± 1.383e − 3	0.944 ± 1.347e − 3	0.765 ± 2.892e − 3	0.951 ± 3.808e − 3
gemini-2.0-flash Trend +TFM timemixer	0.781 ± 1.092e − 3	0.954 ± 2.578e − 3	0.773 ± 1.482e − 3	0.946 ± 4.805e − 3
gemini-2.0-flash Trend lstm	0.801 ± 3.111e − 3	0.938 ± 3.458e − 3	0.781 ± 2.178e − 3	0.935 ± 6.838e − 3
gemini-2.0-flash Trend mlp	0.799 ± 0e0	0.939 ± 0e0	0.777 ± 0e0	0.941 ± 0e0
gemini-2.0-flash Trend patchtsmixer	0.802 ± 0e0	0.945 ± 0e0	0.783 ± 0e0	0.94 ± 0e0
gemini-2.0-flash Trend timemixer	0.8 ± 0e0	0.942 ± 0e0	0.785 ± 0e0	0.938 ± 0e0
gemini-2.0-flash zero_shot +TFM lstm	0.778 ± 2.447e − 3	0.952 ± 3.203e − 3	0.77 ± 1.228e − 3	0.945 ± 2.536e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.778 ± 1.966e − 3	0.957 ± 9.836e − 4	0.761 ± 3.949e − 3	0.956 ± 3.297e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.775 ± 1.757e − 3	0.959 ± 9.166e − 4	0.758 ± 2.899e − 3	0.962 ± 4.923e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.774 ± 3.55e − 3	0.964 ± 6.95e − 3	0.768 ± 3.997e − 3	0.952 ± 4.818e − 3
gemini-2.0-flash zero_shot lstm	0.793 ± 7.071e − 4	0.943 ± 5.19e − 3	0.765 ± 3.309e − 3	0.955 ± 9.115e − 3
gemini-2.0-flash zero_shot mlp	0.794 ± 0e0	0.942 ± 0e0	0.762 ± 0e0	0.957 ± 0e0
gemini-2.0-flash zero_shot patchtsmixer	0.795 ± 0e0	0.953 ± 0e0	0.767 ± 0e0	0.954 ± 0e0
gemini-2.0-flash zero_shot timemixer	0.787 ± 0e0	0.957 ± 0e0	0.767 ± 0e0	0.957 ± 0e0
interp lstm	0.708 ± 1.932e − 3	0.886 ± 6.243e − 3	0.657 ± 9.491e − 3	0.813 ± 2.445e − 2
interp mlp	0.733 ± 1.943e − 3	0.955 ± 3.448e − 3	0.702 ± 7.303e − 3	0.911 ± 2.16e − 2
interp patchtsmixer	0.702 ± 9.913e − 3	0.808 ± 1.83e − 2	0.678 ± 7.828e − 3	0.8 ± 1.262e − 2
interp timemixer	0.641 ± 8.843e − 3	0.654 ± 1.769e − 2	0.638 ± 3.367e − 3	0.733 ± 8.544e − 3
mean lstm	0.649 ± 4.088e − 3	0.8 ± 1.151e − 2	0.623 ± 9.661e − 3	0.726 ± 1.935e − 2
mean mlp	0.691 ± 3.43e − 3	0.885 ± 6.064e − 3	0.642 ± 1.942e − 3	0.772 ± 2.956e − 3
mean patchtsmixer	0.686 ± 1.272e − 2	0.775 ± 2.779e − 2	0.679 ± 7.18e − 3	0.81 ± 1.601e − 2
mean timemixer	0.667 ± 5.537e − 3	0.708 ± 1.183e − 2	0.649 ± 6.057e − 3	0.754 ± 1.358e − 2
No_sum_CoT CoT +TFM lstm	0.795 ± 4.985e − 3	0.936 ± 5.18e − 3	0.775 ± 1.182e − 3	0.943 ± 2.763e − 3
No_sum_CoT CoT +TFM mlp	0.793 ± 2.168e − 3	0.951 ± 5.7e − 4	0.771 ± 1.56e − 3	0.951 ± 2.646e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.788 ± 1.265e − 3	0.947 ± 7.808e − 4	0.769 ± 1.909e − 3	0.945 ± 2.385e − 3
No_sum_CoT CoT +TFM timemixer	0.794 ± 8.153e − 4	0.94 ± 3.354e − 3	0.772 ± 1.13e − 3	0.948 ± 2.62e − 3
No_sum_CoT lstm	0.806 ± 6.248e − 4	0.948 ± 4.444e − 4	0.786 ± 8.689e − 4	0.951 ± 5.341e − 4
No_sum_CoT mlp	0.803 ± 0e0	0.948 ± 0e0	0.778 ± 0e0	0.948 ± 0e0
No_sum_CoT patchtsmixer	0.805 ± 1.274e − 4	0.95 ± 1.266e − 3	0.781 ± 4.933e − 4	0.95 ± 1.188e − 3
No_sum_CoT timemixer	0.807 ± 9.986e − 4	0.943 ± 5.289e − 4	0.796 ± 1.695e − 3	0.938 ± 4.941e − 4
No_sum_ICD ICD +TFM lstm	0.798 ± 3.219e − 3	0.937 ± 1.662e − 3	0.783 ± 7.751e − 4	0.938 ± 1.566e − 3
No_sum_ICD ICD +TFM mlp	0.793 ± 6.338e − 4	0.951 ± 1.252e − 3	0.771 ± 2.683e − 3	0.951 ± 2.63e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.787 ± 1.883e − 3	0.95 ± 2.141e − 3	0.773 ± 8.41e − 4	0.942 ± 1.64e − 3
No_sum_ICD ICD +TFM timemixer	0.792 ± 3.412e − 4	0.945 ± 1.421e − 3	0.777 ± 1.567e − 3	0.941 ± 5.876e − 3
No_sum_ICD lstm	0.805 ± 3.356e − 4	0.942 ± 6.258e − 4	0.79 ± 1.22e − 3	0.941 ± 2.926e − 4
No_sum_ICD mlp	0.802 ± 0e0	0.948 ± 0e0	0.782 ± 0e0	0.945 ± 0e0
No_sum_ICD patchtsmixer	0.805 ± 9.485e − 4	0.953 ± 8.13e − 4	0.785 ± 1.948e − 3	0.947 ± 7.304e − 4
No_sum_ICD timemixer	0.811 ± 3.35e − 4	0.942 ± 3.166e − 4	0.794 ± 3.853e − 4	0.936 ± 7.954e − 4
No_sum_Trend Trend +TFM lstm	0.8 ± 7.995e − 4	0.936 ± 1.19e − 3	0.776 ± 2.982e − 3	0.939 ± 6.048e − 3
No_sum_Trend Trend +TFM mlp	0.791 ± 2.368e − 3	0.951 ± 1.239e − 3	0.771 ± 4.85e − 3	0.947 ± 2.446e − 3
No_sum_Trend Trend +TFM patchtsmixer	0.789 ± 1.329e − 3	0.946 ± 2.136e − 3	0.769 ± 1.466e − 3	0.945 ± 1.707e − 3
No_sum_Trend Trend +TFM timemixer	0.791 ± 3.189e − 3	0.943 ± 2.13e − 3	0.777 ± 4.916e − 3	0.939 ± 6.115e − 3

Table 32: Cross-site transfer results - Lab. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	f1_micro	recall_micro	f1_micro	recall_micro
No_sum_Trend lstm	0.805 ± 8.298e − 4	0.944 ± 1.776e − 3	0.78 ± 3.98e − 4	0.953 ± 5.442e − 4
No_sum_Trend mlp	0.8 ± 0e0	0.946 ± 0e0	0.779 ± 0e0	0.946 ± 0e0
No_sum_Trend patchtsmixer	0.807 ± 7.472e − 4	0.95 ± 9.666e − 4	0.779 ± 2.562e − 3	0.948 ± 1.084e − 3
No_sum_Trend timemixer	0.808 ± 1.079e − 3	0.941 ± 4.553e − 4	0.789 ± 7.168e − 4	0.937 ± 1.581e − 3
medgemma CoT +TFM lstm	0.779 ± 3.345e − 3	0.955 ± 5.752e − 3	0.775 ± 1.652e − 3	0.94 ± 2.151e − 3
medgemma CoT +TFM mlp	0.783 ± 4.335e − 4	0.957 ± 7.516e − 4	0.771 ± 1.904e − 3	0.949 ± 2.446e − 3
medgemma CoT +TFM patchtsmixer	0.778 ± 4.118e − 3	0.958 ± 2.585e − 3	0.766 ± 2.887e − 3	0.952 ± 4.626e − 3
medgemma CoT +TFM timemixer	0.78 ± 1.849e − 3	0.956 ± 3.352e − 3	0.772 ± 5.854e − 3	0.944 ± 8.061e − 3
medgemma CoT lstm	0.797 ± 2.765e − 3	0.946 ± 1.407e − 3	0.774 ± 2.348e − 3	0.951 ± 9.214e − 3
medgemma CoT mlp	0.797 ± 0e0	0.947 ± 0e0	0.774 ± 0e0	0.949 ± 0e0
medgemma CoT patchtsmixer	0.799 ± 0e0	0.954 ± 0e0	0.774 ± 0e0	0.949 ± 0e0
medgemma CoT timemixer	0.796 ± 0e0	0.952 ± 0e0	0.776 ± 0e0	0.946 ± 0e0
medgemma ICD +TFM lstm	0.77 ± 8.732e − 4	0.966 ± 2.587e − 3	0.772 ± 1.314e − 3	0.944 ± 2.011e − 3
medgemma ICD +TFM mlp	0.774 ± 2.889e − 3	0.963 ± 2.689e − 3	0.764 ± 7.353e − 4	0.957 ± 1.034e − 3
medgemma ICD +TFM patchtsmixer	0.771 ± 8.649e − 4	0.964 ± 1.031e − 3	0.76 ± 3.556e − 3	0.958 ± 4.775e − 3
medgemma ICD +TFM timemixer	0.77 ± 6.629e − 4	0.968 ± 8.445e − 4	0.771 ± 2.011e − 3	0.949 ± 1.917e − 3
medgemma ICD lstm	0.791 ± 3.62e − 3	0.946 ± 1.039e − 3	0.777 ± 5.636e − 3	0.945 ± 4.801e − 3
medgemma ICD mlp	0.789 ± 0e0	0.95 ± 0e0	0.771 ± 0e0	0.95 ± 0e0
medgemma ICD patchtsmixer	0.788 ± 0e0	0.957 ± 0e0	0.775 ± 0e0	0.952 ± 0e0
medgemma ICD timemixer	0.786 ± 0e0	0.958 ± 0e0	0.776 ± 0e0	0.949 ± 0e0
medgemma Trend +TFM lstm	0.779 ± 5.745e − 3	0.954 ± 1.007e − 2	0.773 ± 4.056e − 3	0.94 ± 2.083e − 3
medgemma Trend +TFM mlp	0.78 ± 1.057e − 3	0.958 ± 1.437e − 3	0.773 ± 1.767e − 3	0.944 ± 8.65e − 4
medgemma Trend +TFM patchtsmixer	0.784 ± 6.558e − 4	0.951 ± 1.35e − 4	0.764 ± 1.626e − 3	0.952 ± 3.448e − 3
medgemma Trend +TFM timemixer	0.775 ± 5.554e − 4	0.963 ± 1.068e − 3	0.77 ± 1.111e − 3	0.947 ± 2.364e − 4
medgemma Trend lstm	0.797 ± 2.411e − 3	0.942 ± 2.284e − 3	0.784 ± 6.873e − 3	0.938 ± 1.287e − 3
medgemma Trend mlp	0.793 ± 0e0	0.946 ± 0e0	0.781 ± 0e0	0.936 ± 0e0
medgemma Trend patchtsmixer	0.796 ± 0e0	0.951 ± 0e0	0.785 ± 0e0	0.941 ± 0e0
medgemma Trend timemixer	0.796 ± 0e0	0.95 ± 0e0	0.787 ± 0e0	0.94 ± 0e0
medgemma zero_shot +TFM lstm	0.772 ± 2.638e − 3	0.964 ± 5.331e − 3	0.775 ± 3.988e − 3	0.942 ± 3.711e − 3
medgemma zero_shot +TFM mlp	0.777 ± 9.697e − 4	0.962 ± 1.42e − 3	0.769 ± 1.61e − 3	0.953 ± 1.657e − 3
medgemma zero_shot +TFM patchtsmixer	0.773 ± 3.995e − 4	0.963 ± 5.138e − 4	0.761 ± 1.3e − 3	0.96 ± 2.928e − 4
medgemma zero_shot +TFM timemixer	0.774 ± 6.375e − 3	0.961 ± 7.042e − 3	0.768 ± 3.265e − 3	0.952 ± 2.319e − 3
medgemma zero_shot lstm	0.793 ± 2.093e − 3	0.947 ± 7.361e − 3	0.777 ± 8.91e − 4	0.943 ± 8.026e − 3
medgemma zero_shot mlp	0.791 ± 0e0	0.952 ± 0e0	0.773 ± 0e0	0.95 ± 0e0
medgemma zero_shot patchtsmixer	0.791 ± 0e0	0.958 ± 0e0	0.776 ± 0e0	0.951 ± 0e0
medgemma zero_shot timemixer	0.793 ± 0e0	0.953 ± 0e0	0.78 ± 0e0	0.946 ± 0e0
No_sum_zero_shot zero_shot +TFM lstm	0.797 ± 5.88e − 4	0.936 ± 4.818e − 4	0.772 ± 9.725e − 4	0.945 ± 2.726e − 3
No_sum_zero_shot zero_shot +TFM mlp	0.787 ± 3.481e − 3	0.954 ± 2.038e − 3	0.772 ± 5.512e − 3	0.943 ± 4.855e − 3
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.787 ± 1.369e − 3	0.948 ± 7.006e − 4	0.768 ± 6.24e − 4	0.947 ± 1.247e − 3
No_sum_zero_shot zero_shot +TFM timemixer	0.793 ± 4.514e − 3	0.941 ± 3.638e − 3	0.774 ± 2.981e − 3	0.942 ± 4.471e − 3
No_sum_zero_shot lstm	0.808 ± 9.673e − 4	0.939 ± 1.164e − 3	0.787 ± 4.993e − 4	0.95 ± 8.082e − 4
No_sum_zero_shot mlp	0.8 ± 0e0	0.947 ± 0e0	0.775 ± 0e0	0.946 ± 0e0
No_sum_zero_shot patchtsmixer	0.807 ± 1.189e − 3	0.948 ± 5.173e − 4	0.781 ± 1.305e − 3	0.946 ± 1.944e − 3
No_sum_zero_shot timemixer	0.807 ± 9.704e − 4	0.941 ± 1.385e − 3	0.789 ± 5.1e − 4	0.943 ± 1.549e − 3
right lstm	0.711 ± 1.256e − 3	0.913 ± 1.471e − 3	0.715 ± 7.657e − 3	0.955 ± 1.734e − 2
right mlp	0.723 ± 8.389e − 3	0.938 ± 2.397e − 2	0.663 ± 1.991e − 2	0.813 ± 5.145e − 2
right patchtsmixer	0.719 ± 3.188e − 3	0.879 ± 1.593e − 2	0.692 ± 1.36e − 2	0.823 ± 2.478e − 2
right timemixer	0.689 ± 9.4e − 3	0.8 ± 4.114e − 2	0.671 ± 8.889e − 3	0.797 ± 2.827e − 2

Table 33: Cross-site transfer results - Lab. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	masked_mae	masked_mse	masked_mae	masked_mse	masked_mae	masked_mse
Llama-3.1 CoT +TFM lstm	0.073 ± 1.4e − 3	0.014 ± 3.704e − 4	0.093 ± 8.415e − 4	0.018 ± 2.152e − 4	0.074 ± 8.799e − 4	0.015 ± 2.152e − 4
Llama-3.1 CoT +TFM mlp	0.078 ± 2.696e − 4	0.013 ± 7.234e − 5	0.096 ± 3.704e − 4	0.018 ± 1.193e − 4	0.08 ± 8.415e − 4	0.015 ± 2.801e − 4
Llama-3.1 CoT +TFM patchtsmixer	0.125 ± 7.892e − 4	0.028 ± 3.819e − 4	0.139 ± 9.019e − 4	0.033 ± 4.1e − 4	0.455 ± 7.315e − 2	0.339 ± 9.426e − 2
Llama-3.1 CoT +TFM timemixer	0.077 ± 7.422e − 4	0.013 ± 2.196e − 4	0.094 ± 3.646e − 4	0.018 ± 2.452e − 4	0.109 ± 6.269e − 4	0.021 ± 2.065e − 4
Llama-3.1 CoT lstm	0.075 ± 4.255e − 4	0.015 ± 1.569e − 4	0.094 ± 4.84e − 4	0.019 ± 1.787e − 4	0.077 ± 6.466e − 4	0.017 ± 1.758e − 4
Llama-3.1 CoT mlp	0.077 ± 6.316e − 4	0.015 ± 1.471e − 4	0.095 ± 5.053e − 4	0.019 ± 1.637e − 4	0.078 ± 4.155e − 4	0.016 ± 1.392e − 4
Llama-3.1 CoT patchtsmixer	0.114 ± 1.501e − 3	0.024 ± 4.802e − 4	0.131 ± 1.895e − 4	0.03 ± 2.291e − 4	0.102 ± 5.067e − 3	0.021 ± 1.382e − 3
Llama-3.1 CoT timemixer	0.079 ± 1.151e − 3	0.015 ± 3.293e − 4	0.098 ± 1.675e − 3	0.02 ± 5.097e − 4	0.09 ± 9.862e − 4	0.018 ± 2.485e − 4
Llama-3.1 ICD +TFM lstm	0.074 ± 7.703e − 4	0.014 ± 1.793e − 4	0.094 ± 3.676e − 4	0.019 ± 1.301e − 4	0.074 ± 1.087e − 3	0.015 ± 2.811e − 4
Llama-3.1 ICD +TFM mlp	0.079 ± 1.656e − 4	0.014 ± 3.786e − 5	0.097 ± 4.725e − 4	0.019 ± 1.769e − 4	0.08 ± 1.747e − 4	0.015 ± 6.11e − 5
Llama-3.1 ICD +TFM patchtsmixer	0.125 ± 2.646e − 5	0.028 ± 1.473e − 4	0.14 ± 1.908e − 4	0.034 ± 1.058e − 4	0.453 ± 9.85e − 2	0.338 ± 1.324e − 1
Llama-3.1 ICD +TFM timemixer	0.077 ± 4.784e − 4	0.013 ± 1.877e − 4	0.095 ± 2.836e − 4	0.018 ± 2.001e − 4	0.109 ± 8.773e − 4	0.021 ± 3.139e − 4
Llama-3.1 ICD lstm	0.074 ± 3.806e − 4	0.015 ± 7.848e − 5	0.095 ± 5.013e − 4	0.02 ± 1.673e − 4	0.076 ± 7.454e − 4	0.016 ± 2.512e − 4
Llama-3.1 ICD mlp	0.077 ± 6.549e − 4	0.015 ± 2.034e − 4	0.096 ± 5.709e − 4	0.02 ± 1.431e − 4	0.078 ± 4.151e − 4	0.016 ± 1.066e − 4
Llama-3.1 ICD patchtsmixer	0.115 ± 9.215e − 4	0.024 ± 2.721e − 4	0.13 ± 1.749e − 3	0.03 ± 6.951e − 4	0.1 ± 5.492e − 3	0.021 ± 1.42e − 3
Llama-3.1 ICD timemixer	0.079 ± 5.826e − 4	0.015 ± 2.47e − 4	0.098 ± 8.147e − 4	0.02 ± 1.876e − 4	0.088 ± 4.189e − 4	0.017 ± 1.237e − 4
Llama-3.1 Trend +TFM lstm	0.074 ± 1.042e − 3	0.014 ± 1.662e − 4	0.094 ± 4.75e − 4	0.019 ± 1.803e − 4	0.073 ± 8.15e − 4	0.014 ± 2.974e − 4
Llama-3.1 Trend +TFM mlp	0.079 ± 6.755e − 4	0.014 ± 2.344e − 4	0.097 ± 3.592e − 4	0.019 ± 1.127e − 4	0.08 ± 1.446e − 3	0.015 ± 4.917e − 4
Llama-3.1 Trend +TFM patchtsmixer	0.125 ± 1.351e − 3	0.028 ± 4.681e − 4	0.14 ± 1.02e − 3	0.034 ± 3.47e − 4	0.39 ± 1.236e − 2	0.275 ± 1.994e − 2
Llama-3.1 Trend +TFM timemixer	0.077 ± 3.602e − 4	0.013 ± 4e − 5	0.095 ± 2.159e − 4	0.018 ± 2.517e − 5	0.108 ± 1.084e − 3	0.021 ± 4.368e − 4
Llama-3.1 Trend lstm	0.074 ± 3.203e − 4	0.014 ± 9.574e − 6	0.094 ± 4.096e − 4	0.019 ± 1.576e − 4	0.076 ± 1.132e − 3	0.016 ± 2.965e − 4
Llama-3.1 Trend mlp	0.076 ± 1.679e − 4	0.014 ± 9.179e − 5	0.096 ± 2.505e − 4	0.019 ± 1.044e − 4	0.078 ± 6.638e − 4	0.016 ± 1.73e − 4
Llama-3.1 Trend patchtsmixer	0.114 ± 1.396e − 3	0.024 ± 3.739e − 4	0.13 ± 9.261e − 4	0.029 ± 3.314e − 4	0.095 ± 4.376e − 3	0.019 ± 9.721e − 4
Llama-3.1 Trend timemixer	0.078 ± 1.254e − 3	0.015 ± 2.831e − 4	0.097 ± 1.196e − 3	0.02 ± 2.822e − 4	0.089 ± 9.46e − 4	0.017 ± 2.888e − 4
Llama-3.1 zero_shot +TFM lstm	0.072 ± 7.624e − 4	0.013 ± 2.021e − 4	0.089 ± 1.266e − 4	0.017 ± 6.807e − 5	0.073 ± 4.245e − 4	0.014 ± 1.677e − 4
Llama-3.1 zero_shot +TFM mlp	0.075 ± 1.989e − 3	0.012 ± 6.351e − 4	0.092 ± 1.021e − 3	0.017 ± 2.914e − 4	0.077 ± 8.995e − 4	0.014 ± 3.153e − 4
Llama-3.1 zero_shot +TFM patchtsmixer	0.126 ± 9.015e − 4	0.028 ± 2.178e − 4	0.14 ± 1.091e − 3	0.034 ± 3.46e − 4	0.276 ± 1.287e − 1	0.149 ± 1.026e − 1
Llama-3.1 zero_shot +TFM timemixer	0.074 ± 3.95e − 4	0.012 ± 1.015e − 4	0.091 ± 2.234e − 4	0.017 ± 1.274e − 4	0.107 ± 1.173e − 3	0.02 ± 5.33e − 4
Llama-3.1 zero_shot lstm	0.072 ± 3.809e − 4	0.014 ± 9.106e − 5	0.091 ± 2.923e − 4	0.018 ± 1.134e − 4	0.076 ± 1.052e − 3	0.016 ± 2.92e − 4
Llama-3.1 zero_shot mlp	0.074 ± 4.196e − 4	0.014 ± 9.946e − 5	0.093 ± 3.957e − 4	0.019 ± 1.445e − 4	0.076 ± 4.853e − 4	0.016 ± 1.447e − 4
Llama-3.1 zero_shot patchtsmixer	0.114 ± 1.306e − 3	0.024 ± 5.302e − 4	0.129 ± 2.36e − 3	0.029 ± 8.739e − 4	0.093 ± 3.283e − 3	0.019 ± 5.885e − 4
Llama-3.1 zero_shot timemixer	0.076 ± 1.425e − 3	0.014 ± 3.643e − 4	0.095 ± 1.243e − 3	0.019 ± 2.466e − 4	0.088 ± 1.194e − 3	0.017 ± 3.195e − 4
TFM lstm	0.075 ± 6.816e − 4	0.015 ± 1.303e − 4	0.094 ± 6.191e − 5	0.019 ± 5.715e − 5	0.075 ± 6.752e − 4	0.016 ± 1.608e − 4
TFM mlp	0.077 ± 3.99e − 4	0.015 ± 8.921e − 5	0.095 ± 3.025e − 4	0.019 ± 5.737e − 5	0.076 ± 4.834e − 4	0.016 ± 6.683e − 5
TFM patchtsmixer	0.126 ± 1.558e − 3	0.028 ± 5.498e − 4	0.132 ± 7.469e − 3	0.03 ± 3.232e − 3	0.142 ± 3.861e − 3	0.036 ± 1.391e − 3
TFM timemixer	0.076 ± 5.576e − 4	0.014 ± 1.573e − 4	0.095 ± 3.951e − 4	0.019 ± 4.546e − 5	0.102 ± 3.706e − 3	0.02 ± 1.618e − 3
TSDE lstm	0.066 ± 3.204e − 4	0.012 ± 7.047e − 5	0.072 ± 1.936e − 4	0.012 ± 3.916e − 5	0.052 ± 4.738e − 4	0.009 ± 3.317e − 5
TSDE mlp	0.067 ± 2.139e − 4	0.011 ± 2.16e − 5	0.073 ± 2.428e − 4	0.012 ± 1.414e − 5	0.054 ± 1.744e − 4	0.008 ± 2.754e − 5
TSDE patchtsmixer	0.127 ± 1.484e − 3	0.029 ± 5.634e − 4	0.103 ± 1.141e − 2	0.02 ± 3.797e − 3	0.177 ± 1.168e − 2	0.053 ± 6.641e − 3
TSDE timemixer	0.069 ± 8.416e − 4	0.012 ± 7.416e − 5	0.073 ± 2.546e − 4	0.012 ± 2.944e − 5	0.091 ± 4.927e − 4	0.015 ± 1.533e − 4
gemini-2.0-flash CoT +TFM lstm	0.07 ± 4.102e − 4	0.012 ± 1.716e − 4	0.091 ± 8.102e − 4	0.017 ± 1.607e − 4	0.072 ± 2.073e − 3	0.014 ± 5.859e − 4
gemini-2.0-flash CoT +TFM mlp	0.076 ± 2.954e − 4	0.013 ± 8.888e − 5	0.095 ± 1.234e − 4	0.018 ± 6.429e − 5	0.079 ± 2.364e − 4	0.014 ± 4.933e − 5
gemini-2.0-flash CoT +TFM patchtsmixer	0.125 ± 1.115e − 3	0.028 ± 3.247e − 4	0.133 ± 1.085e − 2	0.031 ± 4.64e − 3	0.425 ± 3.751e − 2	0.288 ± 5.694e − 2
gemini-2.0-flash CoT +TFM timemixer	0.075 ± 4.59e − 4	0.012 ± 1.058e − 4	0.092 ± 8.083e − 4	0.017 ± 2.371e − 4	0.107 ± 8.752e − 4	0.021 ± 4.05e − 4
gemini-2.0-flash CoT lstm	0.072 ± 6.472e − 4	0.013 ± 1.982e − 4	0.091 ± 3.33e − 4	0.018 ± 8.884e − 5	0.072 ± 8.141e − 4	0.014 ± 2.108e − 4
gemini-2.0-flash CoT mlp	0.073 ± 4.444e − 4	0.013 ± 1.389e − 4	0.092 ± 2.359e − 4	0.018 ± 9.815e − 5	0.073 ± 3.309e − 4	0.014 ± 6.551e − 5
gemini-2.0-flash CoT patchtsmixer	0.112 ± 8.465e − 4	0.023 ± 2.616e − 4	0.128 ± 8.426e − 4	0.029 ± 3.549e − 4	0.097 ± 1.014e − 2	0.019 ± 3.068e − 3
gemini-2.0-flash CoT timemixer	0.076 ± 1.872e − 3	0.014 ± 4.269e − 4	0.093 ± 8.366e − 4	0.018 ± 1.911e − 4	0.084 ± 1.191e − 3	0.015 ± 2.974e − 4
gemini-2.0-flash ICD +TFM lstm	0.072 ± 6.045e − 4	0.013 ± 2.179e − 4	0.091 ± 5.595e − 4	0.017 ± 8.963e − 5	0.073 ± 2.059e − 3	0.014 ± 6.668e − 4
gemini-2.0-flash ICD +TFM mlp	0.078 ± 4.842e − 4	0.013 ± 1.528e − 4	0.094 ± 2.538e − 4	0.017 ± 7.55e − 5	0.08 ± 1.065e − 3	0.014 ± 3.148e − 4
gemini-2.0-flash ICD +TFM patchtsmixer	0.124 ± 1.068e − 3	0.028 ± 3.609e − 4	0.139 ± 7.686e − 4	0.033 ± 4.2e − 4	0.379 ± 2.04e − 2	0.229 ± 2.503e − 2
gemini-2.0-flash ICD +TFM timemixer	0.074 ± 4.486e − 4	0.012 ± 2.301e − 4	0.093 ± 7.472e − 4	0.017 ± 2.74e − 4	0.107 ± 1.184e − 3	0.021 ± 5.147e − 4
gemini-2.0-flash ICD lstm	0.072 ± 1.111e − 3	0.013 ± 2.629e − 4	0.092 ± 1.443e − 4	0.018 ± 8.287e − 5	0.073 ± 4.526e − 4	0.015 ± 9.142e − 5
gemini-2.0-flash ICD mlp	0.074 ± 5.099e − 4	0.013 ± 1.425e − 4	0.093 ± 3.266e − 4	0.018 ± 9.032e − 5	0.074 ± 4.188e − 4	0.015 ± 1.209e − 4
gemini-2.0-flash ICD patchtsmixer	0.113 ± 5.735e − 4	0.023 ± 1.782e − 4	0.129 ± 5.676e − 4	0.029 ± 3.551e − 4	0.093 ± 1.037e − 2	0.018 ± 2.615e − 3
gemini-2.0-flash ICD timemixer	0.078 ± 2.33e − 3	0.014 ± 4.438e − 4	0.095 ± 1.849e − 3	0.019 ± 4.16e − 4	0.085 ± 7.941e − 4	0.016 ± 1.919e − 4
gemini-2.0-flash Trend +TFM lstm	0.072 ± 9.879e − 4	0.013 ± 2.553e − 4	0.093 ± 3.412e − 4	0.018 ± 9e − 5	0.073 ± 1.085e − 3	0.014 ± 3.667e − 4
gemini-2.0-flash Trend +TFM mlp	0.077 ± 1.128e − 3	0.013 ± 3.963e − 4	0.096 ± 3.723e − 4	0.018 ± 1.345e − 4	0.08 ± 9.753e − 4	0.014 ± 3.279e − 4
gemini-2.0-flash Trend +TFM patchtsmixer	0.126 ± 4.194e − 4	0.028 ± 2.052e − 4	0.139 ± 3.086e − 4	0.034 ± 2.179e − 4	0.339 ± 3.58e − 2	0.201 ± 3.404e − 2
gemini-2.0-flash Trend +TFM timemixer	0.074 ± 1.176e − 3	0.012 ± 3.675e − 4	0.093 ± 8.504e − 4	0.017 ± 1.986e − 4	0.108 ± 3.383e − 4	0.021 ± 1.557e − 4
gemini-2.0-flash Trend lstm	0.072 ± 8.606e − 4	0.014 ± 3.057e − 4	0.092 ± 4.894e − 4	0.018 ± 1.394e − 4	0.073 ± 6.635e − 4	0.015 ± 1.718e − 4
gemini-2.0-flash Trend mlp	0.073 ± 2.54e − 4	0.013 ± 9.747e − 5	0.093 ± 6.64e − 4	0.018 ± 2.25e − 4	0.074 ± 3.777e − 4	0.015 ± 1.139e − 4
gemini-2.0-flash Trend patchtsmixer	0.114 ± 1.558e − 3	0.024 ± 5.443e − 4	0.129 ± 7.93e − 4	0.029 ± 3.429e − 4	0.089 ± 2.222e − 3	0.017 ± 4.8e − 4
gemini-2.0-flash Trend timemixer	0.077 ± 9.486e − 4	0.014 ± 2.533e − 4	0.094 ± 5.857e − 4	0.018 ± 1.921e − 4	0.086 ± 3.859e − 4	0.016 ± 2.121e − 4
gemini-2.0-flash zero_shot +TFM lstm	0.07 ± 1.208e − 3	0.012 ± 5.456e − 4	0.091 ± 5.216e − 4	0.017 ± 1.674e − 4	0.071 ± 9.767e − 4	0.013 ± 3.25e − 4
gemini-2.0-flash zero_shot +TFM mlp	0.076 ± 7.953e − 4	0.012 ± 2.754e − 4	0.093 ± 5.311e − 4	0.017 ± 1.45e − 4	0.078 ± 6.503e − 4	0.014 ± 1.682e − 4
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.125 ± 1.584e − 3	0.028 ± 4.748e − 4	0.138 ± 1.492e − 3	0.033 ± 6.374e − 4	0.316 ± 1.156e − 2	0.182 ± 1.268e − 2
gemini-2.0-flash zero_shot +TFM timemixer	0.073 ± 4.336e − 4	0.012 ± 1.443e − 4	0.091 ± 2.318e − 4	0.017 ± 1.015e − 4	0.105 ± 9e − 5	0.02 ± 7.506e − 5
gemini-2.0-flash zero_shot lstm	0.07 ± 5.497e − 4	0.013 ± 1.687e − 4	0.09 ± 4.867e − 4	0.018 ± 1.187e − 4	0.071 ± 6.809e − 4	0.014 ± 7.455e − 5
gemini-2.0-flash zero_shot mlp	0.073 ± 1.195e − 3	0.013 ± 3.206e − 4	0.091 ± 4.109e − 4	0.018 ± 1.258e − 4	0.072 ± 7.889e − 4	0.014 ± 2.163e − 4
gemini-2.0-flash zero_shot patchtsmixer	0.112 ± 7.7e − 4	0.023 ± 3.372e − 4	0.128 ± 8.403e − 4	0.028 ± 2.666e − 4	0.088 ± 6.003e − 3	0.017 ± 1.257e − 3
gemini-2.0-flash zero_shot timemixer	0.074 ± 1.52e − 3	0.013 ± 3.207e − 4	0.093 ± 1.556e − 3	0.018 ± 3.88e − 4	0.083 ± 1.11e − 3	0.015 ± 3.31e − 4
interp lstm	0.051 ± 5.768e − 4	0.008 ± 7.958e − 5	0.068 ± 1.974e − 4	0.011 ± 4.992e − 5	0.048 ± 2.052e − 4	0.008 ± 2.217e − 5
interp mlp	0.06 ± 2.003e − 3	0.009 ± 2.888e − 4	0.075 ± 1.632e − 4	0.013 ± 3.916e − 5	0.054 ± 2.749e − 4	0.009 ± 4.243e − 5
interp patchtsmixer	0.184 ± 4.418e − 3	0.435 ± 2.706e − 2	0.142 ± 2.662e − 2	0.103 ± 7.948e − 2	0.117 ± 1.12e − 2	0.133 ± 4.628e − 2
interp timemixer	0.231 ± 1.652e − 2	0.121 ± 1.241e − 2	0.19 ± 4.754e − 3	0.082 ± 3.223e − 3	0.319 ± 3.802e − 3	0.22 ± 4.306e − 3
mean lstm	0.05 ± 3.677e − 4	0.007 ± 6.557e − 5	0.069 ± 1.621e − 4	0.012 ± 6.455e − 5	0.049 ± 4.51e − 4	0.008 ± 4.193e − 5
mean mlp	0.054 ± 3.937e − 4	0.008 ± 3.697e − 5	0.072 ± 6.137e − 5	0.012 ± 2.16e − 5	0.051 ± 7.141e − 5	0.008 ± 1.414e − 5
mean patchtsmixer	0.106 ± 3.928e − 4	0.04 ± 2.229e − 4	0.108 ± 2.447e − 4	0.035 ± 4.372e − 4	0.082 ± 4.109e − 4	0.028 ± 2.391e − 4
mean timemixer	0.195 ± 1.005e − 2	0.09 ± 7.819e − 3	0.165 ± 1.271e − 3	0.06 ± 7.447e − 4	0.262 ± 3.148e − 2	0.167 ± 2.992e − 2
No_sum_CoT CoT +TFM lstm	0.065 ± 5.46e − 4	0.011 ± 9.165e − 5	0.083 ± 5.05e − 4	0.015 ± 1.514e − 4	0.064 ± 1.559e − 3	0.011 ± 4.406e − 4
No_sum_CoT CoT +TFM mlp	0.068 ± 1.489e − 3	0.01 ± 3.9e − 4	0.084 ± 1.162e − 3	0.014 ± 2.957e − 4	0.067 ± 1.8e − 3	0.011 ± 4.652e − 4
No_sum_CoT CoT +TFM patchtsmixer	0.124 ± 4.937e − 4	0.027 ± 5.429e − 4	0.13 ± 1.521e − 2	0.03 ± 6.112e − 3	0.095 ± 2.757e − 2	0.02 ± 1.131e − 2
No_sum_CoT CoT +TFM timemixer	0.069 ± 1.213e − 3	0.011 ± 2.967e − 4	0.084 ± 3.821e − 4	0.014 ± 8.387e − 5	0.1 ± 3.803e − 4	0.018 ± 1.442e − 4
No_sum_CoT lstm	0.068 ± 1.919e − 4	0.012 ± 2.517e − 5	0.087 ± 9.539e − 4	0.017 ± 1.947e − 4	0.066 ± 7.127e − 4	0.013 ± 1.836e − 4
No_sum_CoT mlp	0.068 ± 4.33e − 4	0.011 ± 1.034e − 4	0.087 ± 2.246e − 4	0.016 ± 6.245e − 5	0.068 ± 5.267e − 4	0.013 ± 1.916e − 4
No_sum_CoT patchtsmixer	0.11 ± 4.16e − 4	0.022 ± 2.835e − 4	0.126 ± 3.01e − 4	0.028 ± 1.877e − 4	0.091 ± 7.053e − 3	0.017 ± 1.781e − 3
No_sum_CoT timemixer	0.071 ± 1.913e − 3	0.012 ± 3.332e − 4	0.088 ± 1.554e − 3	0.016 ± 4.282e − 4	0.078 ± 5.167e − 4	0.014 ± 1.441e − 4
No_sum_ICD ICD +TFM lstm	0.064 ± 1.066e − 3	0.011 ± 2.75e − 4	0.082 ± 3.418e − 4	0.015 ± 1.05e − 4	0.063 ± 9.255e − 4	0.011 ± 1.724e − 4
No_sum_ICD ICD +TFM mlp	0.068 ± 4.036e − 4	0.01 ± 1.106e − 4	0.084 ± 3.98e − 4	0.014 ± 1.531e − 4	0.067 ± 1.261e − 3	0.011 ± 3.371e − 4
No_sum_ICD ICD +TFM patchtsmixer	0.123 ± 4.822e − 4	0.027 ± 3.707e − 4	0.123 ± 1.306e − 2	0.027 ± 5.384e − 3	0.282 ± 1.625e − 1	0.164 ± 1.377e − 1
No_sum_ICD ICD +TFM timemixer	0.068 ± 3.485e − 4	0.01 ± 1.882e − 4	0.084 ± 1.168e − 3	0.014 ± 3.993e − 4	0.099 ± 1.084e − 3	0.018 ± 3.851e − 4
No_sum_ICD lstm	0.068 ± 8.225e − 4	0.012 ± 2.161e − 4	0.086 ± 1.076e − 3	0.016 ± 2.791e − 4	0.066 ± 5.363e − 4	0.013 ± 1.179e − 4
No_sum_ICD mlp	0.068 ± 1.764e − 3	0.012 ± 4.181e − 4	0.087 ± 8.077e − 4	0.016 ± 2.247e − 4	0.068 ± 6.574e − 4	0.013 ± 1.47e − 4
No_sum_ICD patchtsmixer	0.112 ± 1.053e − 3	0.023 ± 3.8e − 4	0.126 ± 3.79e − 4	0.028 ± 1.564e − 4	0.089 ± 3.511e − 3	0.017 ± 6.877e − 4
No_sum_ICD timemixer	0.071 ± 1.266e − 3	0.012 ± 3.096e − 4	0.087 ± 8.752e − 4	0.016 ± 2.513e − 4	0.079 ± 7.487e − 4	0.014 ± 2.395e − 4
No_sum_Trend Trend +TFM lstm	0.064 ± 3e − 3	0.01 ± 7.238e − 4	0.082 ± 4.706e − 4	0.015 ± 1.732e − 4	0.064 ± 1.618e − 3	0.011 ± 4.07e − 4
No_sum_Trend Trend +TFM mlp	0.067 ± 1.328e − 3	0.01 ± 3.251e − 4	0.084 ± 8.851e − 4	0.014 ± 2.506e − 4	0.067 ± 4.674e − 4	0.011 ± 1.365e − 4
No_sum_Trend Trend +TFM patchtsmixer	0.124 ± 1.31e − 3	0.027 ± 3.134e − 4	0.131 ± 7.397e − 3	0.03 ± 3.107e − 3	0.264 ± 1.202e − 1	0.145 ± 1.013e − 1
No_sum_Trend Trend +TFM timemixer	0.069 ± 9.413e − 4	0.01 ± 1.45e − 4	0.084 ± 6.694e − 4	0.014 ± 1.818e − 4	0.1 ± 1.34e − 3	0.018 ± 5.112e − 4

Table 34: In-distribution results - Forecast. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	masked_mae	masked_mse	masked_mae	masked_mse	masked_mae	masked_mse
No_sum_Trend lstm	0.068 ± 1.317e − 3	0.012 ± 3.116e − 4	0.085 ± 4.437e − 4	0.016 ± 1.175e − 4	0.066 ± 5.187e − 4	0.013 ± 1.835e − 4
No_sum_Trend mlp	0.068 ± 6.357e − 4	0.011 ± 1.617e − 4	0.087 ± 1.297e − 3	0.016 ± 3.724e − 4	0.067 ± 8.032e − 4	0.012 ± 2.634e − 4
No_sum_Trend patchtsmixer	0.111 ± 8.596e − 4	0.022 ± 2.076e − 4	0.126 ± 1.205e − 3	0.028 ± 5.758e − 4	0.089 ± 6.671e − 3	0.017 ± 1.798e − 3
No_sum_Trend timemixer	0.072 ± 2.463e − 3	0.012 ± 5.846e − 4	0.088 ± 1.163e − 3	0.016 ± 3.008e − 4	0.079 ± 1.129e − 3	0.014 ± 3.756e − 4
medgemma CoT +TFM lstm	0.069 ± 2.873e − 4	0.012 ± 5.686e − 5	0.088 ± 6.381e − 4	0.016 ± 1.65e − 4	0.069 ± 1.11e − 3	0.013 ± 3.317e − 4
medgemma CoT +TFM mlp	0.073 ± 5.965e − 4	0.012 ± 1.747e − 4	0.093 ± 3.061e − 4	0.017 ± 1.136e − 4	0.077 ± 1.177e − 3	0.013 ± 3.378e − 4
medgemma CoT +TFM patchtsmixer	0.126 ± 1.701e − 4	0.028 ± 2.136e − 4	0.138 ± 1.191e − 3	0.033 ± 4.423e − 4	0.426 ± 2.746e − 2	0.295 ± 3.72e − 2
medgemma CoT +TFM timemixer	0.072 ± 1.163e − 3	0.011 ± 2.166e − 4	0.09 ± 7.181e − 4	0.016 ± 2.615e − 4	0.106 ± 1.041e − 3	0.02 ± 4.748e − 4
medgemma CoT lstm	0.069 ± 6.06e − 4	0.013 ± 1.515e − 4	0.088 ± 6.359e − 4	0.017 ± 1.708e − 4	0.07 ± 6.807e − 4	0.014 ± 1.799e − 4
medgemma CoT mlp	0.071 ± 5.875e − 4	0.013 ± 1.431e − 4	0.09 ± 7.285e − 4	0.017 ± 2.234e − 4	0.072 ± 4.666e − 4	0.014 ± 1.45e − 4
medgemma CoT patchtsmixer	0.112 ± 1.6e − 3	0.023 ± 4.853e − 4	0.128 ± 6.378e − 4	0.028 ± 1.309e − 4	0.092 ± 4.877e − 3	0.018 ± 1.024e − 3
medgemma CoT timemixer	0.075 ± 2.101e − 3	0.013 ± 3.557e − 4	0.091 ± 8.831e − 4	0.017 ± 2.701e − 4	0.083 ± 8.314e − 4	0.015 ± 3.74e − 4
medgemma ICD +TFM lstm	0.072 ± 1.752e − 3	0.013 ± 6.251e − 4	0.09 ± 1.406e − 3	0.017 ± 3.769e − 4	0.072 ± 1.18e − 3	0.014 ± 3.703e − 4
medgemma ICD +TFM mlp	0.076 ± 6.813e − 4	0.013 ± 1.93e − 4	0.094 ± 2.524e − 4	0.018 ± 9.504e − 5	0.079 ± 1.08e − 3	0.014 ± 3.46e − 4
medgemma ICD +TFM patchtsmixer	0.125 ± 1.208e − 3	0.028 ± 3.5e − 4	0.139 ± 3.731e − 4	0.033 ± 2.201e − 4	0.419 ± 3.58e − 2	0.279 ± 4.63e − 2
medgemma ICD +TFM timemixer	0.074 ± 1.771e − 3	0.012 ± 3.913e − 4	0.092 ± 7.543e − 4	0.017 ± 2.829e − 4	0.107 ± 3.934e − 4	0.02 ± 1.861e − 4
medgemma ICD lstm	0.071 ± 4.095e − 4	0.013 ± 1.193e − 4	0.091 ± 3.183e − 4	0.018 ± 1.303e − 4	0.072 ± 7.333e − 4	0.015 ± 1.498e − 4
medgemma ICD mlp	0.073 ± 9.869e − 4	0.013 ± 2.471e − 4	0.091 ± 5.297e − 4	0.018 ± 1.841e − 4	0.074 ± 4.049e − 4	0.015 ± 1.207e − 4
medgemma ICD patchtsmixer	0.113 ± 3.482e − 4	0.023 ± 1.571e − 4	0.128 ± 5.202e − 4	0.029 ± 2.037e − 4	0.093 ± 6.342e − 3	0.018 ± 1.394e − 3
medgemma ICD timemixer	0.076 ± 1.618e − 3	0.014 ± 2.87e − 4	0.094 ± 1.292e − 3	0.018 ± 3.181e − 4	0.085 ± 8.767e − 4	0.016 ± 1.905e − 4
medgemma Trend +TFM lstm	0.07 ± 4.701e − 4	0.012 ± 1.778e − 4	0.09 ± 9.335e − 4	0.017 ± 2.538e − 4	0.071 ± 1.38e − 3	0.014 ± 4.46e − 4
medgemma Trend +TFM mlp	0.075 ± 5.048e − 4	0.012 ± 1.498e − 4	0.092 ± 6.178e − 4	0.017 ± 1.562e − 4	0.078 ± 1.407e − 3	0.014 ± 4.779e − 4
medgemma Trend +TFM patchtsmixer	0.126 ± 1.025e − 3	0.028 ± 1.95e − 4	0.134 ± 1.056e − 2	0.031 ± 4.599e − 3	0.375 ± 4.543e − 2	0.238 ± 4.575e − 2
medgemma Trend +TFM timemixer	0.073 ± 1.142e − 3	0.012 ± 4.309e − 4	0.091 ± 3.443e − 4	0.017 ± 1.277e − 4	0.107 ± 1.32e − 3	0.021 ± 5.525e − 4
medgemma Trend lstm	0.071 ± 4.996e − 4	0.013 ± 1.388e − 4	0.09 ± 2.858e − 4	0.018 ± 7.789e − 5	0.073 ± 3.869e − 4	0.015 ± 1.958e − 4
medgemma Trend mlp	0.072 ± 7.023e − 4	0.013 ± 1.992e − 4	0.092 ± 3.242e − 4	0.018 ± 7.724e − 5	0.074 ± 4.473e − 4	0.015 ± 1.184e − 4
medgemma Trend patchtsmixer	0.113 ± 1.433e − 3	0.023 ± 3.029e − 4	0.128 ± 1.315e − 3	0.029 ± 4.507e − 4	0.095 ± 4.215e − 3	0.019 ± 9.673e − 4
medgemma Trend timemixer	0.075 ± 1.406e − 3	0.014 ± 3.356e − 4	0.094 ± 9.093e − 4	0.018 ± 3.172e − 4	0.086 ± 2.637e − 4	0.016 ± 1.499e − 4
medgemma zero_shot +TFM lstm	0.069 ± 2.157e − 4	0.012 ± 7.095e − 5	0.09 ± 2.8e − 4	0.017 ± 8.718e − 5	0.071 ± 2.487e − 3	0.014 ± 6.589e − 4
medgemma zero_shot +TFM mlp	0.075 ± 1.15e − 3	0.012 ± 3.724e − 4	0.093 ± 8.479e − 4	0.017 ± 2.829e − 4	0.078 ± 5.486e − 4	0.014 ± 1.484e − 4
medgemma zero_shot +TFM patchtsmixer	0.125 ± 1.021e − 3	0.028 ± 5.116e − 4	0.13 ± 1.54e − 2	0.03 ± 6.322e − 3	0.439 ± 4.731e − 2	0.31 ± 4.83e − 2
medgemma zero_shot +TFM timemixer	0.073 ± 7.679e − 4	0.012 ± 2.409e − 4	0.091 ± 7.674e − 4	0.017 ± 2.663e − 4	0.105 ± 1.015e − 3	0.02 ± 3.544e − 4
medgemma zero_shot lstm	0.07 ± 3.414e − 4	0.013 ± 6.782e − 5	0.09 ± 1.015e − 3	0.018 ± 2.858e − 4	0.071 ± 5.2e − 4	0.014 ± 1.564e − 4
medgemma zero_shot mlp	0.072 ± 2.818e − 4	0.013 ± 1.03e − 4	0.091 ± 6.299e − 4	0.018 ± 1.88e − 4	0.073 ± 9.287e − 4	0.014 ± 2.866e − 4
medgemma zero_shot patchtsmixer	0.112 ± 1.105e − 3	0.023 ± 2.934e − 4	0.128 ± 9.704e − 4	0.028 ± 2.298e − 4	0.092 ± 8.414e − 3	0.018 ± 2.275e − 3
medgemma zero_shot timemixer	0.074 ± 2.111e − 3	0.013 ± 4.833e − 4	0.093 ± 5.915e − 4	0.018 ± 1.072e − 4	0.084 ± 7.86e − 4	0.016 ± 1.688e − 4
No_sum_zero_shot zero_shot +TFM lstm	0.065 ± 1.6e − 3	0.011 ± 4.01e − 4	0.083 ± 1.504e − 3	0.015 ± 3.972e − 4	0.064 ± 1.191e − 3	0.011 ± 2.845e − 4
No_sum_zero_shot zero_shot +TFM mlp	0.067 ± 6.681e − 4	0.01 ± 1.626e − 4	0.084 ± 8.605e − 4	0.015 ± 2.631e − 4	0.068 ± 1.01e − 3	0.011 ± 2.639e − 4
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.124 ± 1.164e − 3	0.027 ± 5.45e − 4	0.124 ± 1.319e − 2	0.027 ± 5.462e − 3	0.256 ± 1.429e − 1	0.144 ± 1.174e − 1
No_sum_zero_shot zero_shot +TFM timemixer	0.068 ± 1.837e − 3	0.01 ± 4.328e − 4	0.085 ± 1.501e − 3	0.015 ± 3.487e − 4	0.1 ± 7.862e − 4	0.018 ± 2.641e − 4
No_sum_zero_shot lstm	0.067 ± 1.073e − 3	0.012 ± 2.413e − 4	0.086 ± 1.129e − 3	0.016 ± 3.381e − 4	0.066 ± 8.154e − 4	0.013 ± 2.403e − 4
No_sum_zero_shot mlp	0.068 ± 8.283e − 4	0.012 ± 2.077e − 4	0.086 ± 9.896e − 4	0.016 ± 2.701e − 4	0.068 ± 4.37e − 4	0.013 ± 1.22e − 4
No_sum_zero_shot patchtsmixer	0.11 ± 4.895e − 4	0.022 ± 2.851e − 4	0.126 ± 5.631e − 4	0.028 ± 2.757e − 4	0.089 ± 8.783e − 3	0.017 ± 2.339e − 3
No_sum_zero_shot timemixer	0.072 ± 2.877e − 3	0.012 ± 7.297e − 4	0.088 ± 8.869e − 4	0.016 ± 2.265e − 4	0.08 ± 1.347e − 3	0.014 ± 3.821e − 4
right lstm	0.049 ± 3.421e − 4	0.007 ± 5.62e − 5	0.067 ± 9.956e − 4	0.011 ± 1.228e − 4	0.047 ± 1.752e − 4	0.008 ± 2.646e − 5
right mlp	0.056 ± 1.771e − 4	0.008 ± 2.646e − 5	0.072 ± 1.945e − 4	0.012 ± 1.258e − 5	0.052 ± 2.052e − 4	0.008 ± 1.893e − 5
right patchtsmixer	0.109 ± 5.173e − 4	0.041 ± 3.377e − 4	0.11 ± 3.675e − 4	0.037 ± 5.371e − 4	0.083 ± 9.554e − 4	0.031 ± 6.364e − 4
right timemixer	0.156 ± 2.139e − 3	0.068 ± 1.209e − 3	0.134 ± 7.387e − 4	0.046 ± 9.389e − 4	0.228 ± 4.642e − 3	0.141 ± 3.671e − 3

Table 35: In-distribution results - Forecast. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	masked_mae	masked_mse	masked_mae	masked_mse
Llama-3.1 CoT +TFM lstm	0.233 ± 1.151e − 2	0.12 ± 8.28e − 3	0.263 ± 5.289e − 4	0.151 ± 2.991e − 3
Llama-3.1 CoT +TFM mlp	0.257 ± 1.285e − 3	0.138 ± 2.024e − 3	0.266 ± 8.977e − 4	0.155 ± 1.573e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.309 ± 1.042e − 2	0.209 ± 2.272e − 2	0.36 ± 6.736e − 3	0.267 ± 1.614e − 2
Llama-3.1 CoT +TFM timemixer	0.264 ± 5.936e − 3	0.141 ± 4.233e − 3	0.27 ± 1.086e − 3	0.155 ± 1.749e − 3
Llama-3.1 CoT lstm	0.249 ± 3.484e − 3	0.14 ± 4.978e − 3	0.265 ± 1.101e − 3	0.159 ± 1.279e − 3
Llama-3.1 CoT mlp	0.254 ± 1.175e − 3	0.144 ± 1.125e − 3	0.266 ± 4.836e − 4	0.16 ± 5.539e − 4
Llama-3.1 CoT patchtsmixer	0.301 ± 3.518e − 3	0.203 ± 1.327e − 2	0.35 ± 1.025e − 2	0.255 ± 7.716e − 3
Llama-3.1 CoT timemixer	0.257 ± 2.158e − 3	0.147 ± 1.634e − 3	0.273 ± 1.845e − 3	0.165 ± 1.135e − 3
Llama-3.1 ICD +TFM lstm	0.244 ± 3.883e − 3	0.129 ± 2.99e − 3	0.259 ± 6.222e − 3	0.146 ± 7.989e − 3
Llama-3.1 ICD +TFM mlp	0.255 ± 7.535e − 4	0.139 ± 8.615e − 4	0.262 ± 7.825e − 4	0.148 ± 1.253e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.302 ± 7.799e − 3	0.19 ± 1.528e − 2	0.368 ± 1.021e − 2	0.279 ± 2.077e − 2
Llama-3.1 ICD +TFM timemixer	0.258 ± 3.675e − 3	0.139 ± 2.935e − 3	0.268 ± 4.168e − 3	0.154 ± 6.667e − 3
Llama-3.1 ICD lstm	0.248 ± 6.559e − 4	0.14 ± 9.291e − 4	0.265 ± 8.688e − 4	0.16 ± 1.093e − 3
Llama-3.1 ICD mlp	0.248 ± 8.898e − 4	0.14 ± 9.803e − 4	0.267 ± 9.823e − 4	0.161 ± 9.952e − 4
Llama-3.1 ICD patchtsmixer	0.296 ± 4.522e − 3	0.191 ± 1.087e − 2	0.349 ± 5.127e − 3	0.259 ± 1.508e − 2
Llama-3.1 ICD timemixer	0.254 ± 1.418e − 3	0.144 ± 1.737e − 3	0.272 ± 1.191e − 3	0.165 ± 1.344e − 3
Llama-3.1 Trend +TFM lstm	0.241 ± 9.205e − 3	0.127 ± 8.831e − 3	0.265 ± 2.777e − 3	0.154 ± 4.278e − 3
Llama-3.1 Trend +TFM mlp	0.255 ± 2.103e − 3	0.136 ± 3.044e − 3	0.264 ± 1.337e − 3	0.152 ± 1.552e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.309 ± 2.944e − 3	0.212 ± 1.078e − 2	0.35 ± 6.284e − 3	0.244 ± 1.247e − 2
Llama-3.1 Trend +TFM timemixer	0.255 ± 3.634e − 3	0.135 ± 4.983e − 3	0.27 ± 7.901e − 4	0.153 ± 2.791e − 3
Llama-3.1 Trend lstm	0.249 ± 2.012e − 3	0.141 ± 2.458e − 3	0.265 ± 9.183e − 4	0.16 ± 6.172e − 4
Llama-3.1 Trend mlp	0.251 ± 6.926e − 4	0.142 ± 7.046e − 4	0.268 ± 1.058e − 3	0.161 ± 8.152e − 4
Llama-3.1 Trend patchtsmixer	0.299 ± 7.225e − 3	0.207 ± 2.734e − 2	0.343 ± 6.525e − 3	0.242 ± 8.377e − 3
Llama-3.1 Trend timemixer	0.254 ± 2.033e − 3	0.144 ± 1.564e − 3	0.273 ± 1.936e − 3	0.165 ± 1.44e − 3
Llama-3.1 zero_shot +TFM lstm	0.236 ± 9.481e − 3	0.122 ± 9.524e − 3	0.263 ± 2.426e − 3	0.151 ± 2.222e − 3
Llama-3.1 zero_shot +TFM mlp	0.254 ± 9.672e − 4	0.133 ± 2.668e − 3	0.259 ± 1.655e − 3	0.142 ± 3.406e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.307 ± 1.225e − 3	0.208 ± 5.803e − 3	0.359 ± 1.014e − 2	0.259 ± 1.957e − 2
Llama-3.1 zero_shot +TFM timemixer	0.258 ± 4.549e − 3	0.135 ± 3.992e − 3	0.27 ± 2.229e − 3	0.155 ± 3.77e − 3
Llama-3.1 zero_shot lstm	0.248 ± 2.096e − 3	0.139 ± 3.128e − 3	0.265 ± 8.146e − 4	0.16 ± 8.607e − 4
Llama-3.1 zero_shot mlp	0.252 ± 1.013e − 3	0.143 ± 1.042e − 3	0.266 ± 3.319e − 4	0.16 ± 4.706e − 4
Llama-3.1 zero_shot patchtsmixer	0.293 ± 3.942e − 3	0.183 ± 5.569e − 3	0.353 ± 4.407e − 3	0.254 ± 1.169e − 2
Llama-3.1 zero_shot timemixer	0.254 ± 1.517e − 3	0.143 ± 1.546e − 3	0.274 ± 3.347e − 3	0.165 ± 2.725e − 3
TFM lstm	0.256 ± 1.187e − 3	0.148 ± 8.907e − 4	0.271 ± 2.313e − 3	0.165 ± 1.201e − 3
TFM mlp	0.252 ± 2.536e − 4	0.142 ± 3.369e − 4	0.273 ± 7.091e − 4	0.167 ± 7.996e − 4
TFM patchtsmixer	0.313 ± 5.429e − 3	0.217 ± 1.902e − 2	0.365 ± 1.016e − 2	0.284 ± 2.297e − 2
TFM timemixer	0.255 ± 1.952e − 3	0.144 ± 2.339e − 3	0.272 ± 1.852e − 3	0.165 ± 1.846e − 3
TSDE lstm	0.441 ± 5.705e − 2	0.284 ± 7.197e − 2	0.317 ± 7.954e − 2	0.172 ± 7.826e − 2
TSDE mlp	0.329 ± 3.907e − 2	0.169 ± 3.646e − 2	0.425 ± 3.644e − 2	0.313 ± 4.37e − 2
TSDE patchtsmixer	0.311 ± 4.882e − 3	0.209 ± 9.028e − 3	0.361 ± 4.177e − 3	0.269 ± 1.2e − 2
TSDE timemixer	0.379 ± 4.678e − 2	0.23 ± 6.628e − 2	0.309 ± 1.989e − 2	0.167 ± 1.677e − 2
gemini-2.0-flash CoT +TFM lstm	0.245 ± 7.668e − 3	0.129 ± 8.157e − 3	0.264 ± 7.638e − 3	0.152 ± 1.099e − 2
gemini-2.0-flash CoT +TFM mlp	0.256 ± 2.95e − 4	0.137 ± 6.799e − 4	0.271 ± 5.314e − 4	0.158 ± 4.912e − 4
gemini-2.0-flash CoT +TFM patchtsmixer	0.303 ± 5.928e − 3	0.194 ± 1.536e − 2	0.35 ± 1.437e − 2	0.249 ± 1.319e − 2
gemini-2.0-flash CoT +TFM timemixer	0.262 ± 4.926e − 3	0.14 ± 1.162e − 3	0.271 ± 4.388e − 3	0.152 ± 3.192e − 3
gemini-2.0-flash CoT lstm	0.248 ± 2.401e − 3	0.137 ± 2.332e − 3	0.266 ± 6.472e − 4	0.16 ± 1.095e − 3
gemini-2.0-flash CoT mlp	0.251 ± 1.123e − 4	0.142 ± 1.458e − 4	0.267 ± 8.293e − 4	0.16 ± 8.07e − 4
gemini-2.0-flash CoT patchtsmixer	0.297 ± 1.961e − 3	0.194 ± 6.499e − 3	0.345 ± 8.73e − 3	0.243 ± 1.871e − 2
gemini-2.0-flash CoT timemixer	0.259 ± 1.743e − 3	0.147 ± 1.713e − 3	0.272 ± 3.148e − 3	0.163 ± 3.351e − 3
gemini-2.0-flash ICD +TFM lstm	0.242 ± 1.895e − 3	0.128 ± 1.697e − 3	0.263 ± 5.395e − 3	0.148 ± 6.053e − 3
gemini-2.0-flash ICD +TFM mlp	0.254 ± 1.227e − 3	0.131 ± 1.766e − 3	0.269 ± 1.092e − 3	0.156 ± 1.055e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.309 ± 2.339e − 3	0.202 ± 6.27e − 3	0.359 ± 1.048e − 2	0.258 ± 1.893e − 2
gemini-2.0-flash ICD +TFM timemixer	0.263 ± 1.018e − 3	0.136 ± 1.011e − 3	0.267 ± 5.24e − 3	0.149 ± 6.799e − 3
gemini-2.0-flash ICD lstm	0.247 ± 5.091e − 3	0.138 ± 6.117e − 3	0.263 ± 1.068e − 3	0.157 ± 1.731e − 3
gemini-2.0-flash ICD mlp	0.25 ± 1.183e − 3	0.14 ± 1.361e − 3	0.266 ± 8.483e − 4	0.159 ± 8.67e − 4
gemini-2.0-flash ICD patchtsmixer	0.299 ± 8.339e − 3	0.2 ± 1.925e − 2	0.342 ± 6.903e − 3	0.243 ± 1.72e − 2
gemini-2.0-flash ICD timemixer	0.259 ± 2.312e − 3	0.147 ± 2.326e − 3	0.272 ± 3.99e − 3	0.164 ± 3.513e − 3
gemini-2.0-flash Trend +TFM lstm	0.223 ± 2.934e − 3	0.112 ± 1.287e − 3	0.258 ± 5.156e − 3	0.144 ± 6.115e − 3
gemini-2.0-flash Trend +TFM mlp	0.255 ± 7.559e − 4	0.132 ± 2.128e − 3	0.265 ± 6.888e − 4	0.15 ± 1.064e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.304 ± 4.473e − 3	0.195 ± 9.066e − 3	0.362 ± 9.5e − 3	0.267 ± 1.868e − 2
gemini-2.0-flash Trend +TFM timemixer	0.259 ± 2.829e − 3	0.141 ± 5.123e − 3	0.271 ± 2.502e − 3	0.152 ± 4.551e − 3
gemini-2.0-flash Trend lstm	0.246 ± 2.201e − 3	0.138 ± 2.734e − 3	0.264 ± 6.985e − 4	0.158 ± 8.871e − 4
gemini-2.0-flash Trend mlp	0.248 ± 9.524e − 4	0.14 ± 9.351e − 4	0.268 ± 2.051e − 3	0.161 ± 1.963e − 3
gemini-2.0-flash Trend patchtsmixer	0.304 ± 2.698e − 3	0.216 ± 1.305e − 2	0.347 ± 1.166e − 2	0.244 ± 1.992e − 2
gemini-2.0-flash Trend timemixer	0.254 ± 1.467e − 3	0.144 ± 1.907e − 3	0.272 ± 2.424e − 3	0.164 ± 2.789e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.237 ± 8.483e − 3	0.123 ± 8.268e − 3	0.269 ± 7.437e − 3	0.157 ± 8.831e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.257 ± 2.318e − 4	0.14 ± 3.055e − 4	0.27 ± 1.445e − 3	0.158 ± 1.491e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.322 ± 1.483e − 3	0.237 ± 1.498e − 2	0.349 ± 2.874e − 3	0.241 ± 7.466e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.257 ± 2.596e − 3	0.137 ± 4.234e − 3	0.271 ± 1.195e − 3	0.152 ± 1.976e − 3
gemini-2.0-flash zero_shot lstm	0.25 ± 2.993e − 3	0.141 ± 3.455e − 3	0.267 ± 5.898e − 4	0.16 ± 5.229e − 4
gemini-2.0-flash zero_shot mlp	0.253 ± 1.152e − 3	0.144 ± 8.367e − 4	0.27 ± 1.568e − 3	0.163 ± 1.278e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.297 ± 8.402e − 3	0.193 ± 2.072e − 2	0.353 ± 1.37e − 2	0.261 ± 2.067e − 2
gemini-2.0-flash zero_shot timemixer	0.256 ± 6.253e − 4	0.146 ± 4.215e − 4	0.272 ± 1.866e − 3	0.164 ± 1.651e − 3
interp lstm	0.326 ± 7.263e − 3	0.217 ± 5.938e − 3	0.383 ± 4.215e − 3	0.276 ± 2.743e − 3
interp mlp	0.509 ± 5.545e − 2	0.466 ± 9.495e − 2	0.507 ± 1.317e − 2	0.427 ± 1.943e − 2
interp patchtsmixer	0.601 ± 4.662e − 2	24.8 ± 7.965e0	0.83 ± 3.299e − 2	13.411 ± 3.412e0
interp timemixer	0.59 ± 2.228e − 2	0.614 ± 7.488e − 2	0.56 ± 1.332e − 2	0.561 ± 4.048e − 2
mean lstm	0.297 ± 3.83e − 3	0.197 ± 1.58e − 3	0.38 ± 9.415e − 4	0.261 ± 1.091e − 3
mean mlp	0.412 ± 9.945e − 3	0.292 ± 9.692e − 3	0.402 ± 1.338e − 2	0.282 ± 1.586e − 2
mean patchtsmixer	0.259 ± 4.115e − 3	0.306 ± 1.698e − 2	0.448 ± 3.38e − 2	4.839 ± 1.345e0
mean timemixer	0.575 ± 3.447e − 3	0.541 ± 1.402e − 2	0.557 ± 7.886e − 3	0.51 ± 1.44e − 2
No_sum_CoT CoT +TFM lstm	0.241 ± 4.677e − 3	0.125 ± 5.111e − 3	0.263 ± 2.178e − 3	0.15 ± 2.378e − 3
No_sum_CoT CoT +TFM mlp	0.26 ± 3.366e − 3	0.144 ± 4.77e − 3	0.271 ± 3.026e − 3	0.155 ± 4.368e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.31 ± 6.41e − 3	0.208 ± 1.751e − 2	0.361 ± 1.783e − 3	0.269 ± 5.946e − 3
No_sum_CoT CoT +TFM timemixer	0.258 ± 5.222e − 3	0.133 ± 2.969e − 3	0.267 ± 6.478e − 3	0.145 ± 9.273e − 3
No_sum_CoT lstm	0.245 ± 1.906e − 3	0.138 ± 1.487e − 3	0.271 ± 2.812e − 3	0.163 ± 3.417e − 3
No_sum_CoT mlp	0.252 ± 9.53e − 4	0.145 ± 6.666e − 4	0.283 ± 2.47e − 3	0.174 ± 2.155e − 3
No_sum_CoT patchtsmixer	0.302 ± 3.796e − 3	0.214 ± 1.484e − 2	0.351 ± 1.451e − 2	0.25 ± 3.05e − 2
No_sum_CoT timemixer	0.251 ± 2.86e − 3	0.141 ± 2.135e − 3	0.274 ± 4.468e − 3	0.163 ± 5.064e − 3
No_sum_ICD ICD +TFM lstm	0.245 ± 2.003e − 3	0.132 ± 2.688e − 3	0.265 ± 3.819e − 3	0.152 ± 5.225e − 3
No_sum_ICD ICD +TFM mlp	0.264 ± 6.594e − 3	0.148 ± 7.287e − 3	0.277 ± 2.85e − 3	0.163 ± 2.906e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.311 ± 9.231e − 3	0.22 ± 3.153e − 2	0.353 ± 2.39e − 3	0.267 ± 1.882e − 2
No_sum_ICD ICD +TFM timemixer	0.258 ± 3.285e − 3	0.139 ± 5.799e − 3	0.272 ± 5.074e − 3	0.155 ± 8.076e − 3
No_sum_ICD lstm	0.243 ± 10e − 4	0.138 ± 1.574e − 3	0.272 ± 1.379e − 3	0.163 ± 9.315e − 4
No_sum_ICD mlp	0.248 ± 1.423e − 3	0.141 ± 1.563e − 3	0.276 ± 2.058e − 3	0.168 ± 1.711e − 3
No_sum_ICD patchtsmixer	0.3 ± 5.824e − 3	0.216 ± 2.673e − 2	0.356 ± 7.101e − 3	0.263 ± 1.785e − 2
No_sum_ICD timemixer	0.248 ± 2.058e − 3	0.14 ± 1.885e − 3	0.28 ± 4.925e − 3	0.17 ± 3.804e − 3
No_sum_Trend Trend +TFM lstm	0.243 ± 2.09e − 3	0.129 ± 3.814e − 3	0.266 ± 1.709e − 3	0.154 ± 8.954e − 4
No_sum_Trend Trend +TFM mlp	0.261 ± 2.61e − 3	0.145 ± 3.731e − 3	0.277 ± 3.419e − 3	0.163 ± 3.859e − 3
No_sum_Trend Trend +TFM patchtsmixer	0.308 ± 2.055e − 3	0.204 ± 4.42e − 3	0.362 ± 1.132e − 2	0.277 ± 3.123e − 2
No_sum_Trend Trend +TFM timemixer	0.262 ± 2.033e − 3	0.138 ± 4.333e − 3	0.272 ± 1.25e − 3	0.149 ± 1.171e − 3

Table 36: Cross-site transfer results - Forecast. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	masked_mae	masked_mse	masked_mae	masked_mse
No_sum_Trend lstm	0.249 ± 2.623e − 3	0.143 ± 2.817e − 3	0.271 ± 1.111e − 3	0.163 ± 1.111e − 3
No_sum_Trend mlp	0.253 ± 2.225e − 3	0.145 ± 2.026e − 3	0.284 ± 2.136e − 3	0.175 ± 1.803e − 3
No_sum_Trend patchtsmixer	0.293 ± 5.256e − 3	0.19 ± 1.041e − 2	0.345 ± 9.259e − 3	0.243 ± 1.885e − 2
No_sum_Trend timemixer	0.253 ± 2.217e − 3	0.144 ± 2.074e − 3	0.277 ± 2.751e − 3	0.169 ± 3.138e − 3
medgemma CoT +TFM lstm	0.24 ± 6.816e − 3	0.125 ± 8.342e − 3	0.269 ± 7.465e − 4	0.157 ± 3.317e − 3
medgemma CoT +TFM mlp	0.257 ± 1.504e − 3	0.137 ± 9.943e − 4	0.268 ± 1.119e − 3	0.154 ± 1.309e − 3
medgemma CoT +TFM patchtsmixer	0.305 ± 1.024e − 2	0.193 ± 2.069e − 2	0.365 ± 1.438e − 2	0.281 ± 3.208e − 2
medgemma CoT +TFM timemixer	0.262 ± 3.617e − 3	0.138 ± 2.563e − 3	0.267 ± 3.583e − 3	0.15 ± 5.676e − 3
medgemma CoT lstm	0.249 ± 7.646e − 4	0.141 ± 4.029e − 4	0.265 ± 1.296e − 3	0.16 ± 1.675e − 3
medgemma CoT mlp	0.251 ± 2.272e − 4	0.142 ± 1.186e − 4	0.267 ± 1.201e − 3	0.16 ± 1.024e − 3
medgemma CoT patchtsmixer	0.302 ± 5.299e − 3	0.2 ± 5.499e − 3	0.353 ± 9.822e − 3	0.261 ± 1.398e − 2
medgemma CoT timemixer	0.258 ± 2.634e − 3	0.147 ± 7.614e − 4	0.271 ± 1.594e − 3	0.163 ± 1.808e − 3
medgemma ICD +TFM lstm	0.242 ± 3.343e − 3	0.127 ± 5.492e − 3	0.268 ± 2.749e − 3	0.156 ± 3.411e − 3
medgemma ICD +TFM mlp	0.256 ± 1.712e − 3	0.137 ± 1.293e − 3	0.269 ± 5.604e − 4	0.156 ± 6.755e − 4
medgemma ICD +TFM patchtsmixer	0.309 ± 4.25e − 3	0.204 ± 1.313e − 2	0.358 ± 1.76e − 3	0.259 ± 3.414e − 3
medgemma ICD +TFM timemixer	0.264 ± 8.753e − 3	0.143 ± 3.938e − 3	0.265 ± 1.113e − 3	0.149 ± 3.272e − 3
medgemma ICD lstm	0.247 ± 1.964e − 3	0.138 ± 3.509e − 3	0.265 ± 1.511e − 3	0.159 ± 1.315e − 3
medgemma ICD mlp	0.249 ± 6.003e − 4	0.141 ± 6.605e − 4	0.267 ± 1.313e − 3	0.161 ± 1.376e − 3
medgemma ICD patchtsmixer	0.301 ± 8.083e − 3	0.204 ± 2.75e − 2	0.351 ± 1.059e − 2	0.257 ± 2.034e − 2
medgemma ICD timemixer	0.255 ± 2.058e − 3	0.145 ± 2.054e − 3	0.272 ± 2.285e − 3	0.164 ± 2.24e − 3
medgemma Trend +TFM lstm	0.245 ± 4.502e − 3	0.129 ± 5.574e − 3	0.268 ± 8.775e − 4	0.157 ± 3.522e − 3
medgemma Trend +TFM mlp	0.254 ± 4.188e − 4	0.133 ± 1.285e − 3	0.267 ± 1.403e − 3	0.154 ± 1.471e − 3
medgemma Trend +TFM patchtsmixer	0.312 ± 3.208e − 3	0.221 ± 1.173e − 2	0.352 ± 1.223e − 3	0.253 ± 1.052e − 2
medgemma Trend +TFM timemixer	0.265 ± 5.529e − 3	0.139 ± 1.168e − 3	0.271 ± 1.898e − 3	0.154 ± 4.041e − 3
medgemma Trend lstm	0.25 ± 1.1e − 3	0.142 ± 1.041e − 3	0.264 ± 4.763e − 4	0.158 ± 3.932e − 4
medgemma Trend mlp	0.251 ± 7.036e − 4	0.143 ± 5.268e − 4	0.267 ± 8.651e − 4	0.161 ± 6.607e − 4
medgemma Trend patchtsmixer	0.299 ± 6.688e − 4	0.195 ± 4.897e − 3	0.346 ± 1.116e − 2	0.246 ± 2.384e − 2
medgemma Trend timemixer	0.256 ± 1.805e − 3	0.146 ± 1.074e − 3	0.272 ± 2.892e − 3	0.164 ± 2.426e − 3
medgemma zero_shot +TFM lstm	0.25 ± 2.708e − 3	0.139 ± 4.651e − 3	0.264 ± 9.359e − 3	0.151 ± 9.994e − 3
medgemma zero_shot +TFM mlp	0.257 ± 1.904e − 3	0.138 ± 2.683e − 3	0.268 ± 9.304e − 4	0.155 ± 1.016e − 3
medgemma zero_shot +TFM patchtsmixer	0.309 ± 8.308e − 3	0.207 ± 1.436e − 2	0.356 ± 2.022e − 2	0.26 ± 2.808e − 2
medgemma zero_shot +TFM timemixer	0.259 ± 4.9e − 3	0.139 ± 7.448e − 3	0.271 ± 3.455e − 3	0.153 ± 4.72e − 3
medgemma zero_shot lstm	0.248 ± 1.353e − 3	0.141 ± 1.412e − 3	0.267 ± 1.275e − 3	0.16 ± 8.217e − 4
medgemma zero_shot mlp	0.249 ± 7.299e − 4	0.141 ± 5.935e − 4	0.267 ± 1.14e − 3	0.16 ± 1.277e − 3
medgemma zero_shot patchtsmixer	0.295 ± 4.016e − 3	0.192 ± 1.876e − 2	0.349 ± 7.039e − 3	0.247 ± 1.202e − 2
medgemma zero_shot timemixer	0.254 ± 4.558e − 3	0.144 ± 3.587e − 3	0.271 ± 2.25e − 3	0.163 ± 2.291e − 3
No_sum_zero_shot zero_shot +TFM lstm	0.245 ± 4.863e − 3	0.131 ± 6.715e − 3	0.264 ± 2.851e − 3	0.15 ± 4.289e − 3
No_sum_zero_shot zero_shot +TFM mlp	0.262 ± 4.131e − 3	0.146 ± 4.051e − 3	0.273 ± 3.846e − 3	0.158 ± 4.645e − 3
No_sum_zero_shot zero_shot +TFM patchtsmixer	0.304 ± 2.445e − 3	0.194 ± 9.183e − 3	0.356 ± 2.551e − 3	0.265 ± 1.22e − 2
No_sum_zero_shot zero_shot +TFM timemixer	0.26 ± 9.804e − 3	0.134 ± 3.106e − 3	0.273 ± 3.046e − 3	0.156 ± 4.517e − 3
No_sum_zero_shot lstm	0.246 ± 1.023e − 3	0.139 ± 4.635e − 4	0.27 ± 1.73e − 3	0.162 ± 2.37e − 3
No_sum_zero_shot mlp	0.251 ± 2.359e − 3	0.143 ± 1.861e − 3	0.282 ± 2.943e − 3	0.173 ± 2.678e − 3
No_sum_zero_shot patchtsmixer	0.296 ± 1.026e − 2	0.195 ± 2.108e − 2	0.351 ± 3.333e − 3	0.259 ± 2.272e − 2
No_sum_zero_shot timemixer	0.249 ± 3.272e − 3	0.14 ± 3.473e − 3	0.28 ± 2.941e − 3	0.169 ± 1.838e − 3
right lstm	0.32 ± 9.703e − 3	0.212 ± 5.96e − 3	0.379 ± 3.51e − 3	0.27 ± 2.655e − 3
right mlp	0.432 ± 3.094e − 2	0.317 ± 3.616e − 2	0.409 ± 9.342e − 3	0.28 ± 1.177e − 2
right patchtsmixer	0.303 ± 1.23e − 2	0.417 ± 4.725e − 2	0.547 ± 3.435e − 2	6.344 ± 1.676e0
right timemixer	0.537 ± 4.212e − 2	0.593 ± 2.114e − 1	0.487 ± 1.123e − 2	0.431 ± 2.56e − 2

Table 37: Cross-site transfer results - Forecast. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	mae	mse	mae	mse	mae	mse
Llama-3.1 CoT +TFM lstm	0.302 ± 1.48e − 2	0.379 ± 3.635e − 2	0.322 ± 8.197e − 3	0.567 ± 4.639e − 3	0.351 ± 4.62e − 3	0.937 ± 8.33e − 3
Llama-3.1 CoT +TFM mlp	0.304 ± 9.901e − 3	0.372 ± 1.076e − 2	0.304 ± 3.403e − 3	0.505 ± 1.562e − 2	0.343 ± 9.457e − 3	0.922 ± 1.08e − 2
Llama-3.1 CoT +TFM patchtsmixer	0.524 ± 2.093e − 2	0.691 ± 3.479e − 2	0.47 ± 2.87e − 2	0.766 ± 1.297e − 1	0.416 ± 9.236e − 3	1.011 ± 3.858e − 3
Llama-3.1 CoT +TFM timemixer	0.348 ± 3.617e − 3	0.38 ± 1.789e − 3	0.338 ± 1.84e − 2	0.519 ± 3.783e − 2	0.337 ± 5.955e − 3	0.924 ± 7.734e − 3
Llama-3.1 CoT lstm	0.378 ± 1.686e − 2	0.527 ± 1.747e − 2	0.396 ± 1.467e − 2	0.692 ± 9.725e − 3	0.358 ± 6.408e − 3	0.958 ± 7.056e − 3
Llama-3.1 CoT mlp	0.378 ± 1.342e − 2	0.537 ± 1.634e − 2	0.383 ± 7.868e − 3	0.679 ± 1.381e − 2	0.353 ± 5.766e − 3	0.957 ± 7.114e − 3
Llama-3.1 CoT patchtsmixer	0.45 ± 9.321e − 3	0.603 ± 1.17e − 2	0.436 ± 1.907e − 2	0.726 ± 1.54e − 2	0.371 ± 5.441e − 3	0.948 ± 4.508e − 3
Llama-3.1 CoT timemixer	0.428 ± 9.948e − 3	0.608 ± 1.768e − 2	0.4 ± 1.577e − 2	0.745 ± 4.069e − 2	0.348 ± 1.209e − 2	0.959 ± 3.893e − 3
Llama-3.1 ICD +TFM lstm	0.305 ± 3.202e − 3	0.38 ± 6.206e − 3	0.315 ± 7.487e − 3	0.54 ± 2.976e − 2	0.346 ± 1.467e − 3	0.93 ± 6.941e − 4
Llama-3.1 ICD +TFM mlp	0.291 ± 1.751e − 2	0.338 ± 4.194e − 2	0.307 ± 1.124e − 2	0.512 ± 2.612e − 2	0.338 ± 5.783e − 3	0.907 ± 2.877e − 2
Llama-3.1 ICD +TFM patchtsmixer	0.467 ± 6.059e − 2	0.549 ± 1.337e − 1	0.443 ± 3.829e − 2	0.667 ± 1.191e − 1	0.409 ± 7.221e − 3	0.998 ± 9.691e − 3
Llama-3.1 ICD +TFM timemixer	0.352 ± 2.81e − 3	0.393 ± 3.153e − 2	0.343 ± 9.235e − 3	0.544 ± 2.314e − 2	0.339 ± 8.263e − 3	0.931 ± 1.72e − 3
Llama-3.1 ICD lstm	0.373 ± 1.35e − 2	0.507 ± 1.465e − 2	0.396 ± 4.652e − 3	0.671 ± 1.096e − 2	0.343 ± 1.374e − 2	0.943 ± 1.831e − 2
Llama-3.1 ICD mlp	0.375 ± 7.018e − 3	0.519 ± 1.186e − 2	0.386 ± 1.069e − 2	0.662 ± 2.48e − 2	0.346 ± 1.124e − 2	0.945 ± 7.691e − 3
Llama-3.1 ICD patchtsmixer	0.44 ± 1.366e − 2	0.567 ± 2.372e − 2	0.414 ± 3.289e − 3	0.694 ± 1.099e − 2	0.363 ± 8.136e − 3	0.951 ± 7.857e − 3
Llama-3.1 ICD timemixer	0.415 ± 7.205e − 3	0.598 ± 1.578e − 2	0.394 ± 7.977e − 3	0.735 ± 1.667e − 2	0.346 ± 8.294e − 3	0.952 ± 5.613e − 3
Llama-3.1 Trend +TFM lstm	0.288 ± 4.85e − 3	0.357 ± 2.131e − 2	0.294 ± 9.581e − 3	0.519 ± 2.705e − 2	0.352 ± 1.563e − 3	0.94 ± 9.196e − 4
Llama-3.1 Trend +TFM mlp	0.291 ± 1.039e − 2	0.342 ± 2.252e − 2	0.304 ± 8e − 3	0.539 ± 1.464e − 2	0.354 ± 6.17e − 3	0.919 ± 2.627e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.488 ± 3.853e − 2	0.598 ± 6.89e − 2	0.469 ± 2.486e − 2	0.724 ± 7.989e − 2	0.414 ± 1.17e − 2	0.993 ± 4.875e − 3
Llama-3.1 Trend +TFM timemixer	0.335 ± 3.325e − 3	0.374 ± 9.667e − 3	0.33 ± 8.606e − 3	0.532 ± 1.417e − 2	0.341 ± 1.1e − 2	0.935 ± 1.23e − 2
Llama-3.1 Trend lstm	0.317 ± 7.011e − 3	0.417 ± 1.042e − 2	0.32 ± 7.124e − 3	0.587 ± 1.364e − 2	0.349 ± 4.507e − 3	0.944 ± 1.046e − 2
Llama-3.1 Trend mlp	0.313 ± 1.212e − 2	0.411 ± 2.107e − 2	0.311 ± 4.869e − 3	0.579 ± 1.221e − 2	0.351 ± 3.812e − 3	0.938 ± 1.169e − 2
Llama-3.1 Trend patchtsmixer	0.383 ± 9.965e − 3	0.473 ± 1.785e − 2	0.369 ± 7.19e − 3	0.616 ± 1.752e − 2	0.374 ± 9.925e − 3	0.943 ± 1.132e − 2
Llama-3.1 Trend timemixer	0.359 ± 7.483e − 3	0.479 ± 1.367e − 2	0.338 ± 6.444e − 3	0.625 ± 7.808e − 3	0.353 ± 6.539e − 3	0.952 ± 9.052e − 3
Llama-3.1 zero_shot +TFM lstm	0.31 ± 9.545e − 3	0.398 ± 2.841e − 2	0.32 ± 4.846e − 3	0.57 ± 1.255e − 2	0.342 ± 1.128e − 3	0.93 ± 1.504e − 4
Llama-3.1 zero_shot +TFM mlp	0.307 ± 5.449e − 3	0.383 ± 2.041e − 2	0.315 ± 7.42e − 3	0.545 ± 1.806e − 2	0.348 ± 8.502e − 3	0.917 ± 1.885e − 2
Llama-3.1 zero_shot +TFM patchtsmixer	0.489 ± 2.735e − 2	0.615 ± 6.752e − 2	0.398 ± 1.701e − 2	0.547 ± 3.993e − 2	0.408 ± 6.968e − 3	1.002 ± 2.556e − 3
Llama-3.1 zero_shot +TFM timemixer	0.343 ± 5.445e − 3	0.387 ± 1.905e − 2	0.348 ± 2.792e − 3	0.565 ± 2.677e − 2	0.333 ± 8.801e − 3	0.935 ± 9.816e − 3
Llama-3.1 zero_shot lstm	0.353 ± 7.614e − 3	0.493 ± 1.717e − 2	0.391 ± 1.006e − 2	0.701 ± 2.087e − 2	0.345 ± 6.4e − 3	0.944 ± 4.13e − 3
Llama-3.1 zero_shot mlp	0.36 ± 1.207e − 2	0.495 ± 1.645e − 2	0.377 ± 1.374e − 2	0.659 ± 5.663e − 3	0.338 ± 8.451e − 3	0.944 ± 3.195e − 3
Llama-3.1 zero_shot patchtsmixer	0.421 ± 4.832e − 3	0.549 ± 1.19e − 2	0.411 ± 1.348e − 2	0.709 ± 1.73e − 2	0.36 ± 6.219e − 3	0.947 ± 2.512e − 3
Llama-3.1 zero_shot timemixer	0.397 ± 1.102e − 2	0.568 ± 2.333e − 2	0.392 ± 1.076e − 2	0.707 ± 6.456e − 3	0.333 ± 8.444e − 3	0.95 ± 9.248e − 3
TFM lstm	0.313 ± 6.465e − 3	0.438 ± 6.413e − 3	0.322 ± 4.787e − 3	0.823 ± 1.426e − 3	0.419 ± 1.588e − 2	0.889 ± 1.232e − 2
TFM mlp	0.319 ± 9.517e − 3	0.442 ± 5.719e − 3	0.336 ± 7.05e − 3	0.837 ± 3.961e − 3	0.416 ± 7.481e − 3	0.9 ± 9.927e − 3
TFM patchtsmixer	0.44 ± 3.117e − 2	0.565 ± 4.867e − 2	0.413 ± 8.228e − 3	0.882 ± 1.314e − 2	0.494 ± 1.073e − 2	0.957 ± 2.818e − 2
TFM timemixer	0.345 ± 2.447e − 2	0.49 ± 5.165e − 2	0.355 ± 1.015e − 2	0.848 ± 4.638e − 3	0.418 ± 5.579e − 3	0.89 ± 1.031e − 2
TSDE lstm	0.293 ± 7.301e − 3	0.382 ± 4.77e − 3	0.325 ± 1.299e − 2	0.743 ± 1.35e − 2	0.399 ± 5.013e − 3	0.806 ± 4.848e − 3
TSDE mlp	0.298 ± 6.89e − 3	0.378 ± 5.964e − 3	0.326 ± 1.834e − 2	0.729 ± 4.613e − 3	0.412 ± 6.807e − 3	0.806 ± 6.799e − 3
TSDE patchtsmixer	0.411 ± 2.723e − 2	0.484 ± 5.127e − 2	0.406 ± 1.342e − 2	0.801 ± 2.698e − 2	0.485 ± 5.896e − 3	0.877 ± 1.835e − 2
TSDE timemixer	0.331 ± 1.209e − 2	0.403 ± 1.16e − 2	0.343 ± 1.109e − 2	0.742 ± 8.187e − 3	0.4 ± 9.397e − 3	0.821 ± 2.762e − 2
gemini-2.0-flash CoT +TFM lstm	0.311 ± 5.765e − 3	0.401 ± 1.682e − 2	0.299 ± 4.825e − 3	0.501 ± 2.849e − 3	0.346 ± 1.309e − 3	0.934 ± 5.717e − 4
gemini-2.0-flash CoT +TFM mlp	0.293 ± 1.353e − 2	0.341 ± 2.886e − 2	0.311 ± 1.27e − 2	0.519 ± 3.552e − 2	0.353 ± 6.034e − 3	0.916 ± 1.346e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.515 ± 5.591e − 2	0.648 ± 1.144e − 1	0.473 ± 2.76e − 2	0.744 ± 7.319e − 2	0.417 ± 2.119e − 3	0.994 ± 4.372e − 3
gemini-2.0-flash CoT +TFM timemixer	0.342 ± 2.814e − 3	0.392 ± 7.927e − 3	0.353 ± 1.326e − 3	0.537 ± 3.793e − 2	0.333 ± 2.434e − 3	0.916 ± 3.93e − 3
gemini-2.0-flash CoT lstm	0.364 ± 9.354e − 3	0.503 ± 1.154e − 2	0.364 ± 5.208e − 3	0.641 ± 6.085e − 3	0.348 ± 2.818e − 3	0.944 ± 5.052e − 3
gemini-2.0-flash CoT mlp	0.375 ± 8.365e − 3	0.514 ± 1.245e − 2	0.358 ± 1.419e − 2	0.614 ± 1.426e − 2	0.345 ± 7.564e − 3	0.943 ± 8.451e − 3
gemini-2.0-flash CoT patchtsmixer	0.438 ± 9.928e − 3	0.563 ± 5.535e − 3	0.396 ± 2.093e − 3	0.646 ± 4.094e − 2	0.363 ± 6.436e − 3	0.935 ± 1.274e − 2
gemini-2.0-flash CoT timemixer	0.398 ± 3.311e − 3	0.555 ± 9.959e − 3	0.366 ± 1.037e − 2	0.651 ± 3.209e − 2	0.341 ± 7.12e − 3	0.943 ± 5.792e − 3
gemini-2.0-flash ICD +TFM lstm	0.314 ± 1.145e − 2	0.406 ± 2.45e − 2	0.304 ± 1.338e − 2	0.526 ± 4.775e − 2	0.347 ± 2.497e − 3	0.932 ± 1.617e − 3
gemini-2.0-flash ICD +TFM mlp	0.295 ± 1.323e − 2	0.341 ± 3.493e − 2	0.309 ± 7.168e − 3	0.511 ± 3.592e − 2	0.353 ± 5.594e − 3	0.916 ± 1.94e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.499 ± 3.616e − 2	0.626 ± 8.963e − 2	0.445 ± 3.464e − 2	0.661 ± 1.563e − 1	0.409 ± 5.918e − 3	0.998 ± 8.959e − 3
gemini-2.0-flash ICD +TFM timemixer	0.351 ± 8.9e − 3	0.395 ± 1.877e − 2	0.342 ± 3.224e − 2	0.508 ± 4.002e − 2	0.331 ± 1.488e − 2	0.937 ± 3.433e − 3
gemini-2.0-flash ICD lstm	0.377 ± 7.773e − 3	0.51 ± 1.163e − 2	0.379 ± 9.69e − 3	0.655 ± 2.066e − 2	0.345 ± 4.67e − 3	0.943 ± 6.364e − 3
gemini-2.0-flash ICD mlp	0.37 ± 6.721e − 3	0.508 ± 8.322e − 3	0.384 ± 1.658e − 2	0.666 ± 2.546e − 2	0.347 ± 7.267e − 3	0.932 ± 1.746e − 2
gemini-2.0-flash ICD patchtsmixer	0.425 ± 1.65e − 2	0.554 ± 2.298e − 2	0.436 ± 5.942e − 3	0.72 ± 1.483e − 2	0.368 ± 1.732e − 2	0.943 ± 1.599e − 2
gemini-2.0-flash ICD timemixer	0.412 ± 4.38e − 3	0.578 ± 1.692e − 2	0.395 ± 1.056e − 2	0.709 ± 1.757e − 2	0.337 ± 1.046e − 2	0.943 ± 4.178e − 3
gemini-2.0-flash Trend +TFM lstm	0.301 ± 4.708e − 3	0.384 ± 1.742e − 2	0.307 ± 6.442e − 3	0.55 ± 2.156e − 2	0.344 ± 6.558e − 3	0.92 ± 8.059e − 3
gemini-2.0-flash Trend +TFM mlp	0.305 ± 7.051e − 3	0.385 ± 3.715e − 3	0.307 ± 1.223e − 2	0.523 ± 2.544e − 2	0.347 ± 6.139e − 3	0.903 ± 1.67e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.492 ± 6.942e − 2	0.6 ± 1.178e − 1	0.491 ± 3.313e − 2	0.798 ± 9.88e − 2	0.41 ± 4.623e − 3	0.999 ± 7.724e − 3
gemini-2.0-flash Trend +TFM timemixer	0.336 ± 1.257e − 3	0.396 ± 2.114e − 2	0.337 ± 7.226e − 3	0.54 ± 1.294e − 2	0.338 ± 4.963e − 3	0.928 ± 9.852e − 4
gemini-2.0-flash Trend lstm	0.321 ± 4.403e − 3	0.436 ± 4.094e − 3	0.324 ± 8.987e − 3	0.603 ± 6.258e − 3	0.343 ± 5.967e − 3	0.937 ± 7.98e − 4
gemini-2.0-flash Trend mlp	0.333 ± 5.068e − 3	0.453 ± 6.094e − 3	0.32 ± 4.738e − 3	0.592 ± 1.639e − 2	0.348 ± 1.258e − 2	0.929 ± 1.13e − 2
gemini-2.0-flash Trend patchtsmixer	0.379 ± 3.72e − 3	0.479 ± 3.638e − 3	0.37 ± 5.582e − 3	0.615 ± 9.403e − 3	0.366 ± 8.984e − 3	0.934 ± 9.468e − 3
gemini-2.0-flash Trend timemixer	0.35 ± 8.609e − 3	0.471 ± 1.277e − 2	0.349 ± 1.801e − 2	0.615 ± 1.997e − 2	0.326 ± 5.127e − 3	0.921 ± 6.568e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.31 ± 6.208e − 3	0.399 ± 1.677e − 2	0.308 ± 8.873e − 3	0.556 ± 1.445e − 2	0.348 ± 3.44e − 3	0.93 ± 5.467e − 4
gemini-2.0-flash zero_shot +TFM mlp	0.299 ± 4.766e − 3	0.351 ± 4.897e − 3	0.302 ± 7.655e − 3	0.52 ± 3.221e − 2	0.344 ± 1.348e − 2	0.891 ± 3.461e − 2
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.516 ± 4.704e − 2	0.659 ± 9.077e − 2	0.438 ± 3.328e − 2	0.635 ± 1.248e − 1	0.407 ± 1.049e − 2	0.975 ± 1.336e − 2
gemini-2.0-flash zero_shot +TFM timemixer	0.339 ± 6.364e − 3	0.391 ± 3.247e − 2	0.333 ± 4.202e − 3	0.544 ± 9.495e − 3	0.325 ± 8.249e − 3	0.944 ± 1.064e − 2
gemini-2.0-flash zero_shot lstm	0.343 ± 6.846e − 3	0.464 ± 5.354e − 3	0.338 ± 2.042e − 3	0.597 ± 6.937e − 3	0.352 ± 1.166e − 2	0.936 ± 9.862e − 3
gemini-2.0-flash zero_shot mlp	0.347 ± 8.917e − 3	0.459 ± 1.166e − 2	0.339 ± 3.727e − 3	0.59 ± 8.758e − 3	0.347 ± 1.1e − 2	0.941 ± 5.464e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.414 ± 1.191e − 2	0.509 ± 2.031e − 2	0.378 ± 1.284e − 2	0.631 ± 2.204e − 2	0.364 ± 6.278e − 3	0.922 ± 8.645e − 3
gemini-2.0-flash zero_shot timemixer	0.378 ± 1.103e − 2	0.51 ± 2.382e − 2	0.35 ± 1.269e − 2	0.624 ± 1.051e − 2	0.347 ± 8.378e − 3	0.939 ± 8.521e − 3
interp lstm	0.29 ± 1.648e − 3	0.38 ± 6.123e − 3	0.309 ± 1.547e − 2	0.756 ± 4.973e − 3	0.412 ± 5.721e − 3	0.861 ± 3.341e − 2
interp mlp	0.301 ± 7.762e − 3	0.366 ± 5.698e − 3	0.332 ± 7.038e − 3	0.775 ± 3.284e − 3	0.425 ± 1.774e − 2	0.79 ± 1.397e − 2
interp patchtsmixer	0.37 ± 1.346e − 3	0.46 ± 1.034e − 2	0.43 ± 1.757e − 2	0.914 ± 1.501e − 2	0.493 ± 9.374e − 3	0.891 ± 2.526e − 2
interp timemixer	0.337 ± 1.468e − 2	0.426 ± 1.816e − 3	0.373 ± 9.808e − 3	0.86 ± 9.56e − 3	0.437 ± 1.081e − 2	0.889 ± 2.57e − 2
mean lstm	0.278 ± 6.395e − 3	0.364 ± 5.147e − 3	0.336 ± 1.411e − 2	0.786 ± 9.499e − 3	0.402 ± 9.46e − 3	0.858 ± 2.289e − 2
mean mlp	0.289 ± 7.444e − 3	0.357 ± 2.237e − 3	0.343 ± 7.519e − 3	0.822 ± 6.117e − 3	0.393 ± 2.354e − 3	0.801 ± 1.183e − 2
mean patchtsmixer	0.38 ± 4.793e − 3	0.486 ± 1.508e − 2	0.447 ± 9.717e − 3	0.928 ± 9.854e − 3	0.528 ± 8.409e − 3	0.954 ± 1.272e − 2
mean timemixer	0.328 ± 8.1e − 3	0.409 ± 3.781e − 3	0.357 ± 8.922e − 3	0.82 ± 1.152e − 2	0.439 ± 1.902e − 2	0.848 ± 1.813e − 2
No_sum_CoT CoT +TFM lstm	0.29 ± 7.114e − 3	0.369 ± 1.166e − 2	0.281 ± 1.775e − 3	0.532 ± 2.267e − 3	0.327 ± 8.959e − 4	0.908 ± 1.299e − 3
No_sum_CoT CoT +TFM mlp	0.283 ± 9.061e − 3	0.337 ± 1.212e − 2	0.265 ± 2.853e − 3	0.45 ± 1.279e − 2	0.332 ± 5.21e − 3	0.849 ± 2.953e − 3
No_sum_CoT CoT +TFM patchtsmixer	0.5 ± 4.685e − 2	0.628 ± 1.227e − 1	0.44 ± 4.554e − 2	0.686 ± 1.143e − 1	0.423 ± 1.273e − 2	0.986 ± 1.659e − 2
No_sum_CoT CoT +TFM timemixer	0.324 ± 4.28e − 3	0.377 ± 7.914e − 3	0.31 ± 6.657e − 3	0.545 ± 1.486e − 2	0.336 ± 1.661e − 2	0.894 ± 1.348e − 2
No_sum_CoT lstm	0.301 ± 5.112e − 3	0.4 ± 1.102e − 2	0.288 ± 1.314e − 2	0.532 ± 1.009e − 2	0.34 ± 8.861e − 3	0.919 ± 7.784e − 3
No_sum_CoT mlp	0.299 ± 3.81e − 3	0.396 ± 7.231e − 3	0.28 ± 3.344e − 3	0.537 ± 7.668e − 3	0.342 ± 4.967e − 3	0.917 ± 1.435e − 2
No_sum_CoT patchtsmixer	0.353 ± 1.188e − 2	0.432 ± 1.909e − 2	0.328 ± 7.745e − 3	0.562 ± 1.088e − 2	0.364 ± 9.251e − 3	0.927 ± 2.058e − 2
No_sum_CoT timemixer	0.338 ± 1.152e − 2	0.436 ± 7.67e − 3	0.308 ± 1.879e − 2	0.56 ± 2.424e − 2	0.321 ± 6.086e − 3	0.92 ± 7.261e − 3
No_sum_ICD ICD +TFM lstm	0.256 ± 1.28e − 2	0.284 ± 2.642e − 2	0.274 ± 1.719e − 3	0.494 ± 1.944e − 2	0.324 ± 3.583e − 3	0.905 ± 2.221e − 3
No_sum_ICD ICD +TFM mlp	0.269 ± 9.898e − 3	0.309 ± 1.739e − 2	0.249 ± 1.39e − 2	0.39 ± 8.124e − 2	0.331 ± 1.195e − 2	0.846 ± 2.891e − 3
No_sum_ICD ICD +TFM patchtsmixer	0.506 ± 5.949e − 2	0.628 ± 1.194e − 1	0.448 ± 4.995e − 2	0.701 ± 1.086e − 1	0.428 ± 1.259e − 2	0.999 ± 1.091e − 2
No_sum_ICD ICD +TFM timemixer	0.324 ± 6.1e − 4	0.37 ± 1.659e − 2	0.313 ± 1.024e − 2	0.522 ± 1.691e − 2	0.338 ± 1.838e − 2	0.909 ± 4.097e − 3
No_sum_ICD lstm	0.297 ± 5.216e − 3	0.386 ± 6.314e − 3	0.298 ± 1.215e − 2	0.546 ± 1.543e − 2	0.336 ± 1.041e − 2	0.918 ± 5.768e − 3
No_sum_ICD mlp	0.309 ± 4.016e − 3	0.411 ± 6.651e − 3	0.291 ± 8.175e − 3	0.547 ± 7.513e − 3	0.348 ± 1.28e − 2	0.915 ± 9.047e − 3
No_sum_ICD patchtsmixer	0.347 ± 5.559e − 3	0.432 ± 7.443e − 3	0.339 ± 9.858e − 3	0.565 ± 3.923e − 3	0.364 ± 7.45e − 3	0.915 ± 1.019e − 2
No_sum_ICD timemixer	0.362 ± 5.646e − 3	0.459 ± 1.286e − 2	0.32 ± 6.855e − 3	0.574 ± 8.25e − 3	0.321 ± 1.326e − 2	0.931 ± 5.679e − 3
No_sum_Trend Trend +TFM lstm	0.279 ± 9.944e − 3	0.343 ± 1.686e − 2	0.283 ± 8.561e − 3	0.53 ± 1.1e − 2	0.329 ± 1.937e − 3	0.908 ± 9.878e − 4
No_sum_Trend Trend +TFM mlp	0.271 ± 1.802e − 2	0.307 ± 4.645e − 2	0.267 ± 5.543e − 3	0.469 ± 2.384e − 2	0.33 ± 7.231e − 3	0.819 ± 5.575e − 2
No_sum_Trend Trend +TFM patchtsmixer	0.522 ± 7.286e − 2	0.65 ± 1.428e − 1	0.478 ± 5.07e − 2	0.731 ± 8.215e − 2	0.404 ± 8.409e − 3	0.985 ± 1.298e − 2
No_sum_Trend Trend +TFM timemixer	0.321 ± 6.699e − 3	0.349 ± 1.516e − 2	0.307 ± 4.901e − 3	0.531 ± 1.704e − 2	0.347 ± 7.004e − 3	0.904 ± 1.368e − 3

Table 38: In-distribution results - Los. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	mae	mse	mae	mse	mae	mse
No_sum_Trend lstm	$0.301 \pm 6.829e-3$	$0.399 \pm 1.76e-2$	$0.283 \pm 5.464e-3$	$0.538 \pm 7.988e-3$	$0.339 \pm 5.337e-3$	$0.92 \pm 6.172e-3$
No_sum_Trend mlp	$0.3 \pm 9.631e-3$	$0.386 \pm 2.206e-2$	$0.294 \pm 1.381e-2$	$0.539 \pm 1.816e-2$	$0.343 \pm 1.163e-2$	$0.929 \pm 2.489e-3$
No_sum_Trend patchtsmixer	$0.348 \pm 3.213e-3$	$0.428 \pm 8.989e-3$	$0.329 \pm 9.072e-3$	$0.563 \pm 6.203e-3$	$0.37 \pm 1.121e-2$	$0.93 \pm 1.805e-2$
No_sum_Trend timemixer	$0.363 \pm 1.044e-2$	$0.457 \pm 1.057e-2$	$0.326 \pm 1.872e-2$	$0.575 \pm 8.601e-3$	$0.325 \pm 9.516e-3$	$0.916 \pm 6.562e-3$
medgemma CoT +TFM lstm	$0.312 \pm 6.866e-3$	$0.405 \pm 2.334e-2$	$0.312 \pm 1.294e-2$	$0.569 \pm 2.844e-2$	$0.346 \pm 3.115e-3$	$0.929 \pm 4.153e-4$
medgemma CoT +TFM mlp	$0.299 \pm 1.399e-2$	$0.355 \pm 3.591e-2$	$0.288 \pm 9.295e-3$	$0.486 \pm 3.863e-2$	$0.339 \pm 2.546e-3$	$0.912 \pm 1.611e-3$
medgemma CoT +TFM patchtsmixer	$0.485 \pm 3.325e-2$	$0.601 \pm 9.309e-2$	$0.403 \pm 1.52e-2$	$0.543 \pm 1.788e-2$	$0.412 \pm 6.571e-3$	$0.993 \pm 1.227e-2$
medgemma CoT +TFM timemixer	$0.344 \pm 5.178e-3$	$0.405 \pm 1.659e-2$	$0.339 \pm 2.904e-3$	$0.542 \pm 7.279e-3$	$0.32 \pm 6.392e-3$	$0.933 \pm 1.025e-2$
medgemma CoT lstm	$0.341 \pm 4.005e-3$	$0.463 \pm 7.062e-3$	$0.332 \pm 7.19e-3$	$0.603 \pm 8.312e-3$	$0.347 \pm 5.449e-3$	$0.939 \pm 6.47e-3$
medgemma CoT mlp	$0.343 \pm 6.417e-3$	$0.477 \pm 1.229e-2$	$0.333 \pm 1.285e-2$	$0.609 \pm 8.441e-3$	$0.343 \pm 5.861e-3$	$0.937 \pm 5.103e-3$
medgemma CoT patchtsmixer	$0.408 \pm 8.074e-3$	$0.518 \pm 2.414e-2$	$0.386 \pm 3.992e-3$	$0.647 \pm 9.512e-3$	$0.359 \pm 3.952e-3$	$0.929 \pm 8.375e-3$
medgemma CoT timemixer	$0.387 \pm 1.108e-2$	$0.529 \pm 2.851e-2$	$0.347 \pm 7.617e-3$	$0.615 \pm 1.594e-2$	$0.336 \pm 7.898e-3$	$0.946 \pm 2.82e-3$
medgemma ICD +TFM lstm	$0.319 \pm 1.329e-2$	$0.422 \pm 2.758e-2$	$0.307 \pm 6.127e-3$	$0.528 \pm 2.618e-2$	$0.348 \pm 2.002e-3$	$0.935 \pm 6.341e-4$
medgemma ICD +TFM mlp	$0.322 \pm 3.012e-3$	$0.421 \pm 1.392e-2$	$0.3 \pm 1.038e-2$	$0.498 \pm 3.683e-2$	$0.349 \pm 6.55e-3$	$0.916 \pm 1.648e-3$
medgemma ICD +TFM patchtsmixer	$0.484 \pm 4.294e-2$	$0.587 \pm 9.281e-2$	$0.492 \pm 2.757e-2$	$0.799 \pm 6.441e-2$	$0.415 \pm 5.771e-3$	$1.012 \pm 9.506e-3$
medgemma ICD +TFM timemixer	$0.345 \pm 6.85e-3$	$0.405 \pm 2.674e-2$	$0.343 \pm 3.796e-3$	$0.543 \pm 1.663e-2$	$0.344 \pm 6.094e-3$	$0.949 \pm 8.038e-3$
medgemma ICD lstm	$0.355 \pm 2.561e-3$	$0.494 \pm 1.149e-2$	$0.356 \pm 6.237e-3$	$0.609 \pm 1.937e-2$	$0.35 \pm 9.343e-3$	$0.938 \pm 6.572e-3$
medgemma ICD mlp	$0.352 \pm 5.407e-3$	$0.499 \pm 1.063e-2$	$0.349 \pm 6.459e-3$	$0.618 \pm 1.178e-2$	$0.349 \pm 6.77e-3$	$0.944 \pm 8.925e-3$
medgemma ICD patchtsmixer	$0.424 \pm 1.153e-2$	$0.549 \pm 1.159e-2$	$0.399 \pm 2.139e-2$	$0.653 \pm 3.121e-2$	$0.368 \pm 1.218e-2$	$0.942 \pm 1.61e-2$
medgemma ICD timemixer	$0.396 \pm 4.963e-3$	$0.549 \pm 8.559e-3$	$0.381 \pm 6.995e-3$	$0.679 \pm 1.414e-2$	$0.334 \pm 9.54e-3$	$0.934 \pm 2.482e-3$
medgemma Trend +TFM lstm	$0.291 \pm 1.051e-2$	$0.366 \pm 1.907e-2$	$0.301 \pm 1.494e-2$	$0.528 \pm 3.878e-2$	$0.346 \pm 5.352e-4$	$0.931 \pm 8.166e-4$
medgemma Trend +TFM mlp	$0.287 \pm 1.007e-2$	$0.346 \pm 2.704e-2$	$0.281 \pm 6.273e-3$	$0.474 \pm 1.746e-2$	$0.344 \pm 1.023e-2$	$0.902 \pm 1.618e-2$
medgemma Trend +TFM patchtsmixer	$0.512 \pm 3.909e-2$	$0.657 \pm 8.499e-2$	$0.456 \pm 3.573e-2$	$0.727 \pm 1.408e-1$	$0.412 \pm 3.71e-3$	$0.999 \pm 2.558e-3$
medgemma Trend +TFM timemixer	$0.345 \pm 5.339e-3$	$0.412 \pm 3.851e-3$	$0.33 \pm 1.176e-2$	$0.508 \pm 2.546e-2$	$0.333 \pm 1.622e-2$	$0.93 \pm 5.742e-3$
medgemma Trend lstm	$0.316 \pm 2.651e-3$	$0.437 \pm 3.236e-3$	$0.316 \pm 7.792e-3$	$0.601 \pm 1.346e-2$	$0.35 \pm 8.86e-3$	$0.94 \pm 6.694e-3$
medgemma Trend mlp	$0.321 \pm 2.304e-3$	$0.433 \pm 9.588e-3$	$0.322 \pm 8.1e-3$	$0.599 \pm 1.109e-2$	$0.342 \pm 9.51e-3$	$0.933 \pm 7.875e-3$
medgemma Trend patchtsmixer	$0.383 \pm 9.782e-3$	$0.478 \pm 8.87e-3$	$0.37 \pm 1.515e-2$	$0.637 \pm 2.943e-2$	$0.36 \pm 9.033e-3$	$0.935 \pm 1.177e-2$
medgemma Trend timemixer	$0.357 \pm 6.686e-3$	$0.481 \pm 8.074e-3$	$0.349 \pm 2.561e-2$	$0.636 \pm 2.673e-2$	$0.33 \pm 6.434e-3$	$0.935 \pm 9.01e-3$
medgemma zero_shot +TFM lstm	$0.306 \pm 1.331e-2$	$0.391 \pm 3.787e-2$	$0.31 \pm 9.843e-3$	$0.555 \pm 1.442e-2$	$0.343 \pm 7.858e-4$	$0.932 \pm 1.795e-3$
medgemma zero_shot +TFM mlp	$0.291 \pm 6.487e-3$	$0.333 \pm 1.188e-2$	$0.304 \pm 1.345e-2$	$0.53 \pm 3.875e-2$	$0.343 \pm 1.195e-2$	$0.902 \pm 2.005e-2$
medgemma zero_shot +TFM patchtsmixer	$0.516 \pm 6.151e-2$	$0.631 \pm 1.278e-1$	$0.452 \pm 2.158e-2$	$0.706 \pm 7.779e-2$	$0.411 \pm 1.256e-2$	$0.995 \pm 1.312e-2$
medgemma zero_shot +TFM timemixer	$0.345 \pm 2.781e-3$	$0.412 \pm 1.626e-2$	$0.33 \pm 4.14e-3$	$0.553 \pm 1.265e-2$	$0.335 \pm 6.177e-3$	$0.918 \pm 1.864e-2$
medgemma zero_shot lstm	$0.345 \pm 8.007e-3$	$0.469 \pm 5.448e-3$	$0.348 \pm 4.531e-3$	$0.608 \pm 1.119e-2$	$0.343 \pm 4.79e-3$	$0.933 \pm 3.1e-3$
medgemma zero_shot mlp	$0.35 \pm 8.235e-3$	$0.473 \pm 2.061e-2$	$0.343 \pm 9.584e-3$	$0.596 \pm 9.739e-3$	$0.341 \pm 8.908e-3$	$0.937 \pm 1.26e-2$
medgemma zero_shot patchtsmixer	$0.422 \pm 8.425e-3$	$0.53 \pm 3.072e-2$	$0.384 \pm 6.472e-3$	$0.633 \pm 1.235e-2$	$0.364 \pm 4.605e-3$	$0.927 \pm 8.179e-3$
medgemma zero_shot timemixer	$0.399 \pm 6.156e-3$	$0.557 \pm 1.604e-2$	$0.359 \pm 1.415e-2$	$0.622 \pm 2.514e-2$	$0.333 \pm 9.835e-3$	$0.935 \pm 4.769e-3$
No_sum_zero_shot zero_shot +TFM lstm	$0.297 \pm 5.776e-3$	$0.373 \pm 1.361e-2$	$0.282 \pm 8.444e-3$	$0.521 \pm 2.444e-2$	$0.327 \pm 1.642e-3$	$0.906 \pm 2.524e-3$
No_sum_zero_shot zero_shot +TFM mlp	$0.267 \pm 1.307e-2$	$0.308 \pm 3.37e-2$	$0.272 \pm 1.423e-2$	$0.478 \pm 2.572e-2$	$0.333 \pm 4.563e-3$	$0.847 \pm 4.354e-3$
No_sum_zero_shot zero_shot +TFM patchtsmixer	$0.511 \pm 6.612e-2$	$0.648 \pm 1.607e-1$	$0.406 \pm 1.438e-2$	$0.597 \pm 4.365e-2$	$0.415 \pm 1.439e-2$	$0.986 \pm 1.714e-2$
No_sum_zero_shot zero_shot +TFM timemixer	$0.324 \pm 4.604e-3$	$0.395 \pm 1.705e-2$	$0.31 \pm 3.484e-3$	$0.52 \pm 1.671e-2$	$0.323 \pm 3.841e-3$	$0.897 \pm 1.225e-2$
No_sum_zero_shot lstm	$0.313 \pm 5.513e-3$	$0.419 \pm 1.07e-2$	$0.28 \pm 1.149e-2$	$0.534 \pm 1.308e-2$	$0.337 \pm 6.522e-3$	$0.921 \pm 5.607e-3$
No_sum_zero_shot mlp	$0.303 \pm 4.134e-3$	$0.398 \pm 8.519e-3$	$0.281 \pm 1.109e-2$	$0.528 \pm 9.66e-3$	$0.338 \pm 1.255e-2$	$0.92 \pm 4.317e-3$
No_sum_zero_shot patchtsmixer	$0.364 \pm 1.151e-2$	$0.445 \pm 1.28e-2$	$0.331 \pm 1.281e-2$	$0.562 \pm 1.346e-2$	$0.358 \pm 9.935e-3$	$0.914 \pm 1.942e-2$
No_sum_zero_shot timemixer	$0.348 \pm 1.641e-2$	$0.451 \pm 1.311e-2$	$0.313 \pm 1.207e-2$	$0.566 \pm 9.761e-3$	$0.324 \pm 6.311e-3$	$0.917 \pm 4.715e-3$
right lstm	$0.28 \pm 7.312e-3$	$0.36 \pm 6.177e-3$	$0.299 \pm 1.479e-2$	$0.751 \pm 8.077e-3$	$0.41 \pm 1.044e-2$	$0.847 \pm 7.172e-3$
right mlp	$0.276 \pm 4.021e-3$	$0.341 \pm 1.953e-3$	$0.295 \pm 1.189e-2$	$0.752 \pm 4.509e-3$	$0.408 \pm 1.236e-2$	$0.786 \pm 1.062e-2$
right patchtsmixer	$0.342 \pm 2.115e-3$	$0.44 \pm 6.682e-3$	$0.409 \pm 2.039e-2$	$0.897 \pm 5.624e-3$	$0.49 \pm 1.099e-2$	$0.918 \pm 8.367e-3$
right timemixer	$0.299 \pm 5.797e-3$	$0.405 \pm 5.026e-3$	$0.333 \pm 1.573e-2$	$0.852 \pm 7.719e-3$	$0.426 \pm 1.061e-2$	$0.881 \pm 1.795e-2$

Table 39: In-distribution results - Los. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	mae	mse	mae	mse
Llama-3.1 CoT +TFM lstm	0.343 ± 6.473e − 3	1.017 ± 4.164e − 3	0.347 ± 1.144e − 3	0.988 ± 2.541e − 3
Llama-3.1 CoT +TFM mlp	0.373 ± 3.751e − 3	1.047 ± 9.349e − 3	0.351 ± 2.08e − 3	1.014 ± 1.484e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.516 ± 1.446e − 2	1.113 ± 1.521e − 2	0.489 ± 1.7e − 2	1.091 ± 1.069e − 2
Llama-3.1 CoT +TFM timemixer	0.425 ± 6.04e − 3	1.043 ± 3.039e − 3	0.392 ± 2.021e − 2	1.015 ± 1.283e − 2
Llama-3.1 CoT lstm	0.795 ± 3.91e − 2	1.544 ± 5.866e − 2	0.558 ± 2.002e − 2	1.267 ± 5.593e − 2
Llama-3.1 CoT mlp	0.788 ± 9.331e − 3	1.522 ± 1.23e − 2	0.57 ± 1.623e − 2	1.299 ± 3.296e − 2
Llama-3.1 CoT patchtsmixer	0.736 ± 2.019e − 2	1.365 ± 3.218e − 2	0.598 ± 4.047e − 2	1.255 ± 5.565e − 2
Llama-3.1 CoT timemixer	0.736 ± 2.372e − 2	1.402 ± 3.504e − 2	0.541 ± 1.175e − 2	1.21 ± 1.633e − 2
Llama-3.1 ICD +TFM lstm	0.355 ± 7.647e − 3	1 ± 1.835e − 3	0.345 ± 2.242e − 3	0.972 ± 5.06e − 3
Llama-3.1 ICD +TFM mlp	0.386 ± 2.413e − 3	1.042 ± 1.282e − 2	0.347 ± 5.083e − 3	0.993 ± 4.004e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.516 ± 1.217e − 2	1.11 ± 5.535e − 3	0.535 ± 3.187e − 3	1.154 ± 1.118e − 2
Llama-3.1 ICD +TFM timemixer	0.427 ± 7.748e − 3	1.037 ± 6.871e − 3	0.425 ± 7.496e − 3	1.016 ± 4.586e − 3
Llama-3.1 ICD lstm	0.824 ± 2.439e − 2	1.715 ± 4.504e − 2	0.683 ± 2.758e − 2	1.474 ± 7.487e − 2
Llama-3.1 ICD mlp	0.803 ± 2.404e − 2	1.632 ± 5.696e − 2	0.677 ± 5.667e − 2	1.452 ± 1.538e − 1
Llama-3.1 ICD patchtsmixer	0.807 ± 2.36e − 2	1.583 ± 4.903e − 2	0.645 ± 2.377e − 2	1.29 ± 4.463e − 2
Llama-3.1 ICD timemixer	0.756 ± 1.606e − 2	1.515 ± 2.635e − 2	0.577 ± 7.892e − 3	1.208 ± 1.812e − 2
Llama-3.1 Trend +TFM lstm	0.366 ± 1.907e − 3	1.004 ± 5.751e − 3	0.363 ± 1.113e − 2	0.976 ± 4.142e − 3
Llama-3.1 Trend +TFM mlp	0.393 ± 3.876e − 3	1.037 ± 1.453e − 2	0.378 ± 4.262e − 3	0.995 ± 1.128e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.563 ± 1.006e − 2	1.152 ± 1.473e − 2	0.543 ± 3.772e − 2	1.125 ± 3.512e − 2
Llama-3.1 Trend +TFM timemixer	0.463 ± 1.657e − 2	1.049 ± 1.219e − 2	0.441 ± 6.405e − 3	1.023 ± 6.917e − 3
Llama-3.1 Trend lstm	1.075 ± 1.984e − 2	2.233 ± 5.925e − 2	0.677 ± 7.822e − 3	1.389 ± 1.461e − 2
Llama-3.1 Trend mlp	1.041 ± 2.621e − 2	2.128 ± 5.849e − 2	0.621 ± 1.222e − 2	1.278 ± 3.591e − 2
Llama-3.1 Trend patchtsmixer	0.985 ± 2.71e − 2	1.874 ± 6.315e − 2	0.663 ± 2.518e − 2	1.285 ± 4.798e − 2
Llama-3.1 Trend timemixer	0.971 ± 4.939e − 2	1.889 ± 1.276e − 1	0.62 ± 1.656e − 2	1.228 ± 2.602e − 2
Llama-3.1 zero_shot +TFM lstm	0.347 ± 3.392e − 3	0.991 ± 1.087e − 2	0.355 ± 3.006e − 3	0.976 ± 4.357e − 3
Llama-3.1 zero_shot +TFM mlp	0.373 ± 5.138e − 3	1.018 ± 7.934e − 3	0.349 ± 6.696e − 3	0.992 ± 3.06e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.517 ± 1.5e − 2	1.107 ± 8.694e − 3	0.482 ± 1.316e − 2	1.071 ± 1.003e − 2
Llama-3.1 zero_shot +TFM timemixer	0.418 ± 9.84e − 3	1.025 ± 8.069e − 3	0.429 ± 6.42e − 3	1.013 ± 7.363e − 3
Llama-3.1 zero_shot lstm	0.834 ± 1.674e − 2	1.618 ± 4.316e − 2	0.722 ± 4.193e − 2	1.446 ± 6.508e − 2
Llama-3.1 zero_shot mlp	0.872 ± 2.293e − 2	1.673 ± 3.933e − 2	0.721 ± 5.726e − 2	1.477 ± 1.215e − 1
Llama-3.1 zero_shot patchtsmixer	0.824 ± 9.857e − 3	1.509 ± 2.648e − 2	0.693 ± 3.927e − 2	1.314 ± 6.808e − 2
Llama-3.1 zero_shot timemixer	0.776 ± 2.656e − 2	1.448 ± 5.02e − 2	0.703 ± 4.272e − 2	1.381 ± 8.666e − 2
TFM lstm	0.908 ± 4.184e − 2	1.523 ± 9.286e − 2	1.082 ± 3.76e − 2	2.178 ± 1.497e − 1
TFM mlp	0.895 ± 1.446e − 2	1.427 ± 3.894e − 2	1.082 ± 4.853e − 2	2.196 ± 2.061e − 1
TFM patchtsmixer	0.797 ± 6.325e − 2	1.206 ± 1.132e − 1	0.929 ± 8.13e − 2	1.585 ± 2.215e − 1
TFM timemixer	0.814 ± 1.252e − 1	1.28 ± 2.451e − 1	0.99 ± 7.457e − 2	1.944 ± 2.091e − 1
TSDE lstm	0.551 ± 2.465e − 2	1.205 ± 1.036e − 1	0.766 ± 1.867e − 1	1.144 ± 2.156e − 1
TSDE mlp	1.454 ± 4.611e − 1	3.135 ± 1.268e0	0.85 ± 2.359e − 1	1.323 ± 3.066e − 1
TSDE patchtsmixer	0.729 ± 7.139e − 2	1.147 ± 5.025e − 2	0.758 ± 3.708e − 2	1.163 ± 3.816e − 2
TSDE timemixer	0.888 ± 2.711e − 1	1.446 ± 3.929e − 1	1.42 ± 6.94e − 2	2.369 ± 1.759e − 1
gemini-2.0-flash CoT +TFM lstm	0.357 ± 3.759e − 3	1.016 ± 4.572e − 3	0.333 ± 4.676e − 3	0.986 ± 4.933e − 3
gemini-2.0-flash CoT +TFM mlp	0.383 ± 5.032e − 4	1.07 ± 1.139e − 2	0.348 ± 4.266e − 3	0.998 ± 4.574e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.524 ± 7.045e − 3	1.119 ± 4.735e − 3	0.506 ± 1.762e − 2	1.093 ± 1.268e − 2
gemini-2.0-flash CoT +TFM timemixer	0.413 ± 6.516e − 3	1.034 ± 8.604e − 3	0.418 ± 3.945e − 3	1.023 ± 2.46e − 3
gemini-2.0-flash CoT lstm	1.031 ± 1.665e − 2	2.036 ± 3.139e − 2	0.51 ± 9.334e − 3	1.167 ± 1.83e − 2
gemini-2.0-flash CoT mlp	1.053 ± 3.363e − 2	2.053 ± 7.408e − 2	0.525 ± 2.298e − 2	1.178 ± 2.304e − 2
gemini-2.0-flash CoT patchtsmixer	0.981 ± 4.818e − 2	1.82 ± 1.047e − 1	0.537 ± 2.603e − 2	1.138 ± 4.755e − 2
gemini-2.0-flash CoT timemixer	1.043 ± 3.713e − 2	2.007 ± 7.049e − 2	0.52 ± 2.384e − 2	1.15 ± 3.186e − 2
gemini-2.0-flash ICD +TFM lstm	0.359 ± 2.204e − 3	1.016 ± 1.085e − 2	0.331 ± 2.327e − 3	0.979 ± 4.021e − 3
gemini-2.0-flash ICD +TFM mlp	0.385 ± 5.553e − 3	1.062 ± 2.03e − 2	0.343 ± 4.33e − 3	0.986 ± 4.732e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.521 ± 1.868e − 2	1.132 ± 2.618e − 2	0.491 ± 7.453e − 3	1.071 ± 1.968e − 2
gemini-2.0-flash ICD +TFM timemixer	0.464 ± 1.75e − 2	1.072 ± 1.602e − 2	0.397 ± 2.892e − 2	0.997 ± 6.933e − 3
gemini-2.0-flash ICD lstm	0.67 ± 7.942e − 2	1.369 ± 1.18e − 1	0.507 ± 8.186e − 3	1.15 ± 1.919e − 2
gemini-2.0-flash ICD mlp	0.657 ± 1.122e − 1	1.317 ± 1.588e − 1	0.502 ± 1.481e − 2	1.123 ± 1.053e − 2
gemini-2.0-flash ICD patchtsmixer	0.688 ± 4.246e − 2	1.277 ± 4.596e − 2	0.545 ± 1.764e − 2	1.125 ± 2.547e − 2
gemini-2.0-flash ICD timemixer	0.706 ± 3.749e − 2	1.324 ± 6.35e − 2	0.506 ± 2.247e − 2	1.113 ± 3.125e − 2
gemini-2.0-flash Trend +TFM lstm	0.366 ± 3.201e − 3	1.004 ± 4.439e − 3	0.356 ± 3.593e − 3	0.972 ± 6.498e − 3
gemini-2.0-flash Trend +TFM mlp	0.384 ± 1.284e − 2	1.035 ± 5.205e − 3	0.361 ± 1.506e − 2	0.987 ± 4.302e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.527 ± 2.383e − 2	1.12 ± 9.917e − 3	0.496 ± 1.71e − 2	1.076 ± 1.597e − 2
gemini-2.0-flash Trend +TFM timemixer	0.448 ± 1.2e − 2	1.042 ± 8.75e − 3	0.433 ± 1.054e − 2	1.016 ± 7.269e − 3
gemini-2.0-flash Trend lstm	1.078 ± 2.93e − 2	2.209 ± 7.425e − 2	0.52 ± 3.029e − 2	1.157 ± 4.75e − 2
gemini-2.0-flash Trend mlp	1.115 ± 3.582e − 2	2.265 ± 9.766e − 2	0.545 ± 2.363e − 2	1.187 ± 3.69e − 2
gemini-2.0-flash Trend patchtsmixer	1.059 ± 4.256e − 2	2.041 ± 1.149e − 1	0.576 ± 2.366e − 2	1.165 ± 3.035e − 2
gemini-2.0-flash Trend timemixer	1.083 ± 5.359e − 2	2.152 ± 1.607e − 1	0.539 ± 1.149e − 2	1.141 ± 2.25e − 2
gemini-2.0-flash zero_shot +TFM lstm	0.361 ± 5.824e − 3	1.017 ± 1.433e − 3	0.353 ± 4.76e − 3	0.971 ± 3.582e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.375 ± 2.239e − 3	1.054 ± 5.46e − 3	0.348 ± 4.035e − 3	0.982 ± 1.185e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.544 ± 1.187e − 2	1.144 ± 2.683e − 2	0.511 ± 2.272e − 2	1.092 ± 2.521e − 2
gemini-2.0-flash zero_shot +TFM timemixer	0.429 ± 5.633e − 3	1.045 ± 5.555e − 3	0.405 ± 9.436e − 4	0.995 ± 3.062e − 3
gemini-2.0-flash zero_shot lstm	0.927 ± 3.896e − 2	1.869 ± 8.006e − 2	0.476 ± 1.601e − 2	1.084 ± 2.417e − 2
gemini-2.0-flash zero_shot mlp	0.973 ± 1.181e − 2	1.93 ± 3.46e − 2	0.481 ± 4.471e − 3	1.067 ± 6.673e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.912 ± 1.512e − 2	1.692 ± 3.892e − 2	0.491 ± 1.926e − 2	1.046 ± 1.578e − 2
gemini-2.0-flash zero_shot timemixer	0.888 ± 3.834e − 2	1.699 ± 9.643e − 2	0.484 ± 2.025e − 2	1.063 ± 2.581e − 2
interp lstm	2.313 ± 1.543e − 1	6.444 ± 9.518e − 1	1.01 ± 1.029e − 1	2.33 ± 3.167e − 1
interp mlp	2.398 ± 6.509e − 2	7.078 ± 3.312e − 1	1.465 ± 1.149e − 1	3.839 ± 6.104e − 1
interp patchtsmixer	0.908 ± 8.378e − 2	1.395 ± 1.569e − 1	0.71 ± 5.453e − 2	1.219 ± 5.595e − 2
interp timemixer	0.845 ± 3.561e − 2	1.263 ± 4.676e − 2	0.556 ± 5.435e − 3	1.122 ± 4.275e − 2
mean lstm	3.13 ± 2.202e − 1	10.631 ± 1.473e0	1.293 ± 3.09e − 1	3.012 ± 1.482e0
mean mlp	2.745 ± 6.051e − 2	8.177 ± 3.628e − 1	2.204 ± 1.503e − 1	6.225 ± 6.756e − 1
mean patchtsmixer	1.092 ± 6.323e − 2	1.822 ± 1.841e − 1	0.849 ± 5.7e − 2	1.393 ± 9.83e − 2
mean timemixer	0.982 ± 4.65e − 2	1.503 ± 1.194e − 1	0.61 ± 3.178e − 2	0.978 ± 3.169e − 2
No_sum_CoT CoT +TFM lstm	0.404 ± 9.289e − 3	0.994 ± 3.007e − 3	0.467 ± 1.512e − 2	1.021 ± 1.776e − 2
No_sum_CoT CoT +TFM mlp	0.411 ± 1.161e − 2	1.019 ± 6.541e − 3	0.435 ± 2.834e − 2	1.032 ± 3.57e − 2
No_sum_CoT CoT +TFM patchtsmixer	0.585 ± 2.176e − 2	1.181 ± 4.043e − 2	0.594 ± 6.436e − 2	1.176 ± 1.01e − 1
No_sum_CoT CoT +TFM timemixer	0.527 ± 2.021e − 2	1.099 ± 2.189e − 2	0.581 ± 4.402e − 2	1.194 ± 7.407e − 2
No_sum_CoT lstm	0.844 ± 5.067e − 2	1.724 ± 1.236e − 1	0.967 ± 3.653e − 2	2.018 ± 1.216e − 1
No_sum_CoT mlp	0.931 ± 4.334e − 2	1.89 ± 1.096e − 1	0.915 ± 2.2e − 2	1.851 ± 5.932e − 2
No_sum_CoT patchtsmixer	0.959 ± 4.675e − 2	1.843 ± 1.004e − 1	0.84 ± 4.783e − 2	1.58 ± 1.046e − 1
No_sum_CoT timemixer	0.917 ± 2.341e − 2	1.848 ± 5.82e − 2	0.81 ± 8.9e − 2	1.578 ± 2.312e − 1
No_sum_ICD ICD +TFM lstm	0.367 ± 1.105e − 2	0.989 ± 7.262e − 4	0.387 ± 1.494e − 2	0.962 ± 9.855e − 4
No_sum_ICD ICD +TFM mlp	0.414 ± 2.747e − 2	1.024 ± 2.568e − 2	0.374 ± 6.777e − 3	1.005 ± 2.132e − 2
No_sum_ICD ICD +TFM patchtsmixer	0.605 ± 1.834e − 2	1.208 ± 5.107e − 2	0.565 ± 2.557e − 2	1.13 ± 3.813e − 2
No_sum_ICD ICD +TFM timemixer	0.533 ± 1.226e − 2	1.112 ± 1.019e − 2	0.508 ± 2.526e − 2	1.065 ± 2.561e − 2
No_sum_ICD lstm	0.886 ± 1.93e − 2	1.811 ± 4.878e − 2	0.897 ± 5.587e − 2	1.733 ± 1.361e − 1
No_sum_ICD mlp	0.989 ± 2.553e − 2	2.024 ± 7.196e − 2	0.873 ± 3.112e − 2	1.675 ± 6.143e − 2
No_sum_ICD patchtsmixer	0.926 ± 3.175e − 2	1.804 ± 8.536e − 2	0.887 ± 3.278e − 2	1.639 ± 6.911e − 2
No_sum_ICD timemixer	0.796 ± 8.071e − 2	1.562 ± 1.722e − 1	0.766 ± 5.493e − 2	1.435 ± 1.114e − 1
No_sum_Trend Trend +TFM lstm	0.379 ± 1.126e − 2	0.978 ± 2.533e − 3	0.452 ± 2.965e − 2	1.01 ± 3.527e − 2
No_sum_Trend Trend +TFM mlp	0.403 ± 2.194e − 2	1.012 ± 2.18e − 3	0.423 ± 1.988e − 2	1.013 ± 4.05e − 2
No_sum_Trend Trend +TFM patchtsmixer	0.594 ± 1.385e − 2	1.177 ± 1.88e − 3	0.605 ± 1.428e − 2	1.179 ± 2.521e − 2
No_sum_Trend Trend +TFM timemixer	0.519 ± 4.6e − 3	1.089 ± 1.016e − 2	0.522 ± 2.55e − 2	1.086 ± 3.986e − 2

Table 40: Cross-site transfer results - Los. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	mae	mse	mae	mse
No_sum_Trend lstm	$0.704 \pm 8.339e-2$	$1.389 \pm 1.546e-1$	$0.742 \pm 4.071e-2$	$1.461 \pm 9.049e-2$
No_sum_Trend mlp	$0.754 \pm 3.342e-2$	$1.496 \pm 4.605e-2$	$0.794 \pm 3.188e-2$	$1.568 \pm 6.687e-2$
No_sum_Trend patchtsmixer	$0.853 \pm 3.164e-2$	$1.606 \pm 6.786e-2$	$0.776 \pm 6.463e-2$	$1.468 \pm 1.277e-1$
No_sum_Trend timemixer	$0.802 \pm 5.126e-2$	$1.562 \pm 1.056e-1$	$0.753 \pm 3.791e-2$	$1.451 \pm 8.137e-2$
medgemma CoT +TFM lstm	$0.369 \pm 5.866e-3$	$1.007 \pm 5.832e-3$	$0.352 \pm 6.795e-3$	$0.963 \pm 3.323e-3$
medgemma CoT +TFM mlp	$0.382 \pm 3.831e-3$	$1.042 \pm 1.398e-2$	$0.343 \pm 6.876e-3$	$0.988 \pm 1.181e-2$
medgemma CoT +TFM patchtsmixer	$0.528 \pm 1.409e-3$	$1.126 \pm 3.894e-3$	$0.48 \pm 6.525e-3$	$1.07 \pm 2.149e-2$
medgemma CoT +TFM timemixer	$0.428 \pm 5.554e-3$	$1.025 \pm 9.87e-3$	$0.416 \pm 1.092e-2$	$1 \pm 6.901e-3$
medgemma CoT lstm	$1.116 \pm 2.962e-2$	$2.313 \pm 8.62e-2$	$0.535 \pm 9.297e-3$	$1.181 \pm 3.171e-2$
medgemma CoT mlp	$1.063 \pm 2.839e-2$	$2.121 \pm 8.412e-2$	$0.539 \pm 2.816e-2$	$1.153 \pm 3.959e-2$
medgemma CoT patchtsmixer	$1 \pm 5.994e-2$	$1.878 \pm 1.566e-1$	$0.554 \pm 2.224e-2$	$1.118 \pm 2.79e-2$
medgemma CoT timemixer	$1.018 \pm 8.012e-2$	$1.97 \pm 2.173e-1$	$0.556 \pm 1.472e-2$	$1.165 \pm 1.833e-2$
medgemma ICD +TFM lstm	$0.367 \pm 2.197e-3$	$1.021 \pm 1.355e-2$	$0.334 \pm 8.282e-3$	$0.972 \pm 2.582e-3$
medgemma ICD +TFM mlp	$0.375 \pm 5.707e-3$	$1.042 \pm 6.535e-3$	$0.345 \pm 7.011e-3$	$0.987 \pm 3.264e-3$
medgemma ICD +TFM patchtsmixer	$0.547 \pm 1.076e-2$	$1.16 \pm 2.469e-2$	$0.507 \pm 6.239e-3$	$1.093 \pm 2.527e-3$
medgemma ICD +TFM timemixer	$0.424 \pm 4.215e-3$	$1.065 \pm 7.981e-3$	$0.4 \pm 1.227e-3$	$1 \pm 6.85e-3$
medgemma ICD lstm	$0.969 \pm 3.308e-2$	$1.936 \pm 8.443e-2$	$0.607 \pm 2.19e-2$	$1.315 \pm 3.69e-2$
medgemma ICD mlp	$0.9 \pm 2.985e-2$	$1.773 \pm 6.322e-2$	$0.59 \pm 1.878e-2$	$1.247 \pm 3.427e-2$
medgemma ICD patchtsmixer	$0.904 \pm 3.217e-2$	$1.681 \pm 5.134e-2$	$0.608 \pm 3.347e-2$	$1.213 \pm 3.086e-2$
medgemma ICD timemixer	$0.892 \pm 2.006e-2$	$1.71 \pm 4.965e-2$	$0.585 \pm 3.202e-2$	$1.21 \pm 4.16e-2$
medgemma Trend +TFM lstm	$0.364 \pm 5.091e-3$	$0.993 \pm 8.769e-3$	$0.37 \pm 2.093e-2$	$0.98 \pm 1.068e-2$
medgemma Trend +TFM mlp	$0.385 \pm 1.412e-3$	$1.03 \pm 8.041e-3$	$0.349 \pm 2.607e-3$	$0.99 \pm 7.04e-3$
medgemma Trend +TFM patchtsmixer	$0.565 \pm 2.129e-2$	$1.149 \pm 1.421e-2$	$0.502 \pm 1.755e-3$	$1.082 \pm 3.905e-3$
medgemma Trend +TFM timemixer	$0.453 \pm 6.806e-3$	$1.033 \pm 6.152e-3$	$0.424 \pm 1.422e-2$	$1.014 \pm 9.086e-3$
medgemma Trend lstm	$0.979 \pm 3.445e-2$	$1.975 \pm 1.007e-1$	$0.628 \pm 1.822e-2$	$1.334 \pm 2.749e-2$
medgemma Trend mlp	$1.054 \pm 2.43e-2$	$2.158 \pm 8.046e-2$	$0.646 \pm 2.782e-2$	$1.358 \pm 6.475e-2$
medgemma Trend patchtsmixer	$0.998 \pm 3.545e-2$	$1.899 \pm 7.864e-2$	$0.641 \pm 1.909e-2$	$1.267 \pm 4.206e-2$
medgemma Trend timemixer	$0.974 \pm 2.663e-2$	$1.885 \pm 5.929e-2$	$0.627 \pm 2.08e-2$	$1.271 \pm 1.604e-2$
medgemma zero_shot +TFM lstm	$0.351 \pm 6.479e-3$	$1.023 \pm 6.784e-3$	$0.342 \pm 4.256e-3$	$0.968 \pm 4.758e-3$
medgemma zero_shot +TFM mlp	$0.385 \pm 8.864e-3$	$1.077 \pm 1.426e-2$	$0.343 \pm 6.275e-4$	$0.979 \pm 1.358e-2$
medgemma zero_shot +TFM patchtsmixer	$0.533 \pm 8.593e-3$	$1.131 \pm 1.83e-2$	$0.501 \pm 2.225e-2$	$1.083 \pm 2.194e-2$
medgemma zero_shot +TFM timemixer	$0.413 \pm 6.214e-3$	$1.045 \pm 2.5e-3$	$0.394 \pm 4.21e-3$	$0.987 \pm 4.679e-3$
medgemma zero_shot lstm	$1.094 \pm 4.891e-2$	$2.266 \pm 1.103e-1$	$0.51 \pm 8.577e-3$	$1.137 \pm 1.391e-2$
medgemma zero_shot mlp	$1.075 \pm 1.523e-2$	$2.132 \pm 6.644e-2$	$0.518 \pm 2.548e-2$	$1.125 \pm 4.125e-2$
medgemma zero_shot patchtsmixer	$1.003 \pm 7.105e-2$	$1.86 \pm 1.89e-1$	$0.52 \pm 1.337e-2$	$1.087 \pm 1.572e-2$
medgemma zero_shot timemixer	$0.956 \pm 7.827e-2$	$1.802 \pm 1.906e-1$	$0.492 \pm 6.012e-3$	$1.088 \pm 7.534e-3$
No_sum_zero_shot zero_shot +TFM lstm	$0.408 \pm 1.453e-2$	$0.996 \pm 1.043e-2$	$0.485 \pm 3.59e-2$	$1.06 \pm 4.806e-2$
No_sum_zero_shot zero_shot +TFM mlp	$0.409 \pm 9.729e-3$	$1.025 \pm 1.152e-2$	$0.456 \pm 3.514e-2$	$1.054 \pm 3.852e-2$
No_sum_zero_shot zero_shot +TFM patchtsmixer	$0.573 \pm 1.885e-2$	$1.159 \pm 2.325e-2$	$0.576 \pm 3.064e-2$	$1.144 \pm 4.679e-2$
No_sum_zero_shot zero_shot +TFM timemixer	$0.533 \pm 1.621e-2$	$1.118 \pm 1.32e-2$	$0.56 \pm 3.409e-2$	$1.15 \pm 5.566e-2$
No_sum_zero_shot lstm	$0.907 \pm 2.519e-2$	$1.867 \pm 5.855e-2$	$0.973 \pm 3.373e-2$	$2.067 \pm 9.612e-2$
No_sum_zero_shot mlp	$0.978 \pm 2.412e-2$	$2.032 \pm 7.958e-2$	$0.979 \pm 2.353e-2$	$2.071 \pm 7.347e-2$
No_sum_zero_shot patchtsmixer	$0.983 \pm 5.443e-2$	$1.898 \pm 1.248e-1$	$0.931 \pm 4.189e-2$	$1.836 \pm 9.39e-2$
No_sum_zero_shot timemixer	$0.824 \pm 5.054e-2$	$1.623 \pm 1.238e-1$	$0.873 \pm 3.693e-2$	$1.753 \pm 9.175e-2$
right lstm	$2.116 \pm 4.78e-2$	$6.35 \pm 2.965e-1$	$0.692 \pm 6.11e-2$	$1.542 \pm 2.375e-1$
right mlp	$2.011 \pm 1.23e-1$	$5.654 \pm 7.344e-1$	$1.09 \pm 8.253e-2$	$2.449 \pm 3.128e-1$
right patchtsmixer	$1.03 \pm 6.371e-2$	$1.643 \pm 1.166e-1$	$0.759 \pm 7.229e-2$	$1.268 \pm 5.882e-2$
right timemixer	$0.963 \pm 6.711e-2$	$1.487 \pm 9.134e-2$	$0.599 \pm 3.729e-2$	$1.03 \pm 2.153e-2$

Table 41: Cross-site transfer results - Los. (part 2/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	auprc	auroc	auprc	auroc	auprc	auroc
Llama-3.1 CoT +TFM lstm	0.434 ± 4.047e − 2	0.888 ± 1.253e − 2	0.539 ± 1.934e − 2	0.851 ± 8.073e − 3	0.573 ± 2.554e − 2	0.796 ± 1.254e − 2
Llama-3.1 CoT +TFM mlp	0.443 ± 2.504e − 2	0.89 ± 1.262e − 2	0.514 ± 2.709e − 2	0.847 ± 1.626e − 2	0.546 ± 4.932e − 2	0.789 ± 3.076e − 2
Llama-3.1 CoT +TFM patchtsmixer	0.351 ± 4.992e − 3	0.847 ± 2.807e − 3	0.412 ± 4.796e − 3	0.776 ± 4.092e − 3	0.247 ± 1.763e − 3	0.506 ± 1.556e − 3
Llama-3.1 CoT +TFM timemixer	0.343 ± 1.168e − 2	0.847 ± 4.997e − 3	0.488 ± 4.005e − 3	0.834 ± 2.875e − 3	0.537 ± 3.13e − 2	0.788 ± 1.751e − 2
Llama-3.1 CoT lstm	0.341 ± 9.298e − 3	0.843 ± 5.597e − 3	0.451 ± 5.876e − 3	0.815 ± 3.777e − 3	0.457 ± 2.569e − 2	0.731 ± 1.79e − 2
Llama-3.1 CoT mlp	0.346 ± 1.218e − 2	0.844 ± 7.496e − 3	0.441 ± 7.156e − 3	0.81 ± 4.386e − 3	0.481 ± 1.72e − 2	0.751 ± 1.171e − 2
Llama-3.1 CoT patchtsmixer	0.271 ± 4.8e − 2	0.779 ± 4.136e − 2	0.27 ± 8.385e − 2	0.671 ± 6.974e − 2	0.254 ± 1.391e − 2	0.519 ± 2.568e − 2
Llama-3.1 CoT timemixer	0.346 ± 8.836e − 3	0.838 ± 3.985e − 3	0.438 ± 1.006e − 2	0.801 ± 6.004e − 3	0.427 ± 3.391e − 2	0.707 ± 2.318e − 2
Llama-3.1 ICD +TFM lstm	0.413 ± 1.303e − 2	0.882 ± 5.344e − 3	0.499 ± 2.606e − 2	0.829 ± 1.53e − 2	0.589 ± 8.374e − 3	0.793 ± 2.423e − 3
Llama-3.1 ICD +TFM mlp	0.431 ± 1.249e − 2	0.885 ± 5.047e − 3	0.494 ± 2.567e − 2	0.832 ± 1.578e − 2	0.563 ± 2.534e − 2	0.795 ± 1.56e − 2
Llama-3.1 ICD +TFM patchtsmixer	0.367 ± 1.159e − 2	0.85 ± 4.922e − 3	0.414 ± 1.644e − 2	0.764 ± 1.01e − 2	0.252 ± 2.609e − 3	0.515 ± 3.33e − 3
Llama-3.1 ICD +TFM timemixer	0.316 ± 4.6e − 2	0.833 ± 3.057e − 2	0.467 ± 9.826e − 3	0.816 ± 7.494e − 3	0.529 ± 1.666e − 2	0.772 ± 9.703e − 3
Llama-3.1 ICD lstm	0.358 ± 5.344e − 3	0.851 ± 3.623e − 3	0.426 ± 5.02e − 3	0.786 ± 4.125e − 3	0.485 ± 1.091e − 2	0.735 ± 6.719e − 3
Llama-3.1 ICD mlp	0.36 ± 1.804e − 2	0.85 ± 9.91e − 3	0.405 ± 3.751e − 3	0.777 ± 2.018e − 3	0.484 ± 1.949e − 2	0.74 ± 1.339e − 2
Llama-3.1 ICD patchtsmixer	0.184 ± 1.02e − 1	0.668 ± 1.099e − 1	0.265 ± 6.944e − 2	0.672 ± 4.457e − 2	0.276 ± 2.049e − 2	0.56 ± 3.181e − 2
Llama-3.1 ICD timemixer	0.373 ± 5.533e − 3	0.848 ± 2.639e − 3	0.416 ± 8.855e − 3	0.777 ± 5.424e − 3	0.437 ± 1.335e − 2	0.707 ± 7.682e − 3
Llama-3.1 Trend +TFM lstm	0.477 ± 1.975e − 2	0.905 ± 5.623e − 3	0.516 ± 1.085e − 2	0.848 ± 6.583e − 3	0.553 ± 1.918e − 2	0.782 ± 1.068e − 2
Llama-3.1 Trend +TFM mlp	0.488 ± 1.448e − 2	0.907 ± 5.459e − 3	0.5 ± 1.744e − 3	0.845 ± 3.669e − 4	0.55 ± 1.294e − 2	0.786 ± 9.126e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.397 ± 2.262e − 2	0.875 ± 1.143e − 2	0.422 ± 7.062e − 3	0.786 ± 4.757e − 3	0.249 ± 1.022e − 3	0.507 ± 3.733e − 3
Llama-3.1 Trend +TFM timemixer	0.383 ± 3.465e − 2	0.867 ± 1.702e − 2	0.451 ± 6.875e − 3	0.819 ± 6.305e − 3	0.529 ± 1.622e − 2	0.776 ± 8.389e − 3
Llama-3.1 Trend lstm	0.379 ± 9.086e − 3	0.861 ± 5.678e − 3	0.432 ± 9.522e − 3	0.808 ± 8.114e − 3	0.509 ± 1.237e − 2	0.755 ± 9.144e − 3
Llama-3.1 Trend mlp	0.383 ± 1.052e − 2	0.86 ± 5.351e − 3	0.437 ± 5.673e − 3	0.813 ± 4.773e − 3	0.494 ± 1.749e − 2	0.748 ± 1.242e − 2
Llama-3.1 Trend patchtsmixer	0.32 ± 6.834e − 2	0.817 ± 5.205e − 2	0.246 ± 1.042e − 2	0.669 ± 1.561e − 2	0.256 ± 1.661e − 2	0.525 ± 2.969e − 2
Llama-3.1 Trend timemixer	0.414 ± 2.107e − 2	0.866 ± 9.814e − 3	0.429 ± 1.088e − 2	0.8 ± 8e − 3	0.487 ± 8.158e − 3	0.739 ± 5.146e − 3
Llama-3.1 zero_shot +TFM lstm	0.479 ± 1.003e − 2	0.904 ± 2.183e − 3	0.57 ± 1.765e − 2	0.867 ± 6.438e − 3	0.564 ± 1.818e − 2	0.787 ± 7.777e − 3
Llama-3.1 zero_shot +TFM mlp	0.493 ± 7.385e − 3	0.907 ± 1.986e − 3	0.498 ± 1.421e − 2	0.84 ± 8.639e − 3	0.54 ± 3.381e − 2	0.783 ± 1.955e − 2
Llama-3.1 zero_shot +TFM patchtsmixer	0.359 ± 2.186e − 2	0.863 ± 1.464e − 2	0.428 ± 8.876e − 3	0.786 ± 5.617e − 3	0.251 ± 2.901e − 3	0.511 ± 7.165e − 3
Llama-3.1 zero_shot +TFM timemixer	0.37 ± 5.431e − 3	0.867 ± 1.671e − 3	0.479 ± 2.191e − 2	0.833 ± 1.254e − 2	0.48 ± 1.758e − 2	0.749 ± 1.067e − 2
Llama-3.1 zero_shot lstm	0.398 ± 2.089e − 2	0.873 ± 9.164e − 3	0.437 ± 1.317e − 2	0.813 ± 9.215e − 3	0.492 ± 2.239e − 2	0.746 ± 1.226e − 2
Llama-3.1 zero_shot mlp	0.407 ± 3.909e − 3	0.876 ± 1.509e − 3	0.441 ± 6.955e − 3	0.816 ± 4.737e − 3	0.465 ± 2.571e − 2	0.736 ± 1.543e − 2
Llama-3.1 zero_shot patchtsmixer	0.316 ± 8.626e − 2	0.823 ± 6.232e − 2	0.281 ± 8.333e − 2	0.689 ± 6.299e − 2	0.262 ± 1.707e − 2	0.537 ± 3.236e − 2
Llama-3.1 zero_shot timemixer	0.394 ± 8.982e − 3	0.866 ± 1.953e − 3	0.434 ± 1.654e − 2	0.805 ± 8.895e − 3	0.473 ± 1.179e − 2	0.736 ± 6.462e − 3
TFM lstm	0.363 ± 8.238e − 3	0.856 ± 1.688e − 3	0.485 ± 2.578e − 3	0.797 ± 2.742e − 3	0.402 ± 1.312e − 2	0.676 ± 1.54e − 2
TFM mlp	0.361 ± 1.466e − 2	0.851 ± 3.699e − 3	0.474 ± 6.051e − 3	0.792 ± 3.642e − 3	0.412 ± 1.503e − 2	0.689 ± 1.612e − 2
TFM patchtsmixer	0.266 ± 8.085e − 3	0.793 ± 3.754e − 3	0.435 ± 6.064e − 3	0.753 ± 3.762e − 3	0.279 ± 1.976e − 3	0.511 ± 3.903e − 3
TFM timemixer	0.263 ± 7.734e − 3	0.815 ± 3.212e − 3	0.417 ± 4.331e − 2	0.77 ± 1.75e − 2	0.377 ± 1.22e − 3	0.664 ± 2.151e − 3
TSDE lstm	0.539 ± 1.805e − 2	0.92 ± 3.511e − 3	0.708 ± 2.806e − 3	0.913 ± 3.496e − 4	0.745 ± 7.439e − 3	0.885 ± 4.187e − 3
TSDE mlp	0.528 ± 1.921e − 2	0.921 ± 2.489e − 3	0.705 ± 2.821e − 3	0.913 ± 1.258e − 3	0.751 ± 4.275e − 3	0.888 ± 1.822e − 3
TSDE patchtsmixer	0.324 ± 1.226e − 1	0.806 ± 1.361e − 1	0.494 ± 2.136e − 1	0.793 ± 1.772e − 1	0.583 ± 2.046e − 1	0.765 ± 1.73e − 1
TSDE timemixer	0.48 ± 2.2e − 2	0.909 ± 2.952e − 3	0.679 ± 1.438e − 2	0.906 ± 4.163e − 3	0.741 ± 3.528e − 3	0.884 ± 3.234e − 3
gemini-2.0-flash CoT +TFM lstm	0.545 ± 1.623e − 2	0.92 ± 5.141e − 3	0.57 ± 3.002e − 3	0.868 ± 1.086e − 3	0.667 ± 2.369e − 2	0.843 ± 1.084e − 2
gemini-2.0-flash CoT +TFM mlp	0.533 ± 1.684e − 2	0.913 ± 5.039e − 3	0.556 ± 1.724e − 2	0.865 ± 8.301e − 3	0.629 ± 2.831e − 3	0.83 ± 1.045e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.464 ± 2.736e − 2	0.894 ± 6.52e − 3	0.469 ± 2.338e − 2	0.811 ± 1.472e − 2	0.568 ± 2.321e − 2	0.782 ± 9.886e − 3
gemini-2.0-flash CoT +TFM timemixer	0.463 ± 2.377e − 2	0.897 ± 7.053e − 3	0.506 ± 1.608e − 2	0.845 ± 6.596e − 3	0.611 ± 1.684e − 2	0.827 ± 7.697e − 3
gemini-2.0-flash CoT lstm	0.488 ± 1.318e − 2	0.898 ± 5.601e − 3	0.535 ± 1.128e − 2	0.849 ± 6.038e − 3	0.609 ± 2.214e − 2	0.813 ± 1.378e − 2
gemini-2.0-flash CoT mlp	0.495 ± 5.355e − 3	0.9 ± 2.406e − 3	0.543 ± 4.201e − 3	0.855 ± 2.708e − 3	0.599 ± 2.296e − 2	0.811 ± 1.515e − 2
gemini-2.0-flash CoT patchtsmixer	0.46 ± 7.063e − 3	0.886 ± 2.332e − 3	0.352 ± 8.72e − 2	0.737 ± 5.631e − 2	0.358 ± 1.421e − 1	0.627 ± 1.101e − 1
gemini-2.0-flash CoT timemixer	0.515 ± 1.074e − 2	0.897 ± 3.754e − 3	0.527 ± 5.431e − 3	0.84 ± 3.986e − 3	0.583 ± 4.305e − 3	0.795 ± 2.4e − 3
gemini-2.0-flash ICD +TFM lstm	0.542 ± 1.859e − 2	0.92 ± 5.373e − 3	0.575 ± 3.127e − 3	0.869 ± 1.216e − 3	0.617 ± 1.63e − 2	0.819 ± 8.116e − 3
gemini-2.0-flash ICD +TFM mlp	0.539 ± 1.848e − 2	0.917 ± 5.173e − 3	0.558 ± 1.31e − 2	0.865 ± 7.865e − 3	0.656 ± 1.286e − 2	0.849 ± 9.153e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.463 ± 4.079e − 2	0.893 ± 1.05e − 2	0.494 ± 4.833e − 3	0.819 ± 5.417e − 3	0.446 ± 1.682e − 1	0.683 ± 1.509e − 1
gemini-2.0-flash ICD +TFM timemixer	0.414 ± 2.549e − 2	0.884 ± 1.156e − 2	0.521 ± 1.206e − 2	0.849 ± 8.343e − 3	0.592 ± 2.55e − 2	0.817 ± 1.571e − 2
gemini-2.0-flash ICD lstm	0.491 ± 9.256e − 3	0.903 ± 3.847e − 3	0.519 ± 2.924e − 3	0.844 ± 1.835e − 3	0.601 ± 1.177e − 2	0.812 ± 9.131e − 3
gemini-2.0-flash ICD mlp	0.488 ± 9.744e − 3	0.902 ± 4.768e − 3	0.522 ± 6.623e − 3	0.847 ± 4.112e − 3	0.582 ± 2.098e − 2	0.804 ± 1.415e − 2
gemini-2.0-flash ICD patchtsmixer	0.372 ± 1.737e − 1	0.798 ± 1.752e − 1	0.384 ± 1.383e − 1	0.749 ± 9.318e − 2	0.265 ± 1.408e − 2	0.542 ± 2.714e − 2
gemini-2.0-flash ICD timemixer	0.507 ± 1.638e − 2	0.9 ± 5.618e − 3	0.518 ± 1.333e − 2	0.839 ± 7.76e − 3	0.552 ± 2.55e − 2	0.781 ± 1.502e − 2
gemini-2.0-flash Trend +TFM lstm	0.506 ± 4.189e − 3	0.911 ± 1.254e − 3	0.547 ± 1.392e − 2	0.86 ± 5.845e − 3	0.659 ± 2.17e − 2	0.839 ± 7.601e − 3
gemini-2.0-flash Trend +TFM mlp	0.526 ± 2.121e − 2	0.917 ± 7.035e − 3	0.541 ± 6.958e − 3	0.863 ± 2.501e − 3	0.611 ± 1.525e − 2	0.827 ± 9.683e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.389 ± 3.568e − 2	0.866 ± 1.646e − 2	0.454 ± 5.615e − 3	0.805 ± 6.018e − 3	0.448 ± 1.73e − 1	0.686 ± 1.545e − 1
gemini-2.0-flash Trend +TFM timemixer	0.428 ± 6.544e − 2	0.883 ± 2.307e − 2	0.504 ± 1.568e − 2	0.848 ± 8.365e − 3	0.569 ± 3.539e − 2	0.809 ± 1.971e − 2
gemini-2.0-flash Trend lstm	0.455 ± 2.345e − 2	0.891 ± 1.043e − 2	0.502 ± 8.363e − 3	0.845 ± 3.571e − 3	0.557 ± 2.188e − 2	0.791 ± 1.465e − 2
gemini-2.0-flash Trend mlp	0.453 ± 3.441e − 2	0.888 ± 1.365e − 2	0.496 ± 2.96e − 3	0.842 ± 1.153e − 3	0.556 ± 2.65e − 2	0.793 ± 1.715e − 2
gemini-2.0-flash Trend patchtsmixer	0.345 ± 1.025e − 1	0.826 ± 7.356e − 2	0.304 ± 1.152e − 1	0.701 ± 8.825e − 2	0.318 ± 1.327e − 1	0.581 ± 1.204e − 1
gemini-2.0-flash Trend timemixer	0.46 ± 2.204e − 2	0.884 ± 8.693e − 3	0.499 ± 9.007e − 3	0.838 ± 4.022e − 3	0.537 ± 7.438e − 3	0.775 ± 4.428e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.559 ± 1.48e − 2	0.92 ± 4.149e − 3	0.59 ± 1.774e − 2	0.877 ± 8.688e − 3	0.671 ± 1.047e − 2	0.848 ± 3.791e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.588 ± 1.514e − 2	0.926 ± 4.635e − 3	0.577 ± 2.385e − 3	0.875 ± 9.789e − 4	0.661 ± 1.407e − 2	0.85 ± 9.313e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.529 ± 3.47e − 2	0.913 ± 7.443e − 3	0.481 ± 3.513e − 2	0.818 ± 1.429e − 2	0.585 ± 3.289e − 2	0.801 ± 1.514e − 2
gemini-2.0-flash zero_shot +TFM timemixer	0.482 ± 4.511e − 2	0.897 ± 1.444e − 2	0.545 ± 1.157e − 2	0.859 ± 1.057e − 2	0.61 ± 1.279e − 2	0.825 ± 6.485e − 3
gemini-2.0-flash zero_shot lstm	0.52 ± 6.046e − 3	0.909 ± 1.438e − 3	0.548 ± 5.271e − 3	0.858 ± 3.121e − 3	0.611 ± 1.238e − 2	0.822 ± 6.855e − 3
gemini-2.0-flash zero_shot mlp	0.514 ± 8.721e − 3	0.905 ± 2.253e − 3	0.55 ± 6.146e − 3	0.862 ± 3.469e − 3	0.59 ± 1.971e − 2	0.811 ± 1.093e − 2
gemini-2.0-flash zero_shot patchtsmixer	0.502 ± 6.089e − 3	0.903 ± 3.096e − 3	0.367 ± 1.597e − 1	0.724 ± 1.452e − 1	0.252 ± 6.236e − 3	0.518 ± 1.4e − 2
gemini-2.0-flash zero_shot timemixer	0.53 ± 7.372e − 3	0.904 ± 2.18e − 3	0.556 ± 8.705e − 3	0.86 ± 5.361e − 3	0.588 ± 3.082e − 3	0.804 ± 1.6e − 3
interp lstm	0.531 ± 8.932e − 2	0.891 ± 4.505e − 2	0.635 ± 9.586e − 3	0.888 ± 5.944e − 3	0.704 ± 4.107e − 2	0.844 ± 4.157e − 2
interp mlp	0.552 ± 9.912e − 3	0.909 ± 1.25e − 2	0.631 ± 5.08e − 3	0.89 ± 2.156e − 3	0.715 ± 7.563e − 3	0.863 ± 4.351e − 3
interp patchtsmixer	0.09 ± 0e0	0.5 ± 0e0	0.161 ± 0e0	0.5 ± 0e0	0.273 ± 0e0	0.5 ± 0e0
interp timemixer	0.44 ± 5.277e − 2	0.873 ± 1.352e − 2	0.536 ± 1.41e − 2	0.842 ± 9.574e − 3	0.614 ± 7.715e − 3	0.804 ± 9.152e − 3
mean lstm	0.522 ± 1.259e − 1	0.898 ± 4.499e − 2	0.59 ± 6.604e − 2	0.862 ± 3.451e − 2	0.678 ± 1.46e − 2	0.843 ± 1.883e − 2
mean mlp	0.549 ± 2.305e − 2	0.912 ± 5.273e − 3	0.583 ± 4.494e − 2	0.863 ± 2.267e − 2	0.683 ± 6.58e − 3	0.852 ± 1.164e − 2
mean patchtsmixer	0.09 ± 7.05e − 4	0.5 ± 3.5e − 4	0.161 ± 0e0	0.5 ± 0e0	0.273 ± 0e0	0.5 ± 0e0
mean timemixer	0.59 ± 8.689e − 3	0.927 ± 2.573e − 3	0.62 ± 1.842e − 2	0.881 ± 9.053e − 3	0.654 ± 1.856e − 2	0.827 ± 1.538e − 2
No_sum_CoT lstm	0.634 ± 8.476e − 3	0.93 ± 2.699e − 3	0.621 ± 3.847e − 3	0.888 ± 1.819e − 3	0.681 ± 5.061e − 3	0.852 ± 3.018e − 3
No_sum_CoT mlp	0.603 ± 1.558e − 2	0.925 ± 3.711e − 3	0.615 ± 6.819e − 3	0.888 ± 2.711e − 3	0.682 ± 8.652e − 3	0.856 ± 5.444e − 3
No_sum_CoT patchtsmixer	0.427 ± 1.858e − 1	0.829 ± 1.538e − 1	0.405 ± 1.571e − 1	0.768 ± 1.107e − 1	0.38 ± 1.891e − 1	0.638 ± 1.454e − 1
No_sum_CoT timemixer	0.595 ± 1.814e − 2	0.92 ± 4.36e − 3	0.593 ± 4.734e − 3	0.876 ± 2.179e − 3	0.653 ± 6.419e − 3	0.833 ± 4.429e − 3
No_sum_ICD lstm	0.601 ± 1.16e − 2	0.924 ± 2.787e − 3	0.611 ± 3.797e − 3	0.885 ± 2.014e − 3	0.664 ± 5.689e − 3	0.843 ± 3.831e − 3
No_sum_ICD mlp	0.587 ± 8.258e − 3	0.922 ± 1.93e − 3	0.592 ± 1.31e − 2	0.88 ± 5.193e − 3	0.665 ± 1.535e − 2	0.847 ± 9.774e − 3
No_sum_ICD patchtsmixer	0.44 ± 1.527e − 1	0.868 ± 7.703e − 2	0.365 ± 1.131e − 1	0.751 ± 6.852e − 2	0.28 ± 2.339e − 2	0.562 ± 2.774e − 2
No_sum_ICD timemixer	0.58 ± 4.962e − 3	0.918 ± 1.255e − 3	0.588 ± 7.78e − 3	0.873 ± 3.758e − 3	0.651 ± 1.578e − 2	0.834 ± 9.552e − 3
No_sum_Trend lstm	0.629 ± 4.602e − 3	0.928 ± 1.935e − 3	0.614 ± 5.217e − 3	0.887 ± 2.218e − 3	0.687 ± 9.539e − 3	0.856 ± 5.409e − 3
No_sum_Trend mlp	0.598 ± 7.807e − 3	0.924 ± 1.88e − 3	0.613 ± 1.319e − 2	0.889 ± 5.065e − 3	0.679 ± 1.49e − 2	0.853 ± 8.805e − 3
No_sum_Trend patchtsmixer	0.467 ± 1.18e − 1	0.883 ± 4.873e − 2	0.359 ± 1.188e − 1	0.746 ± 7.483e − 2	0.258 ± 1.191e − 2	0.532 ± 2.675e − 2
No_sum_Trend timemixer	0.602 ± 8.593e − 3	0.921 ± 2.787e − 3	0.605 ± 1.238e − 2	0.882 ± 5.221e − 3	0.642 ± 2.953e − 3	0.827 ± 1.822e − 3
medgemma CoT +TFM lstm	0.592 ± 1.745e − 2	0.927 ± 4.496e − 3	0.624 ± 1.678e − 2	0.89 ± 6.438e − 3	0.68 ± 2.922e − 2	0.853 ± 1.525e − 2
medgemma CoT +TFM mlp	0.6 ± 3.052e − 2	0.929 ± 9.352e − 3	0.593 ± 1.427e − 3	0.882 ± 1.475e − 3	0.638 ± 1.244e − 2	0.839 ± 7.595e − 3
medgemma CoT +TFM patchtsmixer	0.485 ± 2.384e − 2	0.903 ± 7.847e − 3	0.503 ± 8.196e − 3	0.833 ± 5.26e − 3	0.562 ± 3.098e − 2	0.786 ± 1.93e − 2
medgemma CoT +TFM timemixer	0.523 ± 1.243e − 2	0.91 ± 2.409e − 3	0.568 ± 1.704e − 2	0.874 ± 8.405e − 3	0.615 ± 2.055e − 2	0.832 ± 1.059e − 2
medgemma CoT lstm	0.556 ± 9.648e − 3	0.917 ± 3.49e − 3	0.577 ± 8.543e − 3	0.874 ± 3.529e − 3	0.604 ± 1.558e − 2	0.818 ± 1.096e − 2
medgemma CoT mlp	0.542 ± 2.072e − 2	0.912 ± 6.416e − 3	0.583 ± 1.125e − 2	0.879 ± 5.863e − 3	0.589 ± 6.352e − 3	0.81 ± 4.491e − 3
medgemma CoT patchtsmixer	0.407 ± 1.298e − 1	0.854 ± 7.869e − 2	0.357 ± 9.654e − 2	0.742 ± 6.527e − 2	0.304 ± 1.14e − 2	0.591 ± 1.576e − 2
medgemma CoT timemixer	0.543 ± 1.328e − 2	0.906 ± 3.944e − 3	0.57 ± 1.39e − 2	0.866 ± 6.891e − 3	0.568 ± 1.264e − 2	0.789 ± 7.697e − 3

Table 42: In-distribution results - Mort. (part 1/2)

Method	hirid → hirid		mimic → mimic		ppicu → ppicu	
	auprc	auroc	auprc	auroc	auprc	auroc
medgemma ICD +TFM lstm	$0.582 \pm 6.357e-3$	$0.927 \pm 2.878e-3$	$0.572 \pm 1.236e-2$	$0.87 \pm 5.52e-3$	$0.66 \pm 2.533e-2$	$0.844 \pm 1.297e-2$
medgemma ICD +TFM mlp	$0.582 \pm 2.505e-2$	$0.929 \pm 8.287e-3$	$0.571 \pm 4.68e-3$	$0.874 \pm 4.162e-4$	$0.634 \pm 1.234e-2$	$0.837 \pm 7.092e-3$
medgemma ICD +TFM patchtsmixer	$0.467 \pm 5.098e-3$	$0.899 \pm 2.98e-3$	$0.489 \pm 2.237e-2$	$0.821 \pm 1.109e-2$	$0.582 \pm 1.334e-2$	$0.786 \pm 8.672e-3$
medgemma ICD +TFM timemixer	$0.453 \pm 6.038e-2$	$0.894 \pm 1.792e-2$	$0.526 \pm 1.474e-2$	$0.853 \pm 7.344e-3$	$0.609 \pm 1.786e-2$	$0.827 \pm 1.059e-2$
medgemma ICD lstm	$0.506 \pm 2.012e-2$	$0.906 \pm 7.309e-3$	$0.547 \pm 1.449e-2$	$0.858 \pm 7.275e-3$	$0.594 \pm 8.807e-3$	$0.81 \pm 5.606e-3$
medgemma ICD mlp	$0.492 \pm 1.938e-2$	$0.901 \pm 7.272e-3$	$0.538 \pm 5.229e-3$	$0.858 \pm 2.935e-3$	$0.586 \pm 1.356e-2$	$0.807 \pm 9.684e-3$
medgemma ICD patchtsmixer	$0.381 \pm 1.088e-1$	$0.85 \pm 6.58e-2$	$0.385 \pm 1.089e-1$	$0.764 \pm 6.721e-2$	$0.28 \pm 1.13e-2$	$0.564 \pm 1.614e-2$
medgemma ICD timemixer	$0.504 \pm 1.269e-2$	$0.899 \pm 4.643e-3$	$0.53 \pm 1.274e-2$	$0.847 \pm 6.164e-3$	$0.574 \pm 5.149e-3$	$0.794 \pm 2.91e-3$
medgemma Trend +TFM lstm	$0.524 \pm 3.32e-2$	$0.911 \pm 1.015e-2$	$0.549 \pm 1.372e-2$	$0.864 \pm 5.829e-3$	$0.643 \pm 3.179e-2$	$0.827 \pm 1.416e-2$
medgemma Trend +TFM mlp	$0.573 \pm 1.708e-2$	$0.928 \pm 6.287e-3$	$0.532 \pm 1.378e-2$	$0.86 \pm 7.304e-3$	$0.595 \pm 1.673e-2$	$0.813 \pm 9.188e-3$
medgemma Trend +TFM patchtsmixer	$0.371 \pm 5.494e-3$	$0.863 \pm 3.222e-3$	$0.456 \pm 2.658e-2$	$0.818 \pm 1.328e-2$	$0.25 \pm 5.324e-3$	$0.509 \pm 4.151e-3$
medgemma Trend +TFM timemixer	$0.396 \pm 3.892e-2$	$0.87 \pm 1.642e-2$	$0.499 \pm 1.572e-2$	$0.85 \pm 7.941e-3$	$0.56 \pm 2.041e-2$	$0.796 \pm 1.352e-2$
medgemma Trend lstm	$0.44 \pm 2.584e-2$	$0.884 \pm 8.082e-3$	$0.501 \pm 6.582e-3$	$0.846 \pm 3.181e-3$	$0.565 \pm 1.96e-2$	$0.788 \pm 1.193e-2$
medgemma Trend mlp	$0.454 \pm 2.115e-2$	$0.886 \pm 6.986e-3$	$0.492 \pm 1.114e-2$	$0.843 \pm 5.195e-3$	$0.522 \pm 2.362e-2$	$0.767 \pm 1.326e-2$
medgemma Trend patchtsmixer	$0.365 \pm 4.961e-3$	$0.858 \pm 3.198e-3$	$0.369 \pm 8.7e-2$	$0.769 \pm 6.259e-2$	$0.334 \pm 1.194e-1$	$0.608 \pm 1.03e-1$
medgemma Trend timemixer	$0.441 \pm 1.48e-2$	$0.875 \pm 4.638e-3$	$0.495 \pm 3.069e-3$	$0.843 \pm 2.361e-3$	$0.522 \pm 8.734e-3$	$0.762 \pm 6.247e-3$
medgemma zero_shot +TFM lstm	$0.579 \pm 2.183e-2$	$0.924 \pm 5.858e-3$	$0.602 \pm 1.126e-2$	$0.884 \pm 5.661e-3$	$0.644 \pm 1.1e-2$	$0.836 \pm 4.788e-3$
medgemma zero_shot +TFM mlp	$0.57 \pm 1.762e-2$	$0.921 \pm 5.704e-3$	$0.579 \pm 2.067e-2$	$0.876 \pm 1.088e-2$	$0.65 \pm 2.438e-2$	$0.846 \pm 1.564e-2$
medgemma zero_shot +TFM patchtsmixer	$0.418 \pm 6.342e-2$	$0.881 \pm 2.308e-2$	$0.478 \pm 2.752e-2$	$0.818 \pm 1.679e-2$	$0.572 \pm 1.182e-2$	$0.79 \pm 7.234e-3$
medgemma zero_shot +TFM timemixer	$0.459 \pm 4.196e-2$	$0.892 \pm 1.27e-2$	$0.553 \pm 1.822e-2$	$0.868 \pm 1.061e-2$	$0.614 \pm 1.171e-2$	$0.829 \pm 4.433e-3$
medgemma zero_shot lstm	$0.546 \pm 6.857e-3$	$0.913 \pm 2.359e-3$	$0.551 \pm 9.915e-3$	$0.863 \pm 4.472e-3$	$0.606 \pm 1.032e-2$	$0.818 \pm 6.156e-3$
medgemma zero_shot mlp	$0.545 \pm 8.764e-3$	$0.913 \pm 3.428e-3$	$0.559 \pm 6.325e-3$	$0.868 \pm 2.58e-3$	$0.598 \pm 1.61e-2$	$0.815 \pm 1.05e-2$
medgemma zero_shot patchtsmixer	$0.462 \pm 7.816e-3$	$0.887 \pm 4.008e-3$	$0.463 \pm 8.225e-2$	$0.814 \pm 5.34e-2$	$0.3 \pm 1.767e-2$	$0.592 \pm 1.908e-2$
medgemma zero_shot timemixer	$0.543 \pm 6.09e-3$	$0.908 \pm 2.577e-3$	$0.562 \pm 1.155e-2$	$0.862 \pm 6.347e-3$	$0.585 \pm 5.476e-3$	$0.801 \pm 3.374e-3$
No_sum_zero_shot lstm	$0.615 \pm 2.522e-2$	$0.925 \pm 6.661e-3$	$0.606 \pm 6.168e-3$	$0.883 \pm 2.376e-3$	$0.686 \pm 9.426e-3$	$0.855 \pm 5.876e-3$
No_sum_zero_shot mlp	$0.603 \pm 9.87e-3$	$0.924 \pm 2.557e-3$	$0.616 \pm 1.695e-2$	$0.89 \pm 6.911e-3$	$0.669 \pm 2.192e-2$	$0.847 \pm 1.442e-2$
No_sum_zero_shot patchtsmixer	$0.468 \pm 1.222e-1$	$0.882 \pm 5.065e-2$	$0.479 \pm 1.207e-1$	$0.823 \pm 7.118e-2$	$0.267 \pm 3.135e-2$	$0.54 \pm 4.644e-2$
No_sum_zero_shot timemixer	$0.587 \pm 1.445e-2$	$0.917 \pm 4.193e-3$	$0.579 \pm 2.141e-2$	$0.871 \pm 9.394e-3$	$0.647 \pm 1.746e-2$	$0.83 \pm 1.127e-2$
right lstm	$0.572 \pm 1.18e-1$	$0.898 \pm 6.297e-2$	$0.675 \pm 1.337e-2$	$0.901 \pm 4.097e-3$	$0.726 \pm 2.893e-2$	$0.863 \pm 2.959e-2$
right mlp	$0.625 \pm 2.227e-2$	$0.918 \pm 6.474e-3$	$0.665 \pm 9.48e-3$	$0.896 \pm 4.352e-3$	$0.742 \pm 7.615e-3$	$0.874 \pm 7.008e-3$
right patchtsmixer	$0.09 \pm 0e0$	$0.5 \pm 1.15e-4$	$0.161 \pm 0e0$	$0.5 \pm 0e0$	$0.273 \pm 0e0$	$0.5 \pm 0e0$
right timemixer	$0.54 \pm 2.507e-2$	$0.907 \pm 1.645e-2$	$0.589 \pm 1.16e-2$	$0.867 \pm 3.511e-3$	$0.648 \pm 5.734e-3$	$0.821 \pm 2.964e-3$

Table 43: In-distribution results - Mort. (part 2/2)

Method	hirid → ppicu		mimic → ppicu	
	auprc	auroc	auprc	auroc
Llama-3.1 CoT +TFM lstm	0.365 ± 5.073e − 3	0.656 ± 2.368e − 3	0.392 ± 2.422e − 3	0.681 ± 9.8e − 4
Llama-3.1 CoT +TFM mlp	0.373 ± 8.213e − 4	0.664 ± 1.107e − 3	0.383 ± 7.763e − 3	0.67 ± 6.792e − 3
Llama-3.1 CoT +TFM patchtsmixer	0.342 ± 2.198e − 3	0.625 ± 1.408e − 3	0.361 ± 4.654e − 3	0.633 ± 5.435e − 3
Llama-3.1 CoT +TFM timemixer	0.369 ± 4.226e − 3	0.662 ± 3.855e − 3	0.376 ± 2.861e − 3	0.665 ± 2.893e − 3
Llama-3.1 CoT lstm	0.362 ± 3.588e − 3	0.653 ± 1.988e − 3	0.365 ± 1.458e − 3	0.652 ± 1.682e − 3
Llama-3.1 CoT mlp	0.361 ± 4.063e − 3	0.65 ± 4.195e − 3	0.365 ± 2.772e − 3	0.652 ± 2.356e − 3
Llama-3.1 CoT patchtsmixer	0.32 ± 3.425e − 2	0.602 ± 4.628e − 2	0.314 ± 3.019e − 2	0.6 ± 3.221e − 2
Llama-3.1 CoT timemixer	0.357 ± 1.49e − 3	0.647 ± 1.388e − 3	0.357 ± 1.373e − 3	0.647 ± 1.679e − 3
Llama-3.1 ICD +TFM lstm	0.36 ± 1.051e − 3	0.656 ± 1.813e − 3	0.389 ± 4.147e − 3	0.674 ± 4.169e − 3
Llama-3.1 ICD +TFM mlp	0.367 ± 3.983e − 3	0.661 ± 3.915e − 3	0.393 ± 3.627e − 3	0.676 ± 3.508e − 3
Llama-3.1 ICD +TFM patchtsmixer	0.347 ± 9.868e − 4	0.632 ± 2.735e − 3	0.361 ± 2.17e − 3	0.637 ± 2.782e − 3
Llama-3.1 ICD +TFM timemixer	0.368 ± 2.004e − 3	0.664 ± 6.352e − 3	0.388 ± 1.479e − 3	0.675 ± 1.013e − 3
Llama-3.1 ICD lstm	0.359 ± 1.49e − 3	0.657 ± 9.47e − 4	0.385 ± 4.528e − 3	0.674 ± 2.962e − 3
Llama-3.1 ICD mlp	0.361 ± 2.11e − 3	0.657 ± 7.309e − 4	0.386 ± 1.053e − 3	0.674 ± 4.826e − 4
Llama-3.1 ICD patchtsmixer	0.291 ± 4.389e − 2	0.578 ± 5.288e − 2	0.31 ± 3.568e − 2	0.604 ± 3.397e − 2
Llama-3.1 ICD timemixer	0.362 ± 1.846e − 3	0.658 ± 1.467e − 4	0.388 ± 3.129e − 3	0.672 ± 1.952e − 3
Llama-3.1 Trend +TFM lstm	0.355 ± 3.702e − 3	0.655 ± 3.368e − 3	0.388 ± 7.469e − 3	0.677 ± 5.3e − 3
Llama-3.1 Trend +TFM mlp	0.36 ± 4.772e − 3	0.66 ± 3.541e − 3	0.391 ± 9.703e − 3	0.677 ± 4.798e − 3
Llama-3.1 Trend +TFM patchtsmixer	0.341 ± 5.777e − 3	0.624 ± 4.799e − 3	0.375 ± 2.775e − 3	0.652 ± 3.665e − 3
Llama-3.1 Trend +TFM timemixer	0.364 ± 4.571e − 3	0.663 ± 1.511e − 3	0.385 ± 3.674e − 3	0.672 ± 4.213e − 3
Llama-3.1 Trend lstm	0.366 ± 2.342e − 3	0.663 ± 4.979e − 4	0.385 ± 7.211e − 3	0.669 ± 3.003e − 3
Llama-3.1 Trend mlp	0.369 ± 1.656e − 3	0.667 ± 2.262e − 4	0.384 ± 2.798e − 3	0.669 ± 1.983e − 3
Llama-3.1 Trend patchtsmixer	0.344 ± 3.142e − 2	0.642 ± 4.051e − 2	0.282 ± 7.166e − 3	0.565 ± 1.115e − 2
Llama-3.1 Trend timemixer	0.364 ± 6.515e − 4	0.662 ± 9.032e − 4	0.385 ± 3.325e − 3	0.665 ± 1.772e − 3
Llama-3.1 zero_shot +TFM lstm	0.346 ± 2.309e − 3	0.65 ± 8.195e − 4	0.398 ± 3.998e − 3	0.674 ± 1.957e − 3
Llama-3.1 zero_shot +TFM mlp	0.348 ± 4.011e − 3	0.653 ± 8.969e − 4	0.403 ± 2.312e − 3	0.688 ± 2.442e − 3
Llama-3.1 zero_shot +TFM patchtsmixer	0.333 ± 1.884e − 3	0.612 ± 5.934e − 3	0.381 ± 2.406e − 3	0.652 ± 2.807e − 3
Llama-3.1 zero_shot +TFM timemixer	0.361 ± 1.536e − 3	0.665 ± 1.248e − 3	0.397 ± 1.288e − 3	0.683 ± 5.83e − 3
Llama-3.1 zero_shot lstm	0.362 ± 2.002e − 3	0.654 ± 5.816e − 4	0.39 ± 9.319e − 4	0.671 ± 1.87e − 3
Llama-3.1 zero_shot mlp	0.362 ± 1.834e − 3	0.654 ± 1.988e − 3	0.39 ± 2.984e − 3	0.672 ± 1.651e − 3
Llama-3.1 zero_shot patchtsmixer	0.336 ± 3.817e − 2	0.632 ± 4.132e − 2	0.321 ± 4.617e − 2	0.614 ± 4.224e − 2
Llama-3.1 zero_shot timemixer	0.361 ± 2.158e − 3	0.657 ± 1.538e − 3	0.389 ± 2.366e − 3	0.673 ± 2.809e − 3
TFM lstm	0.408 ± 1.68e − 3	0.686 ± 1.911e − 3	0.393 ± 3.643e − 3	0.683 ± 1.977e − 3
TFM mlp	0.409 ± 4.157e − 3	0.678 ± 3.606e − 3	0.409 ± 4.364e − 3	0.694 ± 2.94e − 3
TFM patchtsmixer	0.375 ± 1.74e − 3	0.637 ± 8.51e − 3	0.415 ± 9.513e − 3	0.688 ± 2.668e − 3
TFM timemixer	0.39 ± 1.764e − 3	0.662 ± 1.212e − 3	0.397 ± 1.115e − 2	0.679 ± 1.147e − 2
TSDE lstm	0.311 ± 9.39e − 2	0.521 ± 1.192e − 1	0.231 ± 2.691e − 2	0.399 ± 6.764e − 2
TSDE mlp	0.412 ± 1.143e − 1	0.596 ± 1.218e − 1	0.273 ± 1.937e − 2	0.466 ± 2.423e − 2
TSDE patchtsmixer	0.265 ± 6.186e − 3	0.49 ± 8.424e − 3	0.274 ± 2.561e − 3	0.503 ± 2.375e − 3
TSDE timemixer	0.255 ± 1.379e − 2	0.466 ± 2.595e − 2	0.276 ± 4.016e − 2	0.5 ± 6.036e − 2
gemini-2.0-flash CoT +TFM lstm	0.44 ± 5.224e − 3	0.711 ± 2.302e − 3	0.481 ± 3.247e − 3	0.743 ± 1.26e − 3
gemini-2.0-flash CoT +TFM mlp	0.443 ± 8.222e − 3	0.714 ± 6.119e − 3	0.478 ± 8.296e − 3	0.742 ± 3.99e − 3
gemini-2.0-flash CoT +TFM patchtsmixer	0.404 ± 2.595e − 3	0.672 ± 3.237e − 3	0.436 ± 2e − 3	0.699 ± 7.007e − 3
gemini-2.0-flash CoT +TFM timemixer	0.439 ± 3.451e − 3	0.717 ± 1.775e − 3	0.466 ± 4.676e − 3	0.736 ± 8.95e − 4
gemini-2.0-flash CoT lstm	0.447 ± 2.148e − 3	0.715 ± 1.988e − 3	0.492 ± 4.484e − 3	0.741 ± 3.722e − 3
gemini-2.0-flash CoT mlp	0.453 ± 1.266e − 3	0.716 ± 1.248e − 3	0.489 ± 3.167e − 3	0.739 ± 1.827e − 3
gemini-2.0-flash CoT patchtsmixer	0.443 ± 1.368e − 3	0.714 ± 1.423e − 3	0.382 ± 5.961e − 2	0.662 ± 4.915e − 2
gemini-2.0-flash CoT timemixer	0.455 ± 2.492e − 3	0.714 ± 5.74e − 4	0.493 ± 2.823e − 3	0.741 ± 1.451e − 3
gemini-2.0-flash ICD +TFM lstm	0.372 ± 1.011e − 2	0.662 ± 8.436e − 3	0.484 ± 3.49e − 3	0.739 ± 1.642e − 3
gemini-2.0-flash ICD +TFM mlp	0.373 ± 5.795e − 3	0.668 ± 2.982e − 3	0.479 ± 4.541e − 3	0.738 ± 4.24e − 3
gemini-2.0-flash ICD +TFM patchtsmixer	0.35 ± 4.79e − 3	0.631 ± 3.774e − 3	0.431 ± 5.018e − 3	0.689 ± 4.109e − 3
gemini-2.0-flash ICD +TFM timemixer	0.393 ± 1.423e − 2	0.675 ± 8.435e − 3	0.47 ± 3.658e − 3	0.73 ± 2.716e − 3
gemini-2.0-flash ICD lstm	0.385 ± 1.991e − 3	0.671 ± 2.031e − 3	0.478 ± 2.194e − 3	0.735 ± 1.021e − 3
gemini-2.0-flash ICD mlp	0.386 ± 4.339e − 3	0.67 ± 3.934e − 3	0.476 ± 2.665e − 3	0.734 ± 1.832e − 3
gemini-2.0-flash ICD patchtsmixer	0.359 ± 7.454e − 2	0.633 ± 8.33e − 2	0.394 ± 7.93e − 2	0.671 ± 6.795e − 2
gemini-2.0-flash ICD timemixer	0.392 ± 1.708e − 3	0.671 ± 1.699e − 3	0.478 ± 2.165e − 3	0.735 ± 1.052e − 3
gemini-2.0-flash Trend +TFM lstm	0.406 ± 1.599e − 3	0.693 ± 8.598e − 4	0.436 ± 7.652e − 3	0.708 ± 5.305e − 3
gemini-2.0-flash Trend +TFM mlp	0.402 ± 2.571e − 3	0.69 ± 1.739e − 3	0.423 ± 4.072e − 3	0.696 ± 2.111e − 3
gemini-2.0-flash Trend +TFM patchtsmixer	0.376 ± 1.085e − 3	0.659 ± 2.212e − 3	0.402 ± 2.707e − 3	0.674 ± 1.87e − 3
gemini-2.0-flash Trend +TFM timemixer	0.403 ± 2.033e − 3	0.692 ± 7.048e − 3	0.422 ± 7.139e − 3	0.698 ± 5.21e − 3
gemini-2.0-flash Trend lstm	0.4 ± 4.035e − 3	0.687 ± 5.04e − 3	0.405 ± 8.401e − 3	0.692 ± 6.798e − 3
gemini-2.0-flash Trend mlp	0.411 ± 1.205e − 3	0.693 ± 4.034e − 3	0.412 ± 4.631e − 3	0.696 ± 3.027e − 3
gemini-2.0-flash Trend patchtsmixer	0.378 ± 7.402e − 2	0.662 ± 7.818e − 2	0.326 ± 6.116e − 2	0.613 ± 5.775e − 2
gemini-2.0-flash Trend timemixer	0.409 ± 2.988e − 3	0.694 ± 2.309e − 3	0.398 ± 7.758e − 3	0.689 ± 5.446e − 3
gemini-2.0-flash zero_shot +TFM lstm	0.46 ± 2.592e − 3	0.724 ± 1.031e − 3	0.486 ± 4.692e − 3	0.75 ± 2.913e − 3
gemini-2.0-flash zero_shot +TFM mlp	0.455 ± 1.591e − 3	0.721 ± 1.511e − 3	0.486 ± 6.096e − 3	0.747 ± 4.234e − 3
gemini-2.0-flash zero_shot +TFM patchtsmixer	0.414 ± 2.556e − 3	0.68 ± 2.316e − 3	0.453 ± 5.93e − 3	0.717 ± 4.403e − 3
gemini-2.0-flash zero_shot +TFM timemixer	0.446 ± 2.195e − 3	0.716 ± 2.61e − 3	0.469 ± 3.844e − 3	0.742 ± 6.946e − 4
gemini-2.0-flash zero_shot lstm	0.462 ± 3.317e − 3	0.725 ± 2.439e − 3	0.492 ± 7.376e − 4	0.75 ± 6.335e − 4
gemini-2.0-flash zero_shot mlp	0.467 ± 1.358e − 3	0.726 ± 1.781e − 3	0.491 ± 3.966e − 3	0.745 ± 2.92e − 3
gemini-2.0-flash zero_shot patchtsmixer	0.454 ± 4.39e − 3	0.723 ± 2.019e − 3	0.39 ± 1.129e − 1	0.66 ± 1.138e − 1
gemini-2.0-flash zero_shot timemixer	0.463 ± 1.015e − 3	0.722 ± 9.899e − 4	0.487 ± 4.665e − 3	0.741 ± 4.708e − 3
interp lstm	0.336 ± 2.386e − 2	0.612 ± 3.01e − 2	0.327 ± 3.616e − 2	0.557 ± 3.184e − 2
interp mlp	0.308 ± 8.673e − 3	0.573 ± 2.125e − 2	0.233 ± 5.897e − 3	0.445 ± 1.573e − 2
interp patchtsmixer	0.273 ± 1.45e − 4	0.5 ± 9.5e − 5	0.273 ± 0e0	0.5 ± 0e0
interp timemixer	0.299 ± 2.289e − 2	0.536 ± 4.692e − 2	0.272 ± 1.66e − 2	0.496 ± 2.654e − 2
mean lstm	0.35 ± 8.786e − 3	0.6 ± 9.442e − 3	0.358 ± 5.789e − 2	0.573 ± 4.881e − 2
mean mlp	0.341 ± 2.137e − 2	0.588 ± 2.935e − 2	0.313 ± 3.221e − 2	0.505 ± 5.15e − 2
mean patchtsmixer	0.273 ± 1.7e − 4	0.5 ± 2.05e − 4	0.273 ± 1.45e − 4	0.5 ± 9.5e − 5
mean timemixer	0.348 ± 1.908e − 2	0.615 ± 1.567e − 2	0.284 ± 1.763e − 2	0.516 ± 2.608e − 2
No_sum_CoT lstm	0.522 ± 7.255e − 3	0.757 ± 3.119e − 3	0.477 ± 3.339e − 3	0.735 ± 3.721e − 3
No_sum_CoT mlp	0.522 ± 3.581e − 4	0.76 ± 7.204e − 4	0.449 ± 7.005e − 3	0.7 ± 6.358e − 3
No_sum_CoT patchtsmixer	0.408 ± 9.767e − 2	0.68 ± 9.853e − 2	0.396 ± 1.008e − 1	0.674 ± 8.421e − 2
No_sum_CoT timemixer	0.502 ± 8.977e − 3	0.75 ± 2.818e − 3	0.446 ± 1.245e − 2	0.702 ± 1.019e − 2
No_sum_ICD lstm	0.509 ± 4.305e − 3	0.753 ± 7.202e − 4	0.492 ± 6.297e − 3	0.744 ± 1.324e − 3
No_sum_ICD mlp	0.504 ± 7.52e − 3	0.755 ± 2.455e − 3	0.512 ± 1.189e − 2	0.751 ± 6.433e − 3
No_sum_ICD patchtsmixer	0.407 ± 7.366e − 2	0.7 ± 7.045e − 2	0.359 ± 8.137e − 2	0.655 ± 7.15e − 2
No_sum_ICD timemixer	0.482 ± 5.592e − 3	0.747 ± 1.144e − 3	0.483 ± 1.094e − 2	0.739 ± 5.264e − 3
No_sum_Trend lstm	0.549 ± 2.184e − 3	0.768 ± 1.26e − 3	0.489 ± 8.991e − 3	0.752 ± 4.14e − 3
No_sum_Trend mlp	0.551 ± 1.711e − 3	0.772 ± 3.869e − 4	0.489 ± 1.771e − 2	0.739 ± 1.1e − 2
No_sum_Trend patchtsmixer	0.461 ± 7.405e − 2	0.723 ± 4.69e − 2	0.379 ± 8.544e − 2	0.672 ± 6.206e − 2
No_sum_Trend timemixer	0.525 ± 6.96e − 3	0.759 ± 3.27e − 3	0.453 ± 1.791e − 2	0.723 ± 1.081e − 2
medgemma CoT +TFM lstm	0.461 ± 2.095e − 3	0.721 ± 6.483e − 4	0.495 ± 7.993e − 3	0.737 ± 4.285e − 3
medgemma CoT +TFM mlp	0.46 ± 2.703e − 3	0.722 ± 5.972e − 4	0.491 ± 3.25e − 3	0.727 ± 3.272e − 3
medgemma CoT +TFM patchtsmixer	0.419 ± 3.118e − 3	0.677 ± 6.207e − 4	0.453 ± 1.679e − 3	0.698 ± 7.701e − 4
medgemma CoT +TFM timemixer	0.451 ± 2.571e − 3	0.721 ± 4.86e − 4	0.481 ± 2.674e − 3	0.722 ± 3.019e − 3
medgemma CoT lstm	0.482 ± 1.635e − 3	0.731 ± 1.264e − 3	0.501 ± 6.132e − 3	0.737 ± 4.484e − 3
medgemma CoT mlp	0.477 ± 2.855e − 3	0.73 ± 1.213e − 3	0.497 ± 6.07e − 3	0.733 ± 3.784e − 3
medgemma CoT patchtsmixer	0.416 ± 8.737e − 2	0.683 ± 7.765e − 2	0.391 ± 6.036e − 2	0.658 ± 5.204e − 2
medgemma CoT timemixer	0.473 ± 2.42e − 3	0.726 ± 1.035e − 3	0.486 ± 4.948e − 3	0.728 ± 2.753e − 3

Table 44: Cross-site transfer results - Mort. (part 1/2)

Method	hirid → ppicu		mimic → ppicu	
	auprc	auroc	auprc	auroc
medgemma ICD +TFM lstm	$0.459 \pm 2.197e-3$	$0.715 \pm 1.475e-3$	$0.489 \pm 4.21e-3$	$0.741 \pm 2.302e-3$
medgemma ICD +TFM mlp	$0.46 \pm 7.566e-3$	$0.717 \pm 3.012e-3$	$0.484 \pm 9.579e-3$	$0.736 \pm 6.33e-3$
medgemma ICD +TFM patchtsmixer	$0.408 \pm 2.124e-3$	$0.675 \pm 2.087e-3$	$0.449 \pm 2.893e-3$	$0.7 \pm 5.408e-4$
medgemma ICD +TFM timemixer	$0.438 \pm 1.577e-2$	$0.712 \pm 4.393e-3$	$0.482 \pm 3.808e-3$	$0.735 \pm 6.012e-4$
medgemma ICD lstm	$0.455 \pm 7.277e-3$	$0.715 \pm 1.228e-3$	$0.49 \pm 4.828e-3$	$0.741 \pm 2.204e-3$
medgemma ICD mlp	$0.461 \pm 2.622e-3$	$0.717 \pm 1.017e-3$	$0.487 \pm 2.734e-3$	$0.74 \pm 1.132e-3$
medgemma ICD patchtsmixer	$0.416 \pm 5.818e-2$	$0.692 \pm 4.581e-2$	$0.403 \pm 7.541e-2$	$0.685 \pm 5.847e-2$
medgemma ICD timemixer	$0.453 \pm 2.641e-3$	$0.714 \pm 7.776e-4$	$0.485 \pm 3.375e-3$	$0.737 \pm 1.735e-3$
medgemma Trend +TFM lstm	$0.411 \pm 3.69e-3$	$0.694 \pm 2.313e-3$	$0.422 \pm 7.35e-3$	$0.695 \pm 4.394e-3$
medgemma Trend +TFM mlp	$0.407 \pm 8.281e-3$	$0.691 \pm 7.451e-3$	$0.423 \pm 4.594e-3$	$0.697 \pm 4.121e-3$
medgemma Trend +TFM patchtsmixer	$0.368 \pm 2.699e-3$	$0.641 \pm 5.89e-3$	$0.382 \pm 2.051e-3$	$0.653 \pm 3.541e-3$
medgemma Trend +TFM timemixer	$0.401 \pm 1.04e-2$	$0.696 \pm 3.943e-3$	$0.409 \pm 7.085e-3$	$0.686 \pm 9.073e-3$
medgemma Trend lstm	$0.417 \pm 5.283e-3$	$0.701 \pm 2.298e-3$	$0.396 \pm 8.327e-3$	$0.694 \pm 2.259e-3$
medgemma Trend mlp	$0.425 \pm 1.544e-3$	$0.705 \pm 1.227e-3$	$0.402 \pm 4.629e-3$	$0.696 \pm 1.47e-3$
medgemma Trend patchtsmixer	$0.411 \pm 1.628e-3$	$0.7 \pm 1.69e-3$	$0.373 \pm 4.873e-2$	$0.666 \pm 4.055e-2$
medgemma Trend timemixer	$0.419 \pm 8.68e-4$	$0.701 \pm 6.009e-4$	$0.389 \pm 2.712e-3$	$0.693 \pm 1.612e-3$
medgemma zero_shot +TFM lstm	$0.463 \pm 3.495e-3$	$0.718 \pm 2.008e-3$	$0.483 \pm 4.484e-3$	$0.736 \pm 1.781e-3$
medgemma zero_shot +TFM mlp	$0.468 \pm 4.851e-4$	$0.726 \pm 7.529e-4$	$0.49 \pm 6.615e-3$	$0.737 \pm 3.856e-3$
medgemma zero_shot +TFM patchtsmixer	$0.424 \pm 1.662e-2$	$0.681 \pm 9.491e-3$	$0.446 \pm 7.676e-3$	$0.698 \pm 7.654e-3$
medgemma zero_shot +TFM timemixer	$0.446 \pm 3.23e-3$	$0.72 \pm 1.113e-3$	$0.476 \pm 1.064e-2$	$0.728 \pm 7.495e-3$
medgemma zero_shot lstm	$0.455 \pm 2.794e-3$	$0.717 \pm 2.637e-4$	$0.496 \pm 6.549e-3$	$0.74 \pm 3.214e-3$
medgemma zero_shot mlp	$0.462 \pm 2.986e-3$	$0.715 \pm 2.225e-3$	$0.492 \pm 2.232e-3$	$0.735 \pm 1.473e-3$
medgemma zero_shot patchtsmixer	$0.458 \pm 3.019e-3$	$0.716 \pm 2.53e-3$	$0.447 \pm 4.454e-2$	$0.709 \pm 3.68e-2$
medgemma zero_shot timemixer	$0.454 \pm 1.164e-3$	$0.713 \pm 1.441e-3$	$0.485 \pm 8.093e-3$	$0.731 \pm 4.914e-3$
No_sum_zero_shot lstm	$0.54 \pm 1.203e-2$	$0.764 \pm 6.374e-3$	$0.477 \pm 3.603e-3$	$0.734 \pm 4.113e-3$
No_sum_zero_shot mlp	$0.548 \pm 6.611e-3$	$0.77 \pm 3.377e-3$	$0.46 \pm 7.96e-3$	$0.718 \pm 6.132e-3$
No_sum_zero_shot patchtsmixer	$0.434 \pm 6.888e-2$	$0.709 \pm 5.228e-2$	$0.44 \pm 7.164e-2$	$0.719 \pm 5.262e-2$
No_sum_zero_shot timemixer	$0.508 \pm 1.576e-2$	$0.75 \pm 6.17e-3$	$0.444 \pm 2.188e-2$	$0.713 \pm 1.891e-2$
right lstm	$0.377 \pm 3.521e-2$	$0.632 \pm 3.204e-2$	$0.369 \pm 4.221e-2$	$0.605 \pm 4.497e-2$
right mlp	$0.38 \pm 1.781e-2$	$0.646 \pm 2.192e-2$	$0.216 \pm 9.502e-3$	$0.383 \pm 2.84e-2$
right patchtsmixer	$0.273 \pm 1.85e-4$	$0.5 \pm 1.2e-4$	$0.273 \pm 0e0$	$0.5 \pm 0e0$
right timemixer	$0.332 \pm 2.488e-2$	$0.59 \pm 3.827e-2$	$0.231 \pm 4.167e-3$	$0.424 \pm 1.226e-2$

Table 45: Cross-site transfer results - Mort. (part 2/2)