Table 6: Notations used in this paper.

Notation	Meaning
$\mathcal{M}, \mathcal{M}^{old}, \mathcal{M}^{new}$	set of all / old / new drugs
v, v_q	(query) patient record
c	disease / procedure
V	total number diseases and procedures in v
m, m_i, m_t	a single drug
$\mathcal{S}, \mathcal{S}_i, \mathcal{S}_t, \mathcal{Q}, \mathcal{Q}_i, \mathcal{Q}_t$	set of support & query set (for drug m_i / m_t)
$\mathcal{S}', \mathcal{S}'_i, \mathcal{Q}', \mathcal{Q}'_i$	set of negative support & query set (for drug m_i)
N_s, N_q	number of supports & queries
$*$	binary label
$\mathcal{L}$	loss
$\mathbf{h}$	ontology-enriched drug representation
$\alpha_{i,j}$	importance of ancestor m_j for drug m_i
$\mathbf{p}, \mathbf{p}^{(l)}, \mathbf{p}', \mathbf{p}'^{(l)}$	(l-th phenotype-driven) positive and negative prototype representation
$\mathcal{G}, \mathbf{g}^{(l)}$	l-th phenotype set and representation
$z^{(l)}, z'^{(l)}, \mathbf{z}, \mathbf{z}'$	distance to (l-th phenotype-driven) prototype of positive / negative supports
$\boldsymbol{\beta}$	drug-dependent importance weights over different phenotypes
$f_\phi(\cdot)$	drug recommendation model
$f_{\phi_a}(\cdot)$	drug-ancestor attention function
$f_{\phi_r}(\cdot)$	medical record embedding function
$f_{\phi_g}(\cdot)$	phenotype down-projection function
$f_{\phi_\beta}(\cdot)$	drug-dependent importance function
$\lvert\cdot\rvert$	cardinality
$\cap$	set intersection
$\emptyset$	empty set
$\sigma(\cdot)$	sigmoid function
$d(\cdot,\cdot)$	distance function
$\oplus$	concatentaion operator

A Detailed Experimental Settings

A.1 Dataset

MIMIC-IV (Johnson et al., 2020) provides the contemporary information which allows us to locate the year of each patient admission. The original data provides the timing information by a three year long date range (e.g., 2011-2013). We uniformly sample a year in the provided range as the admission year. We select our cohort by filtering out the following samples: (1) patients younger than 18-year-old; (2) admissions without disease or procedure; (3) admissions without medication. We filter out drugs with less than 20 admissions.

Table 7: Dataset statistics. We report the statistics for training / validation / testing separately.

Dataset	# Patients	# Admissions	# Drugs	# Diseases	# Procedures
MIMIC-IV	119K / 2K / 12K	216K / 3K / 15K	786 / 15 / 49	22K	12K
Claims	10K / 3K / 9K	37K / 6K / 29K	244 / 36 / 99	11K	-

A.2 Baselines

- **RNN (Choi et al., 2016b)** formulates the task as a multi-label sequence prediction task. The patient record is encoder into vector representation and then make the prediction using a fully connected network.
- **GAMENet (Shang et al., 2019b)** adopts a memory network to model the drug-drug interaction information and to memorize the historical patient condition.

- **SafeDrug (Yang et al., 2021)** leverages dual molecular encoders to capture the global and local molecule patterns.
- **MAML (Finn et al., 2017)** tries to learn good initialization parameters such that it can adapt to new drugs with a few gradient updates.
- **ProtoNet (Snell et al., 2017)** is a metric learning based method which embeds the support and query samples into hidden space such that the samples belong to the same class are closer to each other.
- **FEAT (Ye et al., 2021)** further proposes to use a set-to-set function to encode the support prototype instead of simple average.
- **CTML (Peng & Pan, 2022)** learns the task-specific representation from both support samples and the learning path.
- **MeLU (Lee et al., 2019)** builds upon the MAML (Finn et al., 2017) framework and can adapt to the new drugs with a few local updates.
- **TaNP (Lin et al., 2021)** incorporates a customization module which adapt the predictor parameters based on different tasks.

A.3 HYPERPARAMETERS

We use Adam as the optimizer with learning rate 1e-3. Linear warmup scheduler is used with the first 10% episodes as the warm-up episodes. Dropout is set to 0.5. We implement EDGE using PyTorch 1.11 and Python 3.8. The model is trained on a CentOS Linux 7 machine with 128 AMD EPYC 7513 32-Core Processors, 512 GB memory, and eight NVIDIA RTX A6000 GPUs.

B ADDITIONAL RESULTS

B.1 ADDITIONAL METRICS FOR MAIN RESULTS

Table 8: Additional results on new drug recommendation.

Method	MIMIC-IV			Claims		
	PR-AUC	Precision@500	Recall@500	PR-AUC	Precision@500	Recall@500
RNN	0.0288 ± 0.0057	0.0282 ± 0.0064	0.0243 ± 0.0017	0.0138 ± 0.0015	0.0138 ± 0.0014	0.0242 ± 0.0017
GAMENet	0.0727 ± 0.0076	0.0766 ± 0.0091	0.1913 ± 0.0121	0.0293 ± 0.0042	0.0286 ± 0.0043	0.0619 ± 0.0073
SafeDrug	0.0961 ± 0.0079	0.0887 ± 0.0079	0.2293 ± 0.0143	0.0135 ± 0.0014	0.0128 ± 0.0016	0.0134 ± 0.0009
MAML	0.0590 ± 0.0064	0.0543 ± 0.0059	0.1766 ± 0.0131	0.0234 ± 0.0036	0.0231 ± 0.0042	0.0415 ± 0.0051
ProtoNet	0.1164 ± 0.0098	0.0977 ± 0.0088	0.3371 ± 0.0197	0.0495 ± 0.0056	0.0572 ± 0.0061	0.1154 ± 0.0089
FEAT	0.0992 ± 0.0044	0.0873 ± 0.0061	0.3123 ± 0.0045	0.0473 ± 0.0023	0.0529 ± 0.0042	0.0962 ± 0.0025
CTML	0.0973 ± 0.0100	0.0871 ± 0.0100	0.2571 ± 0.0157	0.0137 ± 0.0012	0.0111 ± 0.0011	0.0249 ± 0.0022
MeLU	0.0577 ± 0.0074	0.0504 ± 0.0066	0.1245 ± 0.0102	0.0224 ± 0.0030	0.0231 ± 0.0038	0.0297 ± 0.0027
TaNP	0.0368 ± 0.0059	0.0280 ± 0.0042	0.0636 ± 0.0060	0.0170 ± 0.0015	0.0140 ± 0.0017	0.0275 ± 0.0034
EDGE	**0.1940 ± 0.0130***	**0.1459 ± 0.0107***	**0.4462 ± 0.0187***	**0.0781 ± 0.0068***	**0.0871 ± 0.0078***	**0.1975 ± 0.0121***

B.2 ANALYSIS OF DIFFERENT BACKBONE ENCODERS

Table 9: Analysis on the influence of different backbone encoders.

Method	ROC-AUC	P@100	R@100
EDGE + MLP	0.7961 ± 0.0065	0.1258 ± 0.0100	0.0838 ± 0.0057
EDGE + Transformer	0.8417 ± 0.0073	0.1684 ± 0.0117	0.1482 ± 0.0109
EDGE + GRU	0.8418 ± 0.0073	0.2024 ± 0.0141	0.1779 ± 0.0119
EDGE + Bi-GRU	0.8608 ± 0.0069	0.2251 ± 0.0139	0.1907 ± 0.0126

We further incorporate EDGE with different backbone encoders, ranging the MLP to Transformer and single layer GRU. Results can be found in table 9. MLP simply feeds the disease through a feed-forward neutral network and then take the average. It gives the lowest result as it cannot capture the

relationship among disease. Transformer leverages the attention mechanism and gets better results. However, it cannot outperform GRU and Bi-GRU. We suspect the main reason is that we skip the pre-training step for the Transformer. Theoretically, the performance of Transformer will be largely improved if we pre-train it via self-supervised learning and fine-tune it in our task, as shown in Li et al. (2020); Rasmy et al. (2021); Choi et al. (2019). We leave Transformer pre-training to future work.

B.3 Analysis of Different Distance Measures

Table 10: Analysis of the influence of different distance measures.

Method	**ROC-AUC**	**P@100**	**R@100**
EDGE + Cosine distance	0.8610 ± 0.0068	0.2118 ± 0.0135	0.1832 ± 0.0124
EDGE + Euclidean distance	0.8608 ± 0.0069	0.2251 ± 0.0139	0.1907 ± 0.0126

We test EDGE with two distance measures: cosine distance and euclidean distance. The results can be found in table 10. The two distance measures perform similarly on the MIMIC-IV dataset.

B.4 Full table of performance by drug category

Full version of table 4 can be found in table 11.

Table 11: Full performance by drug category.

Drug Category (ATC 2nd Level)	**ROC-AUC**
Antineoplastic agents	0.9882
Antivirals for systemic use	0.9754
Urologicals	0.9738
Anti-parkinson drugs	0.9611
Other alimentary tract and metabolism products in atc	0.9412
Cardiac therapy drugs	0.9287
Antiemetics and antinauseants	0.8941
Drugs for constipation	0.8934
Antihemorrhagics	0.8725
Immunosuppressants	0.8721
Mineral supplements	0.8708
Drugs for obstructive airway diseases	0.8674
Agents acting on the renin-angiotensin system	0.8485
Antithrombotic agents	0.8483
Antihistamines for systemic use	0.8198
Psycholeptics	0.8182
Nasal preparations	0.8149
Ophthalmologicals	0.8061
Psychoanaleptics	0.8019
Vitamins	0.7958
Blood substitutes and perfusion solutions	0.7940
Antibacterials for systemic use	0.7888
All other therapeutic products	0.7827
Antidiarrheals, intestinal antiinflammatory/antiinfective agents	0.7367
Calcium channel blockers	0.7285
Beta-adrenergic blocking agents	0.6074