

A Appendix

This is a supplementary document accompanying the main manuscript. Here we expand on some procedures and implementations used and provide additional results.

B Methods

The information reward for acquiring feature i given observed features \mathbf{x}_o is:

$$R(i, \mathbf{x}_o) = \mathbb{E}_{\mathbf{z}_i \sim p(\mathbf{z}_i | \mathbf{x}_o)} D_{KL}[q(\mathbf{z} | \mathbf{x}_i, \mathbf{x}_o) \| q(\mathbf{z} | \mathbf{x}_o)] - \mathbb{E}_{\mathbf{z}_\phi, \mathbf{x}_i \sim p(\mathbf{x}_\phi, \mathbf{x}_i | \mathbf{x}_o)} D_{KL}[q(\mathbf{z} | \mathbf{x}_\phi, \mathbf{x}_i, \mathbf{x}_o) \| q(\mathbf{z} | \mathbf{x}_\phi, \mathbf{x}_o)] \quad (2)$$

- \mathbf{x}_o : Currently observed features
- \mathbf{x}_i : Candidate feature i to be acquired
- \mathbf{x}_ϕ : Target variables (labels)
- $q(\mathbf{z} | \cdot)$: Posterior encoder distribution in VAE
- $p(\mathbf{x}_i | \mathbf{x}_o)$: Conditional distribution of feature i given observed features
- $D_{KL}[\cdot \| \cdot]$: Kullback-Leibler divergence The following algorithm describes our greedy procedure to select the next candidate from the set of unobserved features.

Algorithm 2: EDDI: Algorithm Overview

Require: Training dataset \mathbf{x}_{trn} , which is partially observed;
Test dataset \mathbf{x}_{tst} without any observation; Indices ϕ of target variables.

- 1: **TRAINING PHASE:**
- 2: Train Partial VAE by optimizing partial variational bound with \mathbf{x}_{trn}
- 3: **INFERENCE PHASE (Active Feature Acquisition):**
- 4: **for** each test instance **do**
- 5: $\mathbf{x}_O \leftarrow \emptyset$ ▷ no variable value has been observed for any test point
- 6: **repeat**
- 7: Choose variable x_i from $U \setminus \phi$ to maximize the information reward (Equation (2))
- 8: $\mathbf{x}_O \leftarrow x_i \cup \mathbf{x}_O$
- 9: **until** Stopping criterion reached (e.g. the time budget)
- 10: **end for**

In this algorithm, we use a PartialVAE, but can also be replaced by any other encoder model. For example, hidden representations can be [CLS] token, or target representation token.

C Open-World Active Feature Acquisition

D Experiments

D.1 Datasets

To enable effective pretraining of our universal neural inference model, we use OPENTABS - a comprehensive tabular dataset compiled from public sources including OpenML, UCI, Kaggle, and CATALOG. Following the collection of these datasets, we curate metadata for each of these datasets,

such as feature descriptions, dataset descriptions, and the range of classes for categorical datasets. We plan on open-sourcing this metadata on publication.

Dataset Name	ROC	Samples	Numerical	Categorical	Label Classes	Source
Breast	C	699	9	0	2	https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original
Bone	C	1479	2	7	3	https://archive.ics.uci.edu/dataset/3/connectionist+bench+choice
Diabetes	C	768	8	0	2	https://openml.org/d/37
Vehicle	C	846	18	0	4	https://archive.ics.uci.edu/dataset/149/starlog+vehicle+silhouettes
Satimage	C	6430	36	0	6	https://archive.ics.uci.edu/dataset/146/starlog+handsat+satellite
Sick	C	3772	7	22	2	http://archive.ics.uci.edu/dataset/102/thyroid+disease
Animaldata	C	797	0	4	6	https://pages.stern.nyu.edu/~simenof/AnimalData/Data/
Pcl	C	1109	21	0	2	https://openml.org/d/1068
Adult	C	48842	6	8	2	https://archive.ics.uci.edu/dataset/2/adult
PhishingWebsites	C	11055	0	30	2	https://archive.ics.uci.edu/dataset/327/phishing+websites
Cylinder-bands	C	540	18	21	2	https://archive.ics.uci.edu/dataset/32/cylinder+bands
MiceProtein	C	1080	77	4	8	https://archive.ics.uci.edu/dataset/342/mice+protein+expression
Car	C	1728	0	6	4	https://archive.ics.uci.edu/dataset/19/car+evaluation
Segment	C	2310	19	0	7	https://archive.ics.uci.edu/dataset/50/image+segmentation
Porto-seguro	R	2000	26	31	2	https://openml.org/d/44787
Amazon	C	2000	0	9	2	https://openml.org/d/44712
Elevators	R	16599	18	19	-	https://openml.org/d/216
Yprop	R	8885	251	0	-	https://openml.org/d/416
Topo	R	8885	266	267	-	https://openml.org/d/422
SAT11	R	4400	115	1	-	https://www.cs.ucb.edu/~calbl/algorithms/Projects/SAT11/
Diamonds	R	53940	6	3	-	https://openml.org/d/42225
House_sales	R	21613	20	1	-	https://openml.org/d/42731

Table 4: Table of the downstream datasets in our experiments, along with different information

D.2 Baselines

Here is the overview of the parameter count of our model and other baselines: ASPIRE Model size - 140, 165, 123 parameters CM2 Model size - 53, 784, 576. Our model is $2.6 \times$ of CM2.

D.3 Additional Results

Dataset	CM2 ↓	ASPIRE ↓	XGB ↓	MLP ↓
Elevators	0.292	0.098	0.340	0.360
House Sales	0.139	0.235	0.340	0.380
Diamonds	0.971	0.020	0.203	0.344
yprop	0.978	0.228	0.982	0.987
topo	0.336	0.204	0.922	0.986
Average	0.543	0.157	0.557	0.611

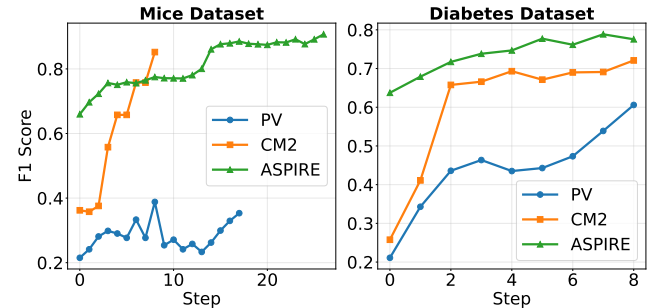


Figure 7: F1 scores at each feature acquisition step. PV indicates Partial VAE in EDDI.

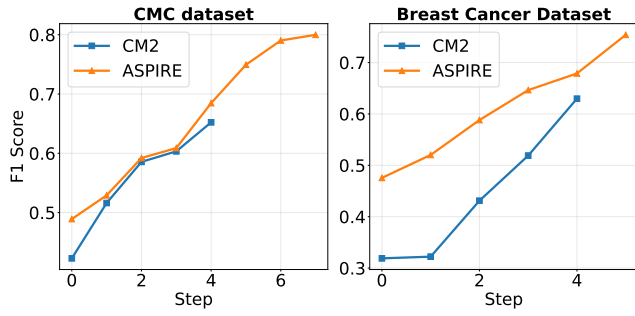


Figure 8: F1 scores at each feature acquisition step. PV indicates Partial VAE in EDDI.

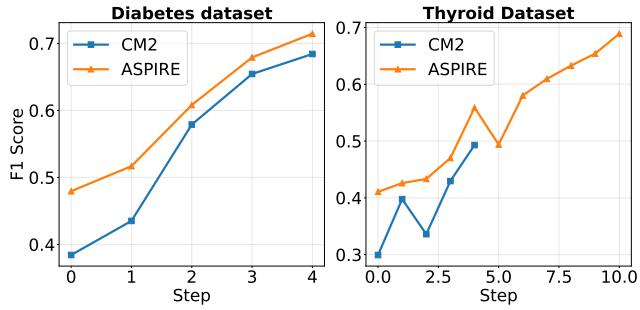


Figure 9: F1 scores at each feature acquisition step. PV indicates Partial VAE in EDDI.

Dataset	MLP	XGB	CM2	ASP.(0)	ASP.(5)
Diabetes	0.727	0.751	0.6963	0.84	0.8545
Vehicle	0.85	0.913	0.588	0.853	0.894
Satimage	0.80	0.89	0.8853	0.89	0.930
Sick	0.55	0.914	0.9357	0.8922	0.950
Pc1	0.476	0.624	0.8158	0.90	0.930
Adult	0.86	0.798	0.911	0.899	0.945
Breast	0.955	0.942	0.930	0.9416	0.966
Cmc	0.67	0.65	0.7265	0.80	0.830
PW	0.948	0.966	0.9912	0.878	0.930
Cylinder	0.532	0.734	0.8138	0.81	0.8829
MiceProtein	0.995	0.99	0.98	0.965	0.999
Car	0.85	0.93	0.99	0.967	0.99
Segment	0.97	0.99	0.9934	0.975	0.9621
Porto-seguro	0.49	0.49	0.725	0.88	0.8923
Amazon	0.87	0.483	0.97	0.7832	0.8756
Average	0.77	0.80	0.863	0.88	0.922

Table 5: Fine-tuning and model performance comparison with additional zero-shot support examples model.

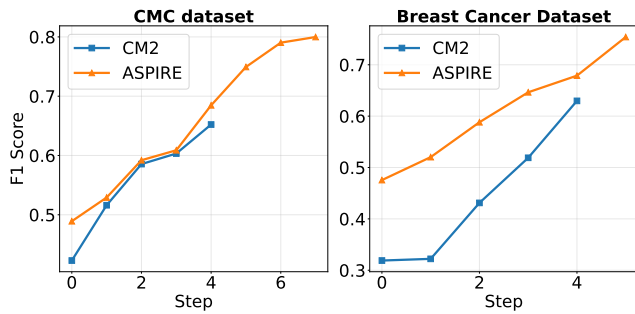


Figure 10: F1 scores at each feature acquisition step. PV indicates Partial VAE in EDDI.