QLabelMIL: Supplementary Material

Pedro C. Neto[1,3,4][0000−0003−1333−4889], Rita N. Lopes[3,4], and Lígia Prado e Castro[2]

[1] Unilabs.AI, Campus Biotech, Geneva, Switzerland
[2] Unilabs, Campus Biotech, Geneva, Switzerland
[3] Faculdade de Engenharia da Universidade do Porto
[4] Equal contribution

1 Additional description on the dataset classes

In contrast to the other conditions, H. pylori is not easily detected on H&E, having other stains utilized in clinical practice. Its prevalence is roughly 44% of the worldwide population [1]. Atrophy has a high prevalence too [7], but is affected by high inter-pathologist variability. It presents several challenges to AI models due to its definition being the absence of glands. It is more prominent in H. pylori-positive patients [7]. Intestinal Metaplasia occurs with the replacement of gastric cells with intestinal cells [3]. Being present in 17% of the worldwide population, it represents a major concern for gastric cancer [5]. The highest risk of progression to cancer is posed by dysplasia [6], characterized by architectural irregularities, cellular crowding, an increased nuclear-to-cytoplasmic ratio, and loss of cellular polarity [6]. It is widely less prominent worldwide [6], and the existence of active inflammation makes it harder to diagnose [2]. Acute inflammation, commonly referred to as activity, indicates active gastritis and is related to the presence of neutrophils in the mucosa. Also primarily associated with H. pylori [4], chronic inflammation is estimated to affect nearly 50% of the global population [4].

2 Hyperparameter Tuning

Each search included 30 trials with a 20-epoch limit using the Tree-structured Parzen Estimator (TPE) algorithm and a Hyperband pruner with a reduction factor of 3 and a grace period of 6 epochs. In order to establish a baseline, the initial trials for each model used the hyperparameters specified in the corresponding literature or author's implementation. Hardcoded parameters were largely excluded from tuning, but hidden dimensions, gradient accumulation steps, and the number of attention branches were made tunable regardless of the original implementation. When applicable, all variants of a model were implemented, and a tunable hyperparameter was introduced to choose between them. For the GCN variant of the proposed architecture, 2 graph convolution layers were used.

3 6-Classes Calibration Results

Results for calibration of the models. Similar to the performance, QLabelMIL seems to be strongly impacted by the lack of data balancing.

References

1. Chen, Y.C., Malfertheiner, P., Yu, H.T., Kuo, C.L., Chang, Y.Y., Meng, F.T., Wu, Y.X., Hsiao, J.L., Chen, M.J., Lin, K.P., et al.: Global prevalence of helicobacter pylori infection and incidence of gastric cancer between 1980 and 2022. Gastroenterology **166**(4), 605–619 (2024)
2. Pennelli, G., Grillo, F., Galuppini, F., Ingravallo, G., Pilozzi, E., Rugge, M., Fiocca, R., Fassan, M., Mastracci, L.: Gastritis: update on etiological features and histological practical approach. Pathologica **112**(3), 153 (2020)
3. Shah, S.C., Gawron, A.J., Mustafa, R.A., Piazuelo, M.B.: Histologic subtyping of gastric intestinal metaplasia: overview and considerations for clinical practice. Gastroenterology **158**(3), 745–750 (2020)
4. Sipponen, P., Maaroos, H.I.: Chronic gastritis. Scandinavian journal of gastroenterology **50**(6), 657–667 (2015)
5. Soroorikia, S., Kazeminia, M., Qaderi, K., Ziapour, A., Hodhodi, T., Javanbakht, Z.: Global prevalence of gastric intestinal metaplasia: a systematic review and meta-analysis. Systematic Reviews **13**(1), 247 (2024)
6. Sung, J.K.: Diagnosis and management of gastric dysplasia. The Korean journal of internal medicine **31**(2), 201 (2016)
7. Yin, Y., Liang, H., Wei, N., Zheng, Z.: Prevalence of chronic atrophic gastritis worldwide from 2010 to 2020: an updated systematic review and meta-analysis. Annals of Palliative Medicine **11**(12), 3697703–3693703 (2022)

Table 1. Comparison of calibration across architectures, measured by ECE for six target pathologies. Best values per column within each pathology are **bolded**, and second best values are underlined. AUROC: Area Under the Receiver Operating Characteristic curve; GCN: Graph Convolutional Network.

Pathology	Architecture	10× Magnification		40× Magnification	
		ResNet18	ResNet50	ResNet18	ResNet50
Intestinal Metaplasia	ABMIL	0.2258	0.2103	0.0632	0.1975
	CLAM	0.0670	0.1067	0.1175	**0.0536**
	ACMIL	0.1056	0.1498	0.1720	0.1020
	MambaMIL	0.1125	0.0606	0.0633	0.1359
	TransMIL	0.2572	0.0776	0.1313	0.1369
	QLabelMIL	**0.0603**	**0.0256**	0.1471	0.1120
	QLabelMIL (GCN)	0.1414	0.1450	**0.0286**	0.1146
Chronic Inflammation	ABMIL	0.4074	0.2473	0.3210	0.2496
	CLAM	0.3701	0.2861	**0.2365**	0.2259
	ACMIL	**0.2001**	0.3160	0.3018	0.2638
	MambaMIL	0.2778	0.2809	0.2996	**0.2174**
	TransMIL	0.2547	0.3890	0.2777	0.2300
	QLabelMIL	0.2609	0.3038	0.2728	0.2427
	QLabelMIL (GCN)	0.3387	**0.1707**	0.3200	0.2879
Acute Inflammation	ABMIL	0.5599	0.6551	0.1464	0.1163
	CLAM	0.1930	0.1830	0.3707	0.3431
	ACMIL	0.0540	0.1831	**0.0239**	**0.0382**
	MambaMIL	**0.0497**	0.0875	0.0387	0.2896
	TransMIL	0.1308	0.1055	0.1981	0.2444
	QLabelMIL	0.1359	**0.0840**	0.1455	0.1412
	QLabelMIL (GCN)	0.2694	0.1339	0.0538	0.0726
Dysplasia	ABMIL	0.0545	0.0409	0.0078	0.0613
	CLAM	0.0110	0.0526	0.1238	0.2079
	ACMIL	**0.0034**	0.0221	**0.0048**	**0.0168**
	MambaMIL	0.0104	0.0293	0.0236	0.3292
	TransMIL	0.0214	0.0293	0.0218	0.1636
	QLabelMIL	0.1717	0.0166	0.0249	0.0240
	QLabelMIL (GCN)	0.1507	**0.0127**	0.0567	0.0214
H. pylori	ABMIL	0.0988	0.0915	0.0594	0.0854
	CLAM	**0.0412**	0.1567	0.1193	**0.0284**
	ACMIL	0.1218	**0.0373**	0.0539	0.0492
	MambaMIL	0.0416	0.0771	0.1559	0.1059
	TransMIL	0.1044	0.0533	0.0797	0.1030
	QLabelMIL	0.0520	0.0499	0.0665	0.1092
	QLabelMIL (GCN)	0.0556	0.1273	**0.0391**	0.0497
Atrophy	ABMIL	**0.0364**	0.1325	0.0528	0.1174
	CLAM	0.0402	0.2143	**0.0356**	0.0809
	ACMIL	0.1130	0.1043	0.1813	0.0931
	MambaMIL	0.1101	0.0667	0.0597	**0.0765**
	TransMIL	0.2010	**0.0396**	0.1005	0.1375
	QLabelMIL	0.0592	0.0877	0.1659	0.1420
	QLabelMIL (GCN)	0.0913	0.2031	0.0397	0.1076
Macro-Average	**ABMIL**	0.2305	0.2296	0.1084	0.1379
	CLAM	0.1204	0.1666	0.1672	0.1566
	ACMIL	**0.0996**	0.1354	0.1230	**0.0939**
	MambaMIL	0.1004	0.1004	0.1068	0.1924
	TransMIL	0.1616	0.1157	0.1349	0.1692
	QLabelMIL	0.1233	**0.0946**	0.1371	0.1285
	QLabelMIL (GCN)	0.1745	0.1321	**0.0896**	0.1090