Supplemental Material for Label-Specific Multi-Semantics Metric Learning for Multi-Label Classification: Global Consideration Helps

Jun-Xiang Mao[1,2], **Wei Wang**[3*] and **Min-Ling Zhang**[1,2†]
[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China
[3]The University of Tokyo, Japan
maojx@seu.edu.cn, wangw@g.ecc.u-tokyo.ac.jp, zhangml@seu.edu.cn

1 Further Experimental Results

1.1 Comparative Studies

We employ ten-fold cross-validation to evaluate the proposed LIMIC approach and three well-established multi-label metric learning approaches $\mathcal{B} \in \{\text{LM}, \text{LJE}, \text{COMMU}\}$ on 8 benchmark data sets. Table 7 reports detailed experimental results in terms of *Average precision*, *Macro-F1*, and *Macro-averaging AUC*, which are not covered in the "Comparative Studies" subsection of the main body due to page limit. Furthermore, pairwise t-test at 0.05 significance level is conducted to demonstrate whether the performance difference between $\mathcal{A}$-LIMIC and $\mathcal{A}$-$\mathcal{B}$ is significant, where the resulting win/tie/loss counts are reported in Table 1, 2, and 3.

1.2 Further Analyses

Pairwise t-test at 0.05 significance level is conducted to analyze whether LIMIC performs statistically better than its variants. Tabel 4, 5, and 6 summarize win/tie/loss counts on each evaluation metric, which demonstrate LIMIC is statistically superior to its variants in terms of all evaluation metrics.

Metrics	BR-KNN-LIMIC **against**			
	BR-KNN	BR-KNN-LM	BR-KNN-LJE	BR-KNN-COMMU
Hamming Loss	7/1/0	5/1/2	7/1/0	5/3/0
Ranking Loss	6/2/0	5/3/0	6/2/0	2/6/0
Coverage	6/2/0	6/2/0	6/2/0	4/4/0
Average precision	6/2/0	8/0/0	6/2/0	4/4/0
Macro-F1	7/1/0	4/3/1	7/1/0	5/3/0
Macro-averaging AUC	6/2/0	6/2/0	6/2/0	3/5/0
In Total	**38/10/0**	**34/11/3**	**38/10/0**	**23/25/0**

Table 1: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for BR-KNN-LIMIC against other compared approaches.

Metrics	ML-KNN-LIMIC **against**			
	ML-KNN	ML-KNN-LM	ML-KNN-LJE	ML-KNN-COMMU
Hamming Loss	5/3/0	4/3/1	5/3/0	4/4/0
Ranking Loss	7/1/0	5/3/0	7/1/0	4/4/0
Coverage	5/3/0	4/4/0	5/3/0	3/4/1
Average precision	6/2/0	6/2/0	6/2/0	5/3/0
Macro-F1	6/2/0	6/2/0	6/2/0	6/2/0
Macro-averaging AUC	6/2/0	5/3/0	6/2/0	4/4/0
In Total	**35/13/0**	**30/17/1**	**35/13/0**	**26/21/1**

Table 2: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for ML-KNN-LIMIC against other compared approaches.

*The work was done when Wei Wang was with Southeast University.

†Corresponding author.

Metrics	RELIAB-KNN-LIMIC **against**			
	RELIAB-KNN	RELIAB-KNN-LM	RELIAB-KNN-LJE	RELIAB-KNN-COMMU
Hamming Loss	6/2/0	6/2/0	6/1/1	6/2/0
Ranking Loss	6/2/0	5/3/0	6/2/0	5/3/0
Coverage	7/1/0	6/0/2	6/1/1	7/1/0
Average precision	5/3/0	7/0/1	7/1/0	6/1/1
Macro-F1	8/0/0	8/0/0	8/0/0	8/0/0
Macro-averaging AUC	7/0/1	8/0/0	8/0/0	6/0/2

Table 3: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for RELIAB-KNN-LIMIC against other compared approaches.

Metrics	BR-KNN-LIMIC **against**		
	BR-KNN-LIMIC-NG	BR-KNN-LIMIC-NL	BR-KNN-LIMIC-NGL
Hamming Loss	4/4/0	3/4/1	4/3/1
Ranking Loss	7/1/0	5/3/0	7/1/0
Coverage	6/1/1	4/3/1	5/2/1
Average precision	7/1/0	6/2/0	8/0/0
Macro-F1	7/0/1	3/4/1	6/1/1
Macro-averaging AUC	6/1/1	4/2/2	8/0/0
In Total	**37/8/3**	**25/18/5**	**38/7/3**

Table 4: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for BR-KNN-LIMIC against its variants coupled with BR-KNN.

Metrics	ML-KNN-LIMIC **against**		
	ML-KNN-LIMIC-NG	ML-KNN-LIMIC-NL	ML-KNN-LIMIC-NGL
Hamming Loss	6/2/0	4/4/0	4/4/0
Ranking Loss	5/2/1	5/3/0	5/2/1
Coverage	4/2/2	3/3/2	4/3/1
Average precision	6/1/1	5/1/2	6/1/1
Macro-F1	7/0/1	3/3/2	5/1/2
Macro-averaging AUC	6/0/2	5/3/0	6/1/1
In Total	**34/7/7**	**25/17/6**	**30/13/6**

Table 5: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for ML-KNN-LIMIC against its variants coupled with ML-KNN.

Metrics	RELIAB-KNN-LIMIC **against**		
	RELIAB-KNN-LIMIC-NG	RELIAB-KNN-LIMIC-NL	RELIAB-KNN-LIMIC-NGL
Hamming Loss	5/3/0	6/2/0	5/3/0
Ranking Loss	4/2/2	5/3/0	6/1/1
Coverage	7/1/0	4/3/1	5/3/0
Average precision	6/1/1	4/3/1	4/4/0
Macro-F1	6/1/1	7/0/1	5/2/1
Macro-averaging AUC	7/0/1	5/3/0	7/1/0
In Total	**35/8/5**	**31/14/3**	**32/14/2**

Table 6: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for RELIAB-KNN-LIMIC against its variants coupled with RELIAB-KNN.

Compared Algorithms	Data Sets							
	CAL500	emotions	birds	genbase	medical	image	scene	yeast
	Average precision ↑							
BR-KNN	0.463±0.009	0.700±0.049•	0.358±0.043•	0.993±0.005	0.801±0.025•	0.788±0.023•	0.850±0.012•	0.762±0.018•
BR-KNN-LM	0.451±0.007•	0.711±0.038•	0.332±0.052•	0.989±0.009•	0.843±0.032•	0.789±0.020•	0.847±0.013•	0.744±0.021•
BR-KNN-LJE	0.456±0.009	0.771±0.042•	0.390±0.042•	0.994±0.005	0.767±0.030•	0.767±0.024•	0.812±0.017•	0.750±0.019•
BR-KNN-COMMU	**0.467±0.010**	0.700±0.049•	0.356±0.043•	0.993±0.005	0.802±0.025•	0.789±0.023•	0.850±0.012	0.763±0.018
BR-KNN-LIMIC	0.464±0.010	**0.800±0.040**	**0.441±0.065**	**0.996±0.005**	**0.881±0.024**	**0.807±0.024**	**0.857±0.019**	**0.767±0.022**
ML-KNN	0.494±0.008	0.712±0.042•	0.414±0.052•	0.989±0.008	0.819±0.020•	0.789±0.021•	0.867±0.017•	0.765±0.018•
ML-KNN-LM	0.493±0.007	0.719±0.019•	0.405±0.056•	0.995±0.006	0.849±0.035•	0.789±0.017•	0.857±0.015•	0.753±0.025•
ML-KNN-LJE	0.491±0.007	0.762±0.036•	0.425±0.039•	0.995±0.004	0.773±0.025•	0.772±0.023•	0.817±0.020•	0.752±0.018•
ML-KNN-COMMU	**0.497±0.010**	0.712±0.042•	0.413±0.052•	0.989±0.008•	0.817±0.020•	0.788±0.021•	0.867±0.016	0.764±0.019
ML-KNN-LIMIC	0.495±0.008	**0.789±0.037**	**0.461±0.054**	**0.996±0.002**	**0.891±0.019**	**0.811±0.025**	**0.868±0.013**	**0.768±0.022**
RELIAB-KNN	0.589±0.010	0.797±0.040•	0.498±0.035•	0.995±0.007	0.857±0.023•	0.831±0.025	0.885±0.015•	0.793±0.020•
RELIAB-KNN-LM	0.583±0.008•	0.816±0.038•	0.531±0.033•	0.987±0.006•	0.873±0.024•	**0.849±0.027**○	0.904±0.013•	0.807±0.024•
RELIAB-KNN-LJE	0.588±0.010	0.835±0.035•	0.524±0.036•	0.984±0.005•	0.826±0.025•	0.816±0.023•	0.852±0.016•	0.784±0.018•
RELIAB-KNN-COMMU	**0.591±0.009**○	0.796±0.040•	0.498±0.035•	**0.996±0.007**•	0.857±0.023•	0.831±0.025	0.885±0.015•	0.794±0.020•
RELIAB-KNN-LIMIC	0.589±0.012	**0.857±0.037**	**0.574±0.041**	**0.996±0.008**	**0.898±0.028**	0.830±0.026	**0.927±0.018**	**0.813±0.024**
	Macro-F1 ↑							
BR-KNN	0.095±0.005•	0.463±0.045	0.038±0.018•	0.583±0.039•	0.169±0.018•	0.591±0.034•	0.721±0.014•	0.416±0.019•
BR-KNN-LM	**0.103±0.006**○	0.490±0.033•	0.060±0.035•	0.645±0.057	0.321±0.051•	0.609±0.033	0.738±0.026	0.389±0.023•
BR-KNN-LJE	0.089±0.005•	0.579±0.049	0.045±0.037•	0.636±0.050•	0.219±0.033•	0.549±0.033•	0.665±0.027•	0.383±0.015•
BR-KNN-COMMU	0.097±0.007	0.463±0.045•	0.038±0.018•	0.583±0.039•	0.170±0.018•	0.591±0.034•	0.721±0.014	0.417±0.021
BR-KNN-LIMIC	0.097±0.006	**0.615±0.016**	**0.103±0.028**	**0.656±0.042**	**0.398±0.043**	**0.621±0.033**	**0.743±0.031**	**0.423±0.027**
ML-KNN	0.053±0.003•	0.361±0.050	0.010±0.007•	0.525±0.034	0.210±0.027•	0.575±0.030•	0.737±0.021•	0.367±0.011•
ML-KNN-LM	0.053±0.003	0.438±0.036•	0.016±0.011•	0.561±0.036•	0.295±0.043•	0.584±0.025•	0.735±0.026	0.342±0.018•
ML-KNN-LJE	0.048±0.002•	0.544±0.043	0.025±0.024•	0.598±0.046	0.214±0.017•	0.521±0.026•	0.656±0.026•	0.332±0.012•
ML-KNN-COMMU	**0.055±0.003**	0.361±0.050•	0.010±0.007•	0.525±0.034•	0.212±0.027•	0.576±0.031•	0.738±0.022	0.367±0.015•
ML-KNN-LIMIC	0.053±0.003	**0.593±0.033**	**0.072±0.030**	**0.609±0.034**	**0.371±0.024**	**0.643±0.036**	**0.747±0.019**	**0.419±0.024**
RELIAB-KNN	0.135±0.004•	0.512±0.034•	0.164±0.019•	0.612±0.036•	0.289±0.022•	0.634±0.032•	0.745±0.015•	0.458±0.012•
RELIAB-KNN-LM	0.147±0.002•	0.577±0.029•	0.194±0.016•	0.679±0.033•	0.358±0.018•	0.629±0.034•	0.752±0.019•	0.436±0.018•
RELIAB-KNN-LJE	0.129±0.003•	0.586±0.042•	0.189±0.022•	0.656±0.032•	0.347±0.020•	0.598±0.027•	0.710±0.026•	0.427±0.015•
RELIAB-KNN-COMMU	0.137±0.003•	0.512±0.034•	0.164±0.019•	0.612±0.036•	0.289±0.022•	0.633±0.032•	0.746±0.015•	0.458±0.012•
RELIAB-KNN-LIMIC	**0.155±0.003**	**0.623±0.033**	**0.234±0.020**	**0.726±0.042**	**0.395±0.031**	**0.682±0.035**	**0.758±0.016**	**0.482±0.016**
	Macro-averaging AUC ↑							
BR-KNN	0.540±0.016	0.737±0.041•	0.670±0.040•	0.985±0.019	0.880±0.032•	0.840±0.019•	**0.936±0.006**•	**0.695±0.021**•
BR-KNN-LM	0.541±0.014	0.742±0.024•	0.658±0.056•	0.973±0.020•	0.879±0.051	0.826±0.020•	0.922±0.009•	0.655±0.030•
BR-KNN-LJE	0.525±0.018	0.807±0.030•	0.678±0.039•	0.986±0.015	0.837±0.033•	0.818±0.018•	0.905±0.012•	0.659±0.014•
BR-KNN-COMMU	**0.542±0.018**	0.737±0.041•	0.670±0.040•	0.986±0.019	0.880±0.031	0.840±0.019•	**0.936±0.006**	**0.695±0.022**
BR-KNN-LIMIC	**0.542±0.017**	**0.826±0.019**	**0.752±0.031**	**0.994±0.012**	**0.892±0.038**	**0.852±0.020**	**0.930±0.008**	**0.694±0.020**
ML-KNN	0.526±0.011	0.720±0.041•	0.666±0.044•	0.985±0.018	0.876±0.032•	0.834±0.019•	**0.933±0.006**•	0.677±0.020•
ML-KNN-LM	0.524±0.016	0.729±0.025•	0.642±0.062•	0.988±0.013	0.878±0.050	0.827±0.021•	0.921±0.009•	0.650±0.034•
ML-KNN-LJE	0.512±0.018	0.787±0.024•	0.664±0.048•	0.986±0.015	0.833±0.034•	0.811±0.018•	0.903±0.013•	0.652±0.014•
ML-KNN-COMMU	**0.530±0.011**	0.720±0.041•	0.664±0.044•	0.985±0.018•	0.875±0.031	0.834±0.019•	**0.933±0.006**	0.679±0.021
ML-KNN-LIMIC	0.525±0.011	**0.818±0.022**	**0.743±0.032**	**0.994±0.012**	**0.892±0.037**	**0.847±0.021**	0.929±0.009	**0.691±0.019**
RELIAB-KNN	0.651±0.013•	0.844±0.033•	0.672±0.042•	0.986±0.019•	0.897±0.035•	0.852±0.016•	**0.945±0.005**○	0.734±0.020•
RELIAB-KNN-LM	0.653±0.010•	0.848±0.028•	0.643±0.035•	0.989±0.020•	0.898±0.029•	0.837±0.013•	0.934±0.004•	0.695±0.018•
RELIAB-KNN-LJE	0.633±0.011•	0.857±0.026•	0.667±0.038•	0.987±0.018•	0.845±0.027•	0.817±0.015•	0.928±0.003•	0.703±0.015•
RELIAB-KNN-COMMU	**0.659±0.012**○	0.844±0.033•	0.672±0.042•	0.985±0.019•	0.897±0.035•	0.852±0.016•	**0.945±0.005**○	0.734±0.020•
RELIAB-KNN-LIMIC	0.657±0.015	**0.896±0.030**	**0.721±0.039**	**0.995±0.018**	**0.913±0.033**	**0.878±0.020**	0.941±0.005	**0.786±0.016**

Table 7: Predictive performance of each compared approach (mean±std) in terms of *Average precision*, *Macro-F1*, and *Macro-average AUC*. ↑ (↓) indicates the larger (smaller) the value, the better the performance. The best results are highlighted in **boldface**. In addition, •/○ indicates whether $\mathcal{A}$-LIMIC ($\mathcal{A} \in \{$BR-KNN, ML-KNN, RELIAB-KNN$\}$) achieves significantly superior/inferior to the compared approach on each data set in terms of different evaluation metrics (pairwise t-test at 0.05 significance level).