Hierarchical Text Classification and its Foundations: a review of current research

Zangari A., Marcuzzo M., Rizzo M., Giudice L.,
Albarelli A., Gasparetto A.

Supplemental Material

Code and datasets

The code for this manuscript can be found at the following link:

- https://gitlab.com/distration/dsi-nlp-publib/-/tree/main/htc-survey-24

The repository also contains the pre-compiled taxonomies for the datasets.

All training splits for the publicly available datasets are available on Zenodo:

- https://zenodo.org/records/7319519

Note that splits for the RCV1 and BGC are not made available as we do not own the copyright to these datasets. However, scripts to replicate the splits are provided in the repository.

Validation procedure

We validate all method on 20% of the training split using the Linux Bugs dataset. We test several combinations of hyperparameters to select the best ones for each method. Because of time constraints, we are unable to perform validation on all datasets, though our experiments provide a decent approximation and, regardless, are not aimed at the best achievable performance, but rather an inter-method comparison. We use 3-fold CV during validation and base our choice of hyperparameters on the average macro F1-score over the three runs. The optimal number of epochs is selected by using early stopping with patience set to 5 (unless specified otherwise by authors). The following tables document the hyperparameters tested, highlighting the best one in bold.

Reading the tables

In each table, parameters utilized during validation are listed on the right. The bold value is the one that has been found to work best.

MATCH [1]

We do not use metadata, nor the joint embedding training (as the first are unavailable, and the second did not work on our datasets). Instead, we use Word2Vec embeddings, and we validate the number of epochs used for training. Other parameters are set to the values used in the original work (Table 1).

Hyperparameter	Value tried
Word2Vec epochs	10, **20**
Learning rate	1e-3, 1e-4, 1**e-5**, 5e-5
Batch Size	32, **64**, 256
N. of layers (n_layers)	**3**, 6
N. probe tokens (n_probes)	**8**
N. attention heads (n_aheads)	**2**
Max sequence length	**500**, 600
Number of epochs (early stopping)	40, early stopping every 100 batches

Table 1. Selected hyperparameters for MATCH on the Linux Bugs dataset

HiAGM [2]

First the best model between HiAGM-TP and HiAGM-LA was chosen. Each was tested with the GCN or TreeLSTM model using the best hyperparameters described in the original work. **HiAGM-TP** with **TreeLSTM** gave the best results on the Bugs dataset; therefore, we fine-tuned the hyperparameters on this combination as described in Table 2.

Hyperparameter	Value tried
Learning rate	1e-3, **2e-4**, 1e-4, 1e-5
Batch Size	**256**
Max sequence length	256, **512**
Number of epochs (early stopping)	150 (patience 5)

Table 2. Selected hyperparameters for HiAGM-TP on the Bugs dataset

HBGL [3]

Hyperparameters for HBGL were tuned on the Bugs dataset as described in Table 3. Because the model's training was particularly slow, we only repeat the 3-fold cross-validation once.

Hyperparameter	Value tried
Learning rate	1e-6, **3e-5**, 1e-4
Batch Size	12, **16**

Table 3. Selected hyperparameters for HBGL on the Bugs dataset

GACaps-HTC [4]

Hyperparameters for GACaps-HTC were tuned on the Bugs dataset as described in Table 4. Like for HBGL, we only reported performance metrics over one 3-fold cross-validation.

Hyperparameter	Value tried
Learning rate	1e-5, **1e-4**, 1e-3
Batch Size	**16**, 32

Table 4. Selected hyperparameters for GACaps-HTC on the Bugs dataset

BERT [5]

With the BERT classifier, we test all possible combinations of hyperparameters defined in Table 5. Batch size is always set to 8 in validation and test. Notably, tests with gradient accumulation that simulate higher batch sizes did not display improvements.

Hyperparameter	Value tried
Learning rate	1e-6, 1e-5, **2e-5**, 5e-5
Batch Size	**8**, 32, 64
Advanced preprocessing	yes, **no**
Number of epochs (early stopping)	1-10, selected 4

Table 5. Selected hyperparameters for BERT on the Bugs dataset

XML-CNN [6]

The straightforward XML-CNN classifier was validated with both GloVe and FastText embeddings, providing best results with the former. The rest of the parameters are standard for hyperparameter tuning, as in Table 6.

Hyperparameter	Value tried
Embeddings initialization (pre-trained)	**GloVe**, FastText
Learning rate	**2e-3**, 2e-4, 2e-5
Batch Size	32, **64**, 128
N. channels convolution (output_channel)	32, **64**
Pooling output size (dynamic_pool_length)	32, **64**
Advanced preprocessing	yes, **no**
Max sequence length	256, **512**
Number of epochs (early stopping)	1-40, selected 5

Table 6. Selected hyperparameters for XML-CNN on the Bugs dataset

SVM [7]

We perform a grid search over the hyper-parameters as described in Table 7. The grid search is performed over a 5-fold CV, and the best-chosen parameters are utilized for a re-training on the whole training split. Note that there are no bold values, as they are dependent on the dataset.

Hyperparameter	Value tried
estimator__C	0.01, 0.1, 1, 5, 10
estimator__max_iter	100, 200, 500, 1000, 2000
objective	multiclass / multilabel

Table 7. Selected hyperparameters for the SVMs on the Bugs dataset

CHAMP / MATCH [8, 9]

When testing models with the CHAMP (CL) and MATCH (ML) losses, we use the default hyper-parameters found in their original implementations. For the MATCH loss, we use values $\lambda_1 = 1e-8$ and $\lambda_2 = 1e-10$. For CHAMP $\beta = 0.2$.

Technical setup

All language models are trained using PyTorch 1.11.0 on a Nvidia RTX 2080 Ti.

Dataset is split into training and testing sets using a 3-fold CV repeated 2 times. The final results reported are the average of metrics over the 6 runs. In order to ensure that algorithms are always tested and validated on the same splits we set the random seed using the "RepeatedStratifiedKFold" utility available in scikit-learn [10]. The splitting strategy also ensures that splits are stratified, meaning that they contain the same proportion of labels as the overall dataset.

Issues with other tested methods

In this section, we briefly address other methods we tested, and the reason why we were unable to reproduce them.

HDLTex [11]

While we do not discuss it directly in our work, HDLTex is utilized as a reference for many other works. This framework is based on a stack of different neural architectures (DNN, RNN, CNN) and is released publicly as a package. Despite the good intent of the authors, there are little to no instructions on how to utilize this code (other than installation directives). Regardless, the main reason we were unable to test this method was based on resources; as it is, the method loads a large quantity of data directly onto memory, which would not work for most of our datasets.

Hi-MATCH [12]

The authors of this work take large inspiration from HiAGM, and their implementation is indeed taken from the ones of the original authors. The most relevant issue we've encountered with this weapon has been that of computational expensiveness; the authors do not give clear instructions on how many epochs the model should be trained, and - because it utilized BERT as one of its components – training times are very consistent. We tried running the method for a number of epochs that is usually standard for BERT-based models (4 to 5 epochs), yet the results were lacklustre. The original configuration file for this method seems to suggest that it should run extensively (up to 150 epochs), which is however unfeasible to do with our hardware limitations.

RLHR [13]

We were able to implement correctly the method proposed by the authors. However, the task it seeks to solve is slightly different: the model is trained for binary classification of a document with a label, while all other models we compare are trained for a multi-label task, making this method unsuitable for our experiments.

HE-HMTC [14]

Though this method is fairly documented, we were unable to run it because of various issues with dependencies and outdated packages. The code largely depends on old versions of the Tensorflow and Keras frameworks and adapting it to newer version proved too cumbersome.

HFT-CNN [15]

The authors of HE-HMTC cite this work as state-of-the-art in HTC; therefore, we explored it as a possible addition to our tests. We ultimately discarded it because of two main reasons: (1) it was originally devised for short texts only, therefore making it inapplicable to a wide range of HTC scenarios, and (2) it is written in a framework we are unfamiliar with (Chainer) and are therefore unable to adapt it to other datasets.

Complete results

Linux Bugs dataset

BUGS									
BERT - epoch 4									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5129	0.5260	0.6246	0.4864	0.6227	0.4885	0.5264	0.4720
	2	0.5063	0.5273	0.6443	0.4898	0.6430	0.4917	0.5277	0.4430
	3	0.5121	0.5294	0.6619	0.4934	0.6610	0.4947	0.5296	0.4512
	avg	0.5104	0.5276	0.6436	0.4899	0.6422	0.4916	0.5279	0.4554
2	1	0.5142	0.5210	0.6252	0.4791	0.6239	0.4811	0.5218	0.4646
	2	0.5119	0.5128	0.6639	0.4750	0.6629	0.4766	0.5134	0.4734
	3	0.4846	0.5059	0.6569	0.4603	0.6550	0.4626	0.5066	0.4597
	avg	0.5036	0.5132	0.6487	0.4715	0.6472	0.4735	0.5139	0.4659
final avg		**0.5070**	**0.5204**	**0.6461**	**0.4807**	**0.6447**	**0.4825**	**0.5209**	**0.4606**
	std	0.0113	0.0093	0.0178	0.0121	0.0180	0.0118	0.0091	0.0119
BUGS									
BERT + CL - epoch 4									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4955	0.4979	0.6518	0.4502	0.6506	0.4514	0.4982	0.4049
	2	0.4787	0.4927	0.6410	0.4466	0.6405	0.4480	0.4934	0.4044
	3	0.4789	0.4957	0.6599	0.4444	0.6596	0.4452	0.4960	0.3930
	avg	0.4844	0.4954	0.6509	0.4471	0.6502	0.4482	0.4959	0.4008
2	1	0.4782	0.5112	0.6455	0.4600	0.6447	0.4613	0.5117	0.4215
	2	0.4808	0.4893	0.6380	0.4361	0.6371	0.4372	0.4896	0.3774
	3	0.4809	0.4920	0.6320	0.4433	0.6308	0.4443	0.4921	0.4269
	avg	0.4799	0.4975	0.6385	0.4465	0.6375	0.4476	0.4978	0.4086
final avg		**0.4822**	**0.4965**	**0.6447**	**0.4468**	**0.6439**	**0.4479**	**0.4968**	**0.4047**
	std	0.0066	0.0078	0.0100	0.0080	0.0102	0.0081	0.0079	0.0182
BUGS									
BERT + ML - epoch 4									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5251	0.5341	0.6414	0.4956	0.6410	0.4976	0.5349	0.4812
	2	0.4999	0.5142	0.6523	0.4735	0.6510	0.4755	0.5152	0.4798
	3	0.4997	0.5177	0.6787	0.4698	0.6771	0.4716	0.5184	0.4361
	avg	0.5082	0.5220	0.6575	0.4796	0.6564	0.4815	0.5228	0.4657
2	1	0.5072	0.5094	0.6209	0.4716	0.6187	0.4733	0.5098	0.4456
	2	0.4991	0.5070	0.6857	0.4556	0.6826	0.4575	0.5078	0.4449
	3	0.5055	0.5165	0.6357	0.4798	0.6341	0.4817	0.5170	0.4926
	avg	0.5039	0.5110	0.6474	0.4690	0.6451	0.4708	0.5115	0.4610
final avg		**0.5061**	**0.5165**	**0.6525**	**0.4743**	**0.6507**	**0.4762**	**0.5172**	**0.4634**
	std	0.0099	0.0095	0.0253	0.0131	0.0249	0.0132	0.0096	0.0238

BUGS									
MATCH									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4042	0.4126	0.5244	0.3685	0.5232	0.3695	0.4127	0.5897
	2	0.3663	0.3631	0.5109	0.3200	0.5101	0.3209	0.3634	0.5712
	3	0.4059	0.4132	0.5455	0.3647	0.5452	0.3666	0.4142	0.6257
	avg	0.3922	0.3963	0.5269	0.3511	0.5262	0.3523	0.3968	0.5955
2	1	0.3945	0.4001	0.5247	0.3531	0.5230	0.3545	0.4003	0.6171
	2	0.2991	0.2848	0.4468	0.2461	0.4458	0.2474	0.2852	0.4181
	3	0.3043	0.3069	0.4484	0.2725	0.4471	0.2734	0.3069	0.4786
	avg	0.3326	0.3306	0.4733	0.2906	0.4720	0.2918	0.3308	0.5046
final avg		**0.3624**	**0.3635**	**0.5001**	**0.3208**	0.4991	**0.3220**	**0.3638**	0.5501
	std	0.0491	0.0559	0.0422	0.0513	0.0423	0.0514	0.0560	0.0834

BUGS									
HiAGM									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4388	0.5233	0.5111	0.5482	0.5067	0.5513	0.5223	0.8404
	2	0.4510	0.5209	0.5375	0.5295	0.5349	0.5335	0.5216	0.7077
	3	0.4474	0.5224	0.5224	0.5472	0.5180	0.5511	0.5217	0.7871
	avg	0.4457	0.5222	0.5237	0.5416	0.5199	0.5453	0.5219	0.7784
2	1	0.4427	0.5293	0.5450	0.5475	0.5417	0.5511	0.5291	0.7529
	2	0.4510	0.5169	0.5233	0.5364	0.5152	0.5418	0.5156	0.8034
	3	0.4585	0.5178	0.5481	0.5260	0.5433	0.5316	0.5182	0.7115
	avg	0.4507	0.5213	0.5388	0.5366	0.5334	0.5415	0.5209	0.7559
final avg		**0.4482**	**0.5218**	**0.5313**	**0.5391**	0.5266	**0.5434**	**0.5214**	0.7672
	std	0.0069	0.0045	0.0146	0.0099	0.0153	0.0092	0.0046	0.0527

BUGS									
HBGL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5766	0.5743	0.6032	0.5587	0.6016	0.5597	0.5740	0.6357
	2	0.5762	0.5705	0.6015	0.5532	0.6012	0.5573	0.5707	0.6320
	3	0.5760	0.5678	0.6003	0.5502	0.5992	0.5519	0.5682	0.6283
final avg		**0.5763**	**0.5709**	**0.6017**	**0.5540**	**0.6007**	**0.5563**	**0.5710**	**0.6320**
	std	0.0003	0.0032	0.0015	0.0043	0.0013	0.0040	0.0029	0.0037

BUGS									
GACaps-HTC									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4861	0.5323	0.5804	0.5280	0.5786	0.5300	0.5324	0.5724
	2	0.4923	0.5425	0.5721	0.5541	0.5631	0.5571	0.5471	0.6531
	3	0.4965	0.5543	0.5532	0.5727	0.5510	0.5746	0.5540	0.7172
final avg		**0.4916**	**0.5430**	**0.5686**	**0.5516**	**0.5642**	**0.5539**	**0.5445**	**0.6476**
	std	0.0052	0.0110	0.0139	0.0225	0.0138	0.0225	0.0110	0.0725

BUGS									
SVM - MultiLABEL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.3073	0.4200	0.5825	0.3462	0.5807	0.3496	0.4212	0.5374
	2	0.2996	0.4172	0.5976	0.3358	0.5961	0.3400	0.4193	0.5220
	3	0.3004	0.4148	0.5850	0.3359	0.5842	0.3401	0.4171	0.5158
	avg	0.3024	0.4173	0.5884	0.3393	0.5870	0.3432	0.4192	0.5251
2	1	0.3031	0.4169	0.5829	0.3399	0.5815	0.3440	0.4188	0.5347
	2	0.2999	0.4164	0.5959	0.3375	0.5949	0.3417	0.4187	0.5197
	3	0.3069	0.4272	0.6005	0.3485	0.5988	0.3527	0.4291	0.5229
	avg	0.3033	0.4202	0.5931	0.3420	0.5917	0.3461	0.4222	0.5258
final avg		**0.3029**	**0.4187**	**0.5907**	**0.3406**	0.5894	**0.3447**	**0.4207**	0.5254
	std	0.0035	0.0044	0.0081	0.0055	0.0081	0.0053	0.0043	0.0086

BUGS					
SVM - Multiclass					
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**
1	1	0.5564	0.4763	0.5559	0.4630
	2	0.5511	0.4724	0.5691	0.4592
	3	0.5525	0.4743	0.5473	0.4605
	avg	0.5534	0.4743	0.5574	0.4609
2	1	0.5466	0.4636	0.5345	0.4511
	2	0.5547	0.4823	0.5612	0.4665
	3	0.5474	0.4714	0.5439	0.4546
	avg	0.5496	0.4724	0.5465	0.4574
final avg		**0.5515**	**0.4734**	**0.5520**	**0.4592**
	std	0.0039	0.0061	0.0125	0.0056

BUGS									
XML-CNN									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.2852	0.2780	0.5104	0.2226	0.5099	0.2231	0.2783	0.2966
	2	0.2908	0.2840	0.4951	0.2295	0.4946	0.2304	0.2845	0.3283
	3	0.2897	0.2796	0.5089	0.2249	0.5085	0.2260	0.2804	0.3172
	avg	0.2886	0.2805	0.5048	0.2257	0.5044	0.2265	0.2810	0.3140
2	1	0.2871	0.2824	0.5237	0.2294	0.5230	0.2300	0.2826	0.3135
	2	0.2603	0.2436	0.4782	0.1932	0.4782	0.1945	0.2448	0.2720
	3	0.2669	0.2587	0.4978	0.2025	0.4977	0.2032	0.2592	0.2721
	avg	0.2714	0.2616	0.4999	0.2084	0.4996	0.2092	0.2622	0.2859
final avg		**0.2800**	**0.2710**	**0.5024**	**0.2170**	0.5020	**0.2179**	**0.2716**	0.3000
	std	0.0130	0.0163	0.0156	0.0154	0.0154	0.0152	0.0160	0.0239

BUGS									
XML-CNN + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.2995	0.2854	0.4859	0.2315	0.4854	0.2327	0.2861	0.3136
	2	0.2740	0.2707	0.5131	0.2145	0.5129	0.2154	0.2713	0.2851
	3	0.2689	0.2787	0.4833	0.2215	0.4822	0.2227	0.2794	0.2641
	avg	0.2808	0.2783	0.4941	0.2225	0.4935	0.2236	0.2789	0.2876
2	1	0.2804	0.2715	0.5040	0.2169	0.5035	0.2175	0.2717	0.2760
	2	0.2809	0.2804	0.5117	0.2256	0.5113	0.2264	0.2807	0.2848
	3	0.2898	0.2736	0.5070	0.2197	0.5162	0.2213	0.2752	0.3075
	avg	0.2837	0.2751	0.5076	0.2208	0.5103	0.2217	0.2759	0.2894
final avg		**0.2823**	**0.2767**	**0.5008**	**0.2216**	0.5019	**0.2227**	**0.2774**	0.2885
	std	0.0110	0.0058	0.0130	0.0062	0.0147	0.0063	0.0057	0.0188

BUGS									
XML-CNN + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.2598	0.2427	0.4771	0.1892	0.4772	0.1900	0.2435	0.2266
	2	0.2600	0.2548	0.4896	0.2018	0.4888	0.2025	0.2554	0.2323
	3	0.2577	0.2489	0.5142	0.1982	0.5141	0.1992	0.2497	0.2646
	avg	0.2591	0.2488	0.4936	0.1964	0.4934	0.1972	0.2495	0.2412
2	1	0.2640	0.2607	0.5127	0.2042	0.5122	0.2050	0.2613	0.2516
	2	0.2554	0.2483	0.4644	0.1955	0.4647	0.1967	0.2495	0.2496
	3	0.2650	0.2538	0.5287	0.1985	0.5287	0.1989	0.2541	0.2412
	avg	0.2615	0.2543	0.5020	0.1994	0.5019	0.2002	0.2550	0.2475
final avg		**0.2603**	**0.2515**	**0.4978**	**0.1979**	0.4976	**0.1987**	**0.2522**	0.2443
	std	0.0037	0.0062	0.0247	0.0052	0.0246	0.0052	0.0061	0.0139

Web of Science dataset

WOS									
BERT									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.7758	0.7960	0.8371	0.7740	0.8368	0.7743	0.7960	0.2115
	2	0.7606	0.7877	0.8409	0.7554	0.8407	0.7556	0.7877	0.2027
	3	0.7658	0.7975	0.8390	0.7729	0.8387	0.7731	0.7975	0.2153
	avg	0.7674	0.7937	0.8390	0.7674	0.8387	0.7677	0.7937	0.2098
2	1	0.7732	0.8022	0.8351	0.7807	0.8348	0.7812	0.8023	0.2127
	2	0.7787	0.8025	0.8362	0.7807	0.8360	0.7809	0.8025	0.2125
	3	0.7730	0.8028	0.8450	0.7761	0.8447	0.7765	0.8029	0.1978
	avg	0.7750	0.8025	0.8388	0.7792	0.8385	0.7795	0.8025	0.2076
final avg		**0.7712**	**0.7981**	**0.8389**	**0.7733**	0.8386	**0.7736**	**0.7981**	0.2087
	std	0.0067	0.0059	0.0036	0.0094	0.0037	0.0094	0.0059	0.0069

WOS									
BERT + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.7726	0.8046	0.8468	0.7753	0.8466	0.7756	0.8046	0.1943
	2	0.7621	0.7899	0.8433	0.7575	0.8431	0.7578	0.7899	0.1935
	3	0.7661	0.7926	0.8479	0.7633	0.8477	0.7635	0.7926	0.1957
	avg	0.7669	0.7957	0.8460	0.7654	0.8458	0.7656	0.7957	0.1945
2	1	0.7549	0.7871	0.8543	0.7446	0.8541	0.7450	0.7871	0.1739
	2	0.7623	0.7874	0.8378	0.7603	0.8376	0.7605	0.7874	0.2150
	3	0.7703	0.7955	0.8426	0.7696	0.8423	0.7700	0.7956	0.1923
	avg	0.7625	0.7900	0.8449	0.7582	0.8447	0.7585	0.7900	0.1937
final avg		**0.7647**	**0.7928**	**0.8455**	**0.7618**	0.8452	**0.7621**	**0.7929**	0.1941
	std	0.0064	0.0066	0.0056	0.0106	0.0056	0.0106	0.0066	0.0130

WOS									
BERT + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.7749	0.8022	0.8451	0.7792	0.8449	0.7796	0.8023	0.2083
	2	0.7678	0.7966	0.8362	0.7717	0.8360	0.7719	0.7967	0.2171
	3	0.7722	0.7937	0.8336	0.7684	0.8334	0.7686	0.7936	0.2070
	avg	0.7717	0.7975	0.8383	0.7731	0.8381	0.7734	0.7975	0.2108
2	1	0.7715	0.7991	0.8332	0.7767	0.8330	0.7771	0.7993	0.2045
	2	0.7742	0.7974	0.8273	0.7824	0.8271	0.7827	0.7974	0.2309
	3	0.7700	0.7952	0.8422	0.7683	0.8420	0.7687	0.7953	0.2041
	avg	0.7719	0.7972	0.8343	0.7758	0.8340	0.7762	0.7973	0.2131
final avg		**0.7718**	**0.7974**	**0.8363**	**0.7745**	0.8361	**0.7748**	**0.7974**	0.2120
	std	0.0027	0.0030	0.0065	0.0059	0.0065	0.0059	0.0030	0.0104

WOS									
MATCH									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6043	0.6776	0.7164	0.6535	0.7159	0.6543	0.6778	0.3949
	2	0.5639	0.6396	0.6798	0.6109	0.6792	0.6118	0.6398	0.4090
	3	0.5925	0.6666	0.7095	0.6361	0.7091	0.6368	0.6667	0.3832
	avg	0.5869	0.6613	0.7019	0.6335	0.7014	0.6343	0.6614	0.3957
2	1	0.5888	0.6663	0.7149	0.6368	0.7145	0.6376	0.6664	0.3824
	2	0.6084	0.6772	0.7253	0.6473	0.7249	0.6481	0.6774	0.3876
	3	0.6010	0.6761	0.7208	0.6482	0.7204	0.6491	0.6763	0.3774
	avg	0.5994	0.6732	0.7204	0.6441	0.7199	0.6449	0.6734	0.3825
final avg		**0.5932**	**0.6672**	**0.7111**	**0.6388**	0.7107	**0.6396**	**0.6674**	0.3891
	std	0.0161	0.0145	0.0163	0.0153	0.0163	0.0153	0.0145	0.0114

WOS									
HiAGM									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6341	0.7474	0.6886	0.8375	0.6868	0.8387	0.7470	0.4830
	2	0.6577	0.7602	0.7335	0.8046	0.7322	0.8056	0.7600	0.3746
	3	0.6516	0.7583	0.7154	0.8249	0.7140	0.8259	0.7581	0.4302
	avg	0.6478	0.7553	0.7125	0.8223	0.7110	0.8234	0.7550	0.4293
2	1	0.6380	0.7494	0.6993	0.8236	0.6977	0.8246	0.7490	0.4299
	2	0.6479	0.7491	0.6914	0.8322	0.6897	0.8333	0.7487	0.4367
	3	0.6784	0.7621	0.7223	0.8224	0.7210	0.8236	0.7620	0.4086
	avg	0.6548	0.7535	0.7043	0.8261	0.7028	0.8272	0.7533	0.4250
final avg		**0.6513**	**0.7544**	**0.7084**	**0.8242**	0.7069	**0.8253**	**0.7541**	0.4272
	std	0.0159	0.0065	0.0181	0.0112	0.0183	0.0113	0.0066	0.0356

WOS									
HBGL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8016	0.8217	0.8448	0.8031	0.8447	0.8032	0.8217	0.2165
	2	0.8014	0.8220	0.8452	0.8037	0.8452	0.8037	0.8220	0.2137
	3	0.8013	0.8225	0.8465	0.8040	0.8464	0.8041	0.8225	0.2117
final avg		**0.8014**	**0.8221**	**0.8455**	**0.8036**	**0.8455**	**0.8037**	**0.8221**	**0.2140**
	std	0.0002	0.0004	0.0009	0.0004	0.0009	0.0005	0.0004	0.0024

WOS									
GACaps-HTC									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.7457	0.8114	0.8111	0.8175	0.8106	0.8181	0.8114	0.2632
	2	0.7369	0.8059	0.7959	0.8195	0.7983	0.8243	0.8073	0.2725
	3	0.7301	0.8014	0.7850	0.8264	0.7843	0.8270	0.8014	0.2901
final avg		**0.7376**	**0.8063**	**0.7973**	**0.8211**	**0.7977**	**0.8231**	**0.8067**	**0.2753**
	std	0.0078	0.0050	0.0131	0.0046	0.0132	0.0046	0.0050	0.0136

WOS									
SVM - MultiLABEL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5057	0.6624	0.8206	0.5683	0.8203	0.5701	0.6632	0.2102
	2	0.5029	0.6608	0.8236	0.5676	0.8234	0.5694	0.6617	0.2151
	3	0.5096	0.6640	0.8143	0.5749	0.8139	0.5766	0.6648	0.2211
	avg	0.5061	0.6624	0.8195	0.5703	0.8192	0.5721	0.6632	0.2155
2	1	0.5023	0.6597	0.8165	0.5676	0.8162	0.5695	0.6606	0.2219
	2	0.5054	0.6607	0.8163	0.5694	0.8160	0.5712	0.6615	0.2245
	3	0.5046	0.6592	0.8231	0.5666	0.8227	0.5681	0.6598	0.2107
	avg	0.5041	0.6599	0.8187	0.5679	0.8183	0.5696	0.6606	0.2190
final avg		**0.5051**	**0.6611**	**0.8191**	**0.5691**	0.8188	**0.5708**	**0.6619**	0.2172
	std	0.0026	0.0018	0.0039	0.0030	0.0039	0.0030	0.0018	0.0061

WOS					
SVM - Multiclass					
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**
1	1	0.7617	0.7436	0.7565	0.7405
	2	0.7556	0.7325	0.7453	0.7323
	3	0.7629	0.7408	0.7471	0.7399
	avg	0.7601	0.7390	0.7496	0.7376
2	1	0.7627	0.7440	0.7530	0.7422
	2	0.7583	0.7364	0.7475	0.7347
	3	0.7643	0.7407	0.7540	0.7382
	avg	0.7618	0.7403	0.7515	0.7384
final avg		**0.7609**	**0.7397**	**0.7506**	**0.7380**
	std	0.0033	0.0045	0.0045	0.0038

WOS									
XML-CNN									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4590	0.5538	0.7918	0.4585	0.7917	0.4591	0.5541	0.1895
	2	0.4851	0.5741	0.7808	0.4865	0.7806	0.4871	0.5744	0.2118
	3	0.4646	0.5601	0.7858	0.4694	0.7856	0.4700	0.5604	0.2033
	avg	0.4695	0.5627	0.7861	0.4715	0.7860	0.4721	0.5630	0.2015
2	1	0.4844	0.5724	0.7714	0.4828	0.7712	0.4834	0.5726	0.2057
	2	0.4613	0.5560	0.7768	0.4592	0.7766	0.4599	0.5563	0.1907
	3	0.4742	0.5603	0.7790	0.4707	0.7788	0.4713	0.5605	0.1858
	avg	0.4733	0.5629	0.7757	0.4709	0.7755	0.4715	0.5631	0.1941
final avg		**0.4714**	**0.5628**	**0.7809**	**0.4712**	0.7807	**0.4718**	**0.5630**	0.1978
	std	0.0115	0.0085	0.0071	0.0116	0.0072	0.0116	0.0085	0.0105

WOS									
XML-CNN + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4665	0.5645	0.7805	0.4715	0.7804	0.4722	0.5648	0.2015
	2	0.4752	0.5691	0.7821	0.4783	0.7819	0.4790	0.5694	0.2046
	3	0.4639	0.5657	0.7807	0.4730	0.7805	0.4738	0.5661	0.2026
	avg	0.4685	0.5664	0.7811	0.4743	0.7810	0.4750	0.5668	0.2029
2	1	0.4796	0.5645	0.7716	0.4754	0.7715	0.4759	0.5647	0.1927
	2	0.4758	0.5695	0.7761	0.4778	0.7758	0.4784	0.5697	0.2222
	3	0.4944	0.5795	0.7771	0.4936	0.7769	0.4944	0.5798	0.2028
	avg	0.4833	0.5711	0.7749	0.4823	0.7748	0.4829	0.5714	0.2059
final avg		**0.4759**	**0.5688**	**0.7780**	**0.4783**	0.7779	**0.4790**	**0.5691**	0.2044
	std	0.0109	0.0057	0.0039	0.0080	0.0039	0.0080	0.0057	0.0097

WOS									
XML-CNN + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4635	0.5586	0.7712	0.4699	0.7712	0.4706	0.5590	0.2014
	2	0.4514	0.5427	0.7768	0.4485	0.7767	0.4491	0.5430	0.1858
	3	0.4437	0.5382	0.7847	0.4436	0.7846	0.4442	0.5385	0.1780
	avg	0.4529	0.5465	0.7776	0.4540	0.7775	0.4546	0.5468	0.1884
2	1	0.4454	0.5355	0.7861	0.4430	0.7859	0.4436	0.5357	0.1876
	2	0.4427	0.5379	0.7980	0.4386	0.7979	0.4391	0.5382	0.1754
	3	0.4460	0.5316	0.7760	0.4392	0.7759	0.4398	0.5319	0.1836
	avg	0.4447	0.5350	0.7867	0.4402	0.7866	0.4408	0.5353	0.1822
final avg		**0.4488**	**0.5408**	**0.7821**	**0.4471**	0.7820	**0.4477**	**0.5410**	0.1853
	std	0.0078	0.0095	0.0096	0.0117	0.0096	0.0118	0.0095	0.0091

RCV1-v2 dataset

RCV1									
BERT									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6417	0.6506	0.7980	0.5876	0.7973	0.5886	0.6508	0.1620
2	1	0.6366	0.6938	0.7772	0.6528	0.7761	0.6539	0.6939	0.2298
final avg		**0.6391**	**0.6722**	**0.7876**	**0.6202**	0.7867	**0.6212**	**0.6723**	0.1959
	std	0.0036	0.0305	0.0147	0.0461	0.0150	0.0462	0.0304	0.0480

RCV1									
BERT + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6412	0.6702	0.7822	0.6268	0.7815	0.6275	0.6702	0.2144
2	1	0.6406	0.6523	0.7975	0.5878	0.7969	0.5886	0.6525	0.1714
final avg		**0.6409**	**0.6612**	**0.7899**	**0.6073**	0.7892	**0.6081**	**0.6613**	0.1929
	std	0.0004	0.0127	0.0108	0.0276	0.0109	0.0274	0.0125	0.0304

RCV1									
BERT + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6341	0.6856	0.7774	0.6467	0.7763	0.6479	0.6858	0.2324
2	1	0.6425	0.6753	0.7678	0.6384	0.7669	0.6394	0.6754	0.2216
final avg		**0.6383**	**0.6805**	**0.7726**	**0.6425**	0.7716	**0.6436**	**0.6806**	0.2270
	std	0.0059	0.0073	0.0068	0.0058	0.0066	0.0060	0.0073	0.0076

RCV1									
MATCH									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5155	0.5347	0.6469	0.4915	0.6458	0.4928	0.5349	0.3481
2	1	0.5142	0.5268	0.6254	0.4864	0.6243	0.4875	0.5269	0.3475
final avg		**0.5149**	**0.5308**	**0.6361**	**0.4890**	0.6351	**0.4902**	**0.5309**	0.3478
	std	0.0010	0.0056	0.0152	0.0036	0.0152	0.0037	0.0056	0.0005

RCV1									
HiAGM									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5989	0.6889	0.6836	0.7304	0.6765	0.7352	0.6888	0.3235
2	1	0.6080	0.6783	0.6682	0.7180	0.6647	0.7196	0.6772	0.3237
final avg		**0.6035**	**0.6836**	**0.6759**	**0.7242**	0.6706	**0.7274**	**0.6830**	0.3236
	std	0.0064	0.0075	0.0109	0.0088	0.0084	0.0111	0.0083	0.0001

RCV1									
HBGL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6600	0.7283	0.7637	0.7069	0.7633	0.7072	0.7283	0.2432
	2	0.6591	0.7283	0.7725	0.7055	0.7723	0.7013	0.7285	0.2325
	3	0.6572	0.7285	0.7798	0.6972	0.7790	0.6981	0.7286	0.2178
final avg		**0.6588**	**0.7284**	**0.7720**	**0.7032**	**0.7715**	**0.7022**	**0.7285**	**0.2312**
	std	0.0015	0.0001	0.0081	0.0052	0.0079	0.0046	0.0002	0.0127

RCV1									
GACaps-HTC									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.6378	0.7153	0.7174	0.7259	0.7154	0.7273	0.7150	0.2746
	2	0.6392	0.7121	0.7253	0.7123	0.7291	0.7111	0.7084	0.2630
	3	0.6401	0.7036	0.7372	0.6889	0.7359	0.6900	0.7036	0.2510
final avg		**0.6390**	**0.7103**	**0.7266**	**0.7090**	**0.7268**	**0.7095**	**0.7090**	**0.2629**
	std	0.0012	0.0060	0.0100	0.0187	0.0104	0.0187	0.0057	0.0118

RCV1									
SVM - MultiLABEL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1		**0.4971**	**0.5456**	**0.7599**	**0.4605**	0.7577	**0.4647**	**0.5472**	0.2340

RCV1					
SVM - Multiclass					
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**
1		**0.7289**	**0.4416**	**0.5069**	**0.4222**

RCV1									
XML-CNN									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5532	0.5011	0.6904	0.4381	0.6893	0.4402	0.5022	0.1884
2	1	0.5499	0.4836	0.7143	0.4218	0.7132	0.4236	0.4842	0.1745
final avg		**0.5516**	**0.4923**	**0.7024**	**0.4300**	0.7012	**0.4319**	**0.4932**	0.1815
	std	0.0023	0.0124	0.0169	0.0115	0.0169	0.0117	0.0127	0.0098

RCV1									
XML-CNN + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5523	0.4946	0.6961	0.4329	0.6948	0.4349	0.4956	0.1807
2	1	0.5396	0.4718	0.6862	0.4205	0.6848	0.4220	0.4724	0.2027
final avg		**0.5460**	**0.4832**	**0.6911**	**0.4267**	0.6898	**0.4285**	**0.4840**	0.1917
	std	0.0089	0.0161	0.0070	0.0088	0.0070	0.0091	0.0164	0.0156

RCV1									
XML-CNN + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5379	0.4857	0.6950	0.4231	0.6935	0.4250	0.4866	0.1854
2	1	0.5520	0.4939	0.7002	0.4301	0.6995	0.4314	0.4944	0.1683
final avg		**0.5449**	**0.4898**	**0.6976**	**0.4266**	0.6965	**0.4282**	**0.4905**	0.1768
	std	0.0099	0.0058	0.0037	0.0049	0.0042	0.0045	0.0055	0.0121

Blurb Genre Collection dataset

BGC									
BERT									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4688	0.6184	0.6849	0.5912	0.6834	0.5925	0.6184	0.3726
2	1	0.4734	0.6016	0.6930	0.5635	0.6917	0.5653	0.6020	0.3254
final avg		**0.4711**	**0.6100**	**0.6889**	**0.5773**	0.6875	**0.5789**	**0.6102**	0.3490
	std	0.0032	0.0119	0.0058	0.0196	0.0059	0.0192	0.0116	0.0334

BGC									
BERT + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4668	0.5948	0.6897	0.5513	0.6888	0.5525	0.5951	0.3315
2	1	0.4679	0.6168	0.6865	0.5886	0.6857	0.5897	0.6170	0.3291
final avg		**0.4674**	**0.6058**	**0.6881**	**0.5699**	0.6873	**0.5711**	**0.6060**	0.3303
	std	0.0008	0.0156	0.0023	0.0264	0.0022	0.0263	0.0155	0.0017

BGC									
BERT + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4737	0.6082	0.6979	0.5679	0.6966	0.5692	0.6083	0.3438
2	1	0.4743	0.6150	0.6916	0.5786	0.6905	0.5801	0.6154	0.3462
final avg		**0.4740**	**0.6116**	**0.6948**	**0.5732**	0.6936	**0.5747**	**0.6119**	0.3450
	std	0.0004	0.0049	0.0045	0.0076	0.0043	0.0077	0.0050	0.0017

BGC									
MATCH									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.3882	0.4790	0.5816	0.4345	0.5807	0.4360	0.4793	0.4751
2	1	0.3870	0.4811	0.5646	0.4424	0.5634	0.4434	0.4810	0.4836
final avg		**0.3876**	**0.4800**	**0.5731**	**0.4385**	0.5720	**0.4397**	**0.4802**	0.4793
	std	0.0009	0.0015	0.0120	0.0056	0.0122	0.0053	0.0012	0.0059

BGC									
HiAGM									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4151	0.5462	0.5638	0.5460	0.5611	0.5499	0.5467	0.5381
2	1	0.4073	0.5504	0.5566	0.5601	0.5533	0.5628	0.5500	0.5585
final avg		**0.4112**	**0.5483**	**0.5602**	**0.5530**	0.5572	**0.5564**	**0.5484**	0.5483
	std	0.0055	0.0030	0.0051	0.0100	0.0055	0.0091	0.0024	0.0144

BGC									
HBGL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.5055	0.6766	0.7083	0.6567	0.7071	0.6574	0.6763	0.3516
	2	0.5058	0.6783	0.7074	0.6608	0.7062	0.6621	0.6781	0.3529
	3	0.5066	0.6797	0.7059	0.6638	0.7044	0.6645	0.6793	0.3576
final avg		**0.5060**	**0.6782**	**0.7072**	**0.6604**	**0.7059**	**0.6614**	**0.6779**	**0.3540**
	std	0.0006	0.0016	0.0012	0.0036	0.0013	0.0036	0.0015	0.0032

BGC									
GACaps-HTC									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.4601	0.6522	0.6423	0.6744	0.6400	0.6759	0.6518	0.4345
	2	0.4521	0.6451	0.6302	0.6800	0.6243	0.6786	0.6421	0.4581
	3	0.4459	0.6405	0.6147	0.6854	0.6124	0.6865	0.6398	0.4624
final avg		**0.4527**	**0.6459**	**0.6291**	**0.6799**	**0.6256**	**0.6803**	**0.6446**	**0.4517**
	std	0.0071	0.0059	0.0138	0.0055	0.0138	0.0055	0.0063	0.0150

BGC									
SVM - MultiLABEL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
final avg		**0.3495**	**0.5129**	**0.7081**	**0.4300**	0.7048	**0.4358**	**0.5148**	0.3985

BGC					
SVM - Multiclass					
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**
final avg		**0.6285**	**0.2792**	**0.3498**	**0.2517**

BGC									
XML-CNN									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.3582	0.3931	0.5844	0.3274	0.5838	0.3285	0.3936	0.3148
2	1	0.3622	0.4061	0.5994	0.3403	0.5983	0.3414	0.4064	0.3045
final avg		**0.3602**	**0.3996**	**0.5919**	**0.3339**	0.5911	**0.3350**	**0.4000**	0.3096
	std	0.0028	0.0092	0.0106	0.0091	0.0103	0.0091	0.0091	0.0073

BGC									
XML-CNN + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.3562	0.3744	0.6068	0.3061	0.6066	0.3069	0.3749	0.2714
2	1	0.3536	0.3851	0.6131	0.3211	0.6124	0.3225	0.3858	0.3036
final avg		**0.3549**	**0.3798**	**0.6100**	**0.3136**	0.6095	**0.3147**	**0.3803**	0.2875
	std	0.0018	0.0076	0.0044	0.0106	0.0042	0.0110	0.0077	0.0227

BGC									
XML-CNN + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.3535	0.3849	0.6136	0.3212	0.6133	0.3220	0.3854	0.2969
2	1	0.3552	0.3891	0.6249	0.3242	0.6244	0.3250	0.3893	0.3140

final avg		**0.3544**	**0.3870**	**0.6193**	**0.3227**	0.6188	**0.3235**	**0.3873**	0.3054
	std	0.0012	0.0029	0.0079	0.0022	0.0078	0.0021	0.0028	0.0121

Amazon 5x5 dataset

AMAZON									
BERT									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8965	0.9217	0.9276	0.9168	0.9275	0.9169	0.9217	0.0671
	2	0.8982	0.9227	0.9291	0.9171	0.9290	0.9172	0.9227	0.0665
	3	0.8939	0.9201	0.9254	0.9161	0.9253	0.9162	0.9201	0.0660
	avg	0.8962	0.9215	0.9274	0.9167	0.9273	0.9168	0.9215	0.0666
2	1	0.8939	0.9195	0.9279	0.9145	0.9278	0.9146	0.9195	0.0662
	2	0.8934	0.9206	0.9289	0.9132	0.9288	0.9133	0.9206	0.0660
	3	0.8967	0.9228	0.9279	0.9183	0.9278	0.9184	0.9228	0.0648
	avg	0.8947	0.9210	0.9282	0.9154	0.9281	0.9155	0.9210	0.0657
final avg		**0.8954**	**0.9212**	**0.9278**	**0.9160**	0.9277	**0.9161**	**0.9212**	0.0661
	std	0.0020	0.0014	0.0013	0.0019	0.0013	0.0019	0.0014	0.0008

AMAZON									
BERT + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8945	0.9214	0.9311	0.9130	0.9310	0.9130	0.9214	0.0619
	2	0.8916	0.9202	0.9317	0.9098	0.9316	0.9099	0.9202	0.0609
	3	0.8890	0.9174	0.9277	0.9093	0.9276	0.9094	0.9174	0.0634
	avg	0.8917	0.9197	0.9301	0.9107	0.9301	0.9108	0.9197	0.0620
2	1	0.8900	0.9186	0.9313	0.9083	0.9313	0.9083	0.9186	0.0648
	2	0.8903	0.9179	0.9297	0.9090	0.9296	0.9091	0.9179	0.0648
	3	0.8915	0.9195	0.9297	0.9117	0.9296	0.9119	0.9195	0.0635
	avg	0.8906	0.9186	0.9302	0.9097	0.9302	0.9098	0.9187	0.0644
final avg		**0.8912**	**0.9192**	**0.9302**	**0.9102**	0.9301	**0.9103**	**0.9192**	0.0632
	std	0.0019	0.0015	0.0015	0.0018	0.0015	0.0018	0.0015	0.0016

AMAZON									
BERT + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8959	0.9214	0.9271	0.9168	0.9271	0.9169	0.9214	0.0659
	2	0.8999	0.9242	0.9294	0.9199	0.9293	0.9200	0.9242	0.0637
	3	0.8957	0.9206	0.9271	0.9163	0.9270	0.9164	0.9206	0.0652
	avg	0.8972	0.9221	0.9279	0.9177	0.9278	0.9178	0.9221	0.0649
2	1	0.8959	0.9215	0.9278	0.9167	0.9277	0.9168	0.9215	0.0651
	2	0.8935	0.9197	0.9257	0.9151	0.9256	0.9153	0.9197	0.0696

	3	0.8950	0.9209	0.9275	0.9158	0.9274	0.9159	0.9210	0.0652
	avg	0.8948	0.9207	0.9270	0.9159	0.9269	0.9160	0.9207	0.0666
final avg		**0.8960**	**0.9214**	**0.9274**	**0.9168**	0.9273	**0.9169**	**0.9214**	0.0658
	std	0.0021	0.0015	0.0012	0.0017	0.0012	0.0016	0.0015	0.0020

AMAZON									
MATCH									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8754	0.9049	0.9125	0.8978	0.9123	0.8979	0.9049	0.0807
	2	0.8541	0.8982	0.9094	0.8889	0.9091	0.8891	0.8981	0.0825
	3	0.8758	0.9054	0.9132	0.8980	0.9131	0.8981	0.9054	0.0785
	avg	0.8684	0.9028	0.9117	0.8949	0.9115	0.8950	0.9028	0.0806
2	1	0.8744	0.9048	0.9130	0.8971	0.9128	0.8972	0.9048	0.0798
	2	0.8755	0.9049	0.9124	0.8978	0.9123	0.8979	0.9049	0.0801
	3	0.8752	0.9052	0.9133	0.8975	0.9132	0.8976	0.9052	0.0787
	avg	0.8750	0.9050	0.9129	0.8975	0.9128	0.8976	0.9049	0.0795
final avg		**0.8717**	**0.9039**	**0.9123**	**0.8962**	0.9121	**0.8963**	**0.9039**	0.0801
	std	0.0087	0.0028	0.0015	0.0036	0.0015	0.0035	0.0028	0.0015

AMAZON									
HiAGM									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8713	0.9026	0.9064	0.8997	0.9062	0.9000	0.9026	0.0832
	2	0.8711	0.9012	0.9056	0.8976	0.9054	0.8979	0.9012	0.0866
	3	0.8786	0.9075	0.9126	0.9034	0.9124	0.9035	0.9074	0.0806
	avg	0.8736	0.9038	0.9082	0.9002	0.9080	0.9004	0.9037	0.0835
2	1	0.8794	0.9056	0.9078	0.9044	0.9075	0.9045	0.9056	0.0915
	2	0.8723	0.9001	0.9061	0.8959	0.9058	0.8960	0.9000	0.0859
	3	0.8686	0.9006	0.9037	0.8981	0.9034	0.8983	0.9005	0.0872
	avg	0.8735	0.9021	0.9058	0.8995	0.9056	0.8996	0.9021	0.0882
final avg		**0.8735**	**0.9029**	**0.9070**	**0.8999**	0.9068	**0.9000**	**0.9029**	0.0858
	std	0.0044	0.0030	0.0031	0.0034	0.0031	0.0033	0.0030	0.0037

AMAZON									
HBGL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8403	0.8804	0.9036	0.8828	0.9035	0.8828	0.8805	0.1134
	2	0.8405	0.8801	0.9021	0.8819	0.8999	0.8825	0.8801	0.1134
	3	0.8406	0.8783	0.8997	0.8816	0.8996	0.8817	0.8783	0.1134
final avg		**0.8405**	**0.8796**	**0.9018**	**0.8821**	**0.9010**	**0.8824**	**0.8796**	**0.1134**
	std	0.0002	0.0012	0.0019	0.0006	0.0022	0.0006	0.0012	0.0000

AMAZON									
GACaps-HTC									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8651	0.9048	0.9189	0.8924	0.9188	0.8925	0.9048	0.0706
	2	0.8677	0.9056	0.9200	0.8933	0.9192	0.8932	0.9053	0.0702
	3	0.8689	0.9069	0.9202	0.8948	0.9201	0.8949	0.9069	0.0695
final avg		**0.8673**	**0.9058**	**0.9197**	**0.8935**	**0.9194**	**0.8936**	**0.9057**	**0.0701**
	std	0.0019	0.0010	0.0007	0.0012	0.0006	0.0012	0.0011	0.0005

AMAZON									
SVM - MultiLABEL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.7321	0.8479	0.9230	0.7907	0.9225	0.7913	0.8480	0.0640
	2	0.7367	0.8509	0.9252	0.7939	0.9248	0.7945	0.8510	0.0616
	3	0.7327	0.8482	0.9247	0.7903	0.9243	0.7910	0.8483	0.0627
	avg	0.7338	0.8490	0.9243	0.7916	0.9239	0.7923	0.8491	0.0628
2	1	0.7337	0.8489	0.9239	0.7918	0.9235	0.7925	0.8491	0.0632
	2	0.7346	0.8491	0.9241	0.7919	0.9236	0.7926	0.8492	0.0636
	3	0.7341	0.8493	0.9255	0.7915	0.9251	0.7921	0.8494	0.0618
	avg	0.7341	0.8491	0.9245	0.7917	0.9241	0.7924	0.8492	0.0628
final avg		**0.7340**	**0.8491**	**0.9244**	**0.7917**	0.9240	**0.7923**	**0.8492**	0.0628
	std	0.0016	0.0010	0.0009	0.0013	0.0009	0.0013	0.0010	0.0010

AMAZON					
SVM - Multiclass					
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**
1	1	0.8664	0.8675	0.8684	0.8672
	2	0.8678	0.8688	0.8697	0.8687
	3	0.8655	0.8668	0.8679	0.8664
	avg	0.8666	0.8677	0.8687	0.8674
2	1	0.8664	0.8676	0.8683	0.8676
	2	0.8665	0.8675	0.8686	0.8672
	3	0.8667	0.8679	0.8689	0.8675
	avg	0.8666	0.8677	0.8686	0.8674
final avg		**0.8666**	**0.8677**	**0.8686**	**0.8674**
	std	0.0007	0.0007	0.0006	0.0008

AMAZON									
XML-CNN									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8276	0.8859	0.9144	0.8610	0.9141	0.8613	0.8859	0.0765
	2	0.8317	0.8870	0.9110	0.8666	0.9107	0.8669	0.8870	0.0808
	3	0.8280	0.8864	0.9140	0.8619	0.9137	0.8623	0.8864	0.0763
	avg	0.8291	0.8864	0.9131	0.8632	0.9128	0.8635	0.8865	0.0778
2	1	0.8258	0.8852	0.9105	0.8635	0.9102	0.8638	0.8852	0.0791
	2	0.8337	0.8898	0.9139	0.8681	0.9136	0.8685	0.8899	0.0763
	3	0.8273	0.8859	0.9144	0.8611	0.9141	0.8614	0.8859	0.0758
	avg	0.8289	0.8870	0.9129	0.8642	0.9127	0.8646	0.8870	0.0770
final avg		**0.8290**	**0.8867**	**0.9130**	**0.8637**	0.9127	**0.8640**	**0.8867**	0.0774
	std	0.0030	0.0016	0.0018	0.0030	0.0018	0.0030	0.0017	0.0020

AMAZON									
XML-CNN + ML									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8250	0.8843	0.9094	0.8627	0.9091	0.8630	0.8844	0.0796
	2	0.8329	0.8887	0.9123	0.8672	0.9120	0.8676	0.8888	0.0780
	3	0.8239	0.8835	0.9115	0.8592	0.9111	0.8595	0.8835	0.0793
	avg	0.8272	0.8855	0.9111	0.8630	0.9108	0.8634	0.8855	0.0790
2	1	0.8315	0.8882	0.9120	0.8671	0.9117	0.8674	0.8882	0.0767
	2	0.8275	0.8854	0.9117	0.8624	0.9114	0.8627	0.8854	0.0780
	3	0.8266	0.8858	0.9113	0.8635	0.9110	0.8638	0.8858	0.0778
	avg	0.8285	0.8865	0.9116	0.8643	0.9114	0.8646	0.8865	0.0775
final avg		**0.8279**	**0.8860**	**0.9113**	**0.8637**	0.9111	**0.8640**	**0.8860**	0.0782
	std	0.0036	0.0021	0.0010	0.0031	0.0010	0.0031	0.0021	0.0011

AMAZON									
XML-CNN + CL									
Run n	**Fold n**	**acc**	**f1**	**prec**	**rec**	**h_prec**	**h_rec**	**h_f1**	**AHC**
1	1	0.8241	0.8851	0.9178	0.8567	0.9177	0.8570	0.8851	0.0704
	2	0.8233	0.8851	0.9156	0.8583	0.9154	0.8586	0.8851	0.0711
	3	0.8278	0.8865	0.9158	0.8607	0.9156	0.8611	0.8866	0.0723
	avg	0.8251	0.8856	0.9164	0.8586	0.9162	0.8589	0.8856	0.0713
2	1	0.8271	0.8860	0.9184	0.8580	0.9181	0.8583	0.8861	0.0708
	2	0.8173	0.8814	0.9197	0.8490	0.9195	0.8493	0.8814	0.0684
	3	0.8254	0.8854	0.9154	0.8589	0.9153	0.8591	0.8854	0.0724
	avg	0.8233	0.8843	0.9178	0.8553	0.9176	0.8556	0.8843	0.0705
final avg		**0.8242**	**0.8849**	**0.9171**	**0.8569**	0.9169	**0.8572**	**0.8850**	0.0709
	std	0.0038	0.0018	0.0018	0.0041	0.0018	0.0041	0.0018	0.0015

Bibliography

[1] Yu Zhang, Zhihong Shen, Yuxiao Dong, Kuansan Wang, and Jiawei Han. 2021. MATCH: Metadata-Aware Text Classification in A Large Hierarchy. In Proceedings of the Web Conference 2021 (WWW '21). Association for Computing Machinery, New York, NY, USA, 3246–3257. https://doi.org/10.1145/3442381.3449979

[2] Jie Zhou, Chunping Ma, Dingkun Long, Guangwei Xu, Ning Ding, Haoyu Zhang, Pengjun Xie, and Gongshen Liu. 2020. Hierarchy-Aware Global Model for Hierarchical Text Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1106–1117, Online. Association for Computational Linguistics.

[3] Ting Jiang, Deqing Wang, Leilei Sun, Zhongzhi Chen, Fuzhen Zhuang, and Qinghong Yang. 2022. Exploiting Global and Local Hierarchies for Hierarchical Text Classification. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 4030–4039, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

[4] Jinhyun Bang, Jonghun Park, and Jonghyuk Park. 2023. GACaps-HTC: graph attention capsule network for hierarchical text classification. Applied Intelligence 53, 17 (Sep 2023), 20577–20594. https://doi.org/10.1007/s10489-023-04585-6

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[6] Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. 2017. Deep Learning for Extreme Multi-label Text Classification. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 115–124. https://doi.org/10.1145/3077136.3080834

[7] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (COLT '92). Association for Computing Machinery, New York, NY, USA, 144–152. https://doi.org/10.1145/130385.130401

[8] Vaswani, A.; Aggarwal, G.; Netrapalli, P.; Hegde, N.G. All Mistakes Are Not Equal: Comprehensive Hierarchy Aware Multi-label 1884 Predictions (CHAMP). arXiv (preprint) 2022. https://doi.org/10.48550/ARXIV.2206.08653.

[9] Zhang, Y.; Shen, Z.; Dong, Y.; Wang, K.; Han, J. MATCH: Metadata-Aware Text Classification in A Large Hierarchy. In 2057 Proceedings of the Proceedings of the Web Conference 2021, New York, NY, USA, 2021; WWW ’21, pp. 3246–3257. https: 2058//doi.org/10.1145/3442381.3449979

[10] Scikit-learn.org [Internet], https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

[11] K. Kowsari, D. E. Brown, M. Heidarysafa, K. Jafari Meimandi, M. S. Gerber, L. E. Barnes, HDLTex: Hierarchical deep learning for text classification, in: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, pp. 364–371. doi:10.1109/ICMLA.2017.0-134.

[12] H. Chen, Q. Ma, Z. Lin, J. Yan, Hierarchy-aware label semantics matching network for hierarchical text classification, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language

Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, Association for Computational Linguistics, 2021, pp. 4370–4379. URL: https://doi.org/10.18653/v1/2021.acl-long.337. doi:10.18653/v1/2021.acl-long.337.

[13] H. Liu, D. Zhang, B. Yin, X. Zhu, Improving pretrained models for zero-shot multi-label text classification through reinforced label hierarchy reasoning, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 1051–1062. URL: https: //aclanthology.org/2021.naacl-main.83. doi:10.18653/v1/2021.naacl-main.83.

[14] Y. Ma, X. Liu, L. Zhao, Y. Liang, P. Zhang, B. Jin, Hybrid embedding-based text representation for hierarchical multi-label text classification, Expert Syst. Appl. 187 (2022) 115905.

[15] J. Liu, W-C. Chang, Y. Wu, and Y. Yang. 2017. Deep Learning for Extreme Multi-Label Text Classification. In Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 115–124.