

Table 1: Common MIL datasets

Name	Instances	Min	Max	Features	Bags	+ bags	- bags
Musk 1 [4] ♣	476	2	40	166	92	47	45
Musk 2 [4] ♣	6598	1	1044	166	102	39	63
Mutagenesis 1 (easy) [6] ♣	10486	28	88	7	188	125	63
Mutagenesis 2 (hard) [6] ♣	2132	26	86	7	42	13	29
Protein [7] ♣	26611	35	189	8	193	25	168
Elephant [1] ♥	1391	2	13	230	200	100	100
Fox [1] ♥	1302	1	13	230	200	100	100
Tiger [1] ♥	1220	2	13	230	200	100	100
Corel, African [3] ♥	7947	2	13	9	2000	100	1900
Corel, Antique [3] ♥	7947	2	13	9	2000	100	1900
Corel, Battleships [3] ♥	7947	2	13	9	2000	100	1900
Corel, Beach [3] ♥	7947	2	13	9	2000	100	1900
Corel, Buses [3] ♥	7947	2	13	9	2000	100	1900
Corel, Cars [3] ♥	7947	2	13	9	2000	100	1900
Corel, Desserts [3] ♥	7947	2	13	9	2000	100	1900
Corel, Dinosaurs [3] ♥	7947	2	13	9	2000	100	1900
Corel, Dogs [3] ♥	7947	2	13	9	2000	100	1900
Corel, Elephants [3] ♥	7947	2	13	9	2000	100	1900
Corel, Fashion [3] ♥	7947	2	13	9	2000	100	1900
Corel, Flowers [3] ♥	7947	2	13	9	2000	100	1900
Corel, Food [3] ♥	7947	2	13	9	2000	100	1900
Corel, Historical [3] ♥	7947	2	13	9	2000	100	1900
Corel, Horses [3] ♥	7947	2	13	9	2000	100	1900
Corel, Lizards [3] ♥	7947	2	13	9	2000	100	1900
Corel, Mountains [3] ♥	7947	2	13	9	2000	100	1900
Corel, Skiing [3] ♥	7947	2	13	9	2000	100	1900
Corel, Sunset [3] ♥	7947	2	13	9	2000	100	1900
Corel, Waterfalls [3] ♥	7947	2	13	9	2000	100	1900
UCSB Breast Cancer [5] ♥	2002	21	40	708	58	26	32
Newsgroups 1, alt.atheism [9] ♠	5443	22	76	200	100	50	50
N.g. 2, comp.graphics [9] ♠	3094	12	58	200	100	50	50
N.g. 3, comp.os.ms-windows.misc [9] ♠	5175	25	82	200	100	50	50
N.g. 4, comp.sys.ibm.pc.hardware [9] ♠	4827	19	74	200	100	50	50
N.g. 5, comp.sys.mac.hardware [9] ♠	4473	17	71	200	100	50	50
N.g. 6, comp.windows.x [9] ♠	3110	12	54	200	100	50	50
N.g. 7, misc.forsale [9] ♠	5306	29	84	200	100	50	50
N.g. 8, rec.autos [9] ♠	3458	15	39	200	100	50	50
N.g. 9, rec.motorcycles [9] ♠	4730	22	73	200	100	50	50
N.g. 10, rec.sport.baseball [9] ♠	3358	15	58	200	100	50	50
N.g. 11, rec.sport.hockey [9] ♠	1982	8	38	200	100	50	50
N.g. 12, sci.crypt [9] ♠	4284	20	71	200	100	50	50
N.g. 13, sci.electronics [9] ♠	3192	12	58	200	100	50	50
N.g. 14, sci.med [9] ♠	3045	11	54	200	100	50	50
N.g. 15, sci.space [9] ♠	3655	16	59	200	100	50	50
N.g. 16, soc.religion.christian [9] ♠	4677	21	71	200	100	50	50
N.g. 17, talk.politics.guns [9] ♠	3558	13	59	200	100	50	50
N.g. 18, talk.politics.mideast [9] ♠	3376	15	55	200	100	50	50
N.g. 19, talk.politics.misc [9] ♠	4788	21	75	200	100	50	50
N.g. 20, talk.religion.misc [9] ♠	4606	25	79	200	100	50	50

MIL application categories: ♣ molecular activity prediction, ♥ image annotation, ♠ text classification, ♦ audio recording classification.

Table 1 continued.

Name	Instances	Min	Max	Features	Bags	+ bags	- bags
Web recommendation 1 [8] ♠	2212	4	131	5863	75	17	58
Web recommendation 2 [8] ♠	2212	5	200	6519	75	18	57
Web recommendation 3 [8] ♠	2212	5	200	6306	75	14	61
Web recommendation 4 [8] ♠	2291	4	200	6059	75	55	20
Web recommendation 5 [8]	2546	5	200	6407	75	61	14
Web recommendation 6 [8] ♠	2462	4	200	6417	75	59	16
Web recommendation 7 [8] ♠	2400	4	200	6450	75	39	36
Web recommendation 8 [8] ♠	2183	4	200	5999	75	35	40
Web recommendation 9 [8] ♠	2321	5	200	6279	75	37	38
Birds, Brown creeper [2] ♦	10232	2	43	38	548	197	351
Birds, Chestnut-backed chickadee [2] ♦	10232	2	43	38	548	117	431
Birds, Dark-eyed junco [2] ♦	10232	2	43	38	548	20	528
Birds, Hammonds flycatcher [2] ♦	10232	2	43	38	548	103	445
Birds, Hermit thrush [2] ♦	10232	2	43	38	548	15	533
Birds, Hermit warbler [2] ♦	10232	2	43	38	548	63	485
Birds, Olive-sided flycatcher [2] ♦	10232	2	43	38	548	90	458
Birds, Pacific slope flycatcher [2] ♦	10232	2	43	38	548	165	383
Birds, Red-breasted nuthatch [2] ♦	10232	2	43	38	548	82	466
Birds, Swainsons thrush [2] ♦	10232	2	43	38	548	79	469
Birds, Varied thrush [2] ♦	10232	2	43	38	548	89	459
Birds, Western tanager [2] ♦	10232	2	43	38	548	46	502
Birds, Winter Wren [2] ♦	10232	2	43	38	548	109	439

MIL application categories: ♣ molecular activity prediction, ♥ image annotation, ♠ text classification, ♦ audio recording classification.

References

- [1] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems 15*, pages 561–568. MIT Press, 2003.
- [2] Forrest Briggs, Balaji Lakshminarayanan, Lawrence Neal, Xiaoli Z Fern, Raviv Raich, Sarah JK Hadley, Adam S Hadley, and Matthew G Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6):4640–4650, 2012.
- [3] Yixin Chen, Jinbo Bi, and James Z Wang. Miles: Multiple-instance learning via embedded instance selection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):1931–1947, 2006.
- [4] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence*, 89(1):31–71, 1997.

- [5] Melih Kandemir, Chong Zhang, and Fred A Hamprecht. Empowering multiple instance histopathology cancer diagnosis by cell graphs. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 228–235. Springer, 2014.
- [6] Ashwin Srinivasan, S Muggleton, and RD King. Comparing the use of background knowledge by inductive logic programming systems. In *Proceedings of the 5th International Workshop on Inductive Logic Programming*, pages 199–230. Department of Computer Science, Katholieke Universiteit Leuven, 1995.
- [7] Qingping Tao, Stephen Scott, NV Vinodchandran, and Thomas Takeo Osugi. Svm-based generalized multiple-instance learning via approximate box counting. In *Proceedings of the twenty-first international conference on Machine learning*, page 101. ACM, 2004.
- [8] Zhi-Hua Zhou, Kai Jiang, and Ming Li. Multi-instance learning based web mining. *Applied Intelligence*, 22(2):135–147, 2005.
- [9] Zhi-Hua Zhou, Yu-Yin Sun, and Yu-Feng Li. Multi-instance learning by treating instances as non-iid samples. In *Proceedings of the 26th annual international conference on machine learning*, pages 1249–1256. ACM, 2009.