1 Supplemental Results

Due to the space limitation, this section provides those experimental results that are not included in the main body of the paper.

1.1 Comparison with Semi-supervised Methods

In this subsection, we present and discuss the comparison results between CDSC-AL and semi-supervised approaches using 5%, 15%, and 20% of labeled data. The best results are in bold-face and the Nemenyi post-hoc test is performed here.

Table 1: Comparison with semi-supervised approaches using 5% labeled data.

Datasets	Metric	LNP	OReSSL	CDSC-AL
Synthetic-1	*BA*	0.8836	0.9214	**0.9323**
	F_{macro}	0.7706	0.9220	**0.9344**
Synthetic-2	*BA*	0.8278	**0.8329**	0.8297
	F_{macro}	0.6078	0.7796	**0.8013**
Sea	*BA*	0.5048	0.8068	**0.9554**
	F_{macro}	0.5879	0.8169	**0.9601**
KDD cup 99	*BA*	0.5125	0.6182	**0.8219**
	F_{macro}	0.5007	0.6205	**0.7829**
Forest covtype	*BA*	0.4838	0.6752	**0.8256**
	F_{macro}	0.4821	0.6755	**0.8066**
Gas Sensor Drift	*BA*	0.5695	0.8474	**0.8866**
	F_{macro}	0.5951	0.8299	**0.8934**
Shuttle	*BA*	0.3635	**0.4475**	0.4459
	F_{macro}	0.3661	**0.4508**	0.4457
MNIST	*BA*	0.7231	0.8322	**0.9307**
	F_{macro}	0.7286	0.8327	**0.9313**
CIFAR-10	*BA*	0.4114	0.6304	**0.7625**
	F_{macro}	0.4075	0.6198	**0.7619**

(a) BA (b) F_{macro}

Figure 1: Comparison of CDSC-AL against semi-supervised methods with the Nemenyi test with $\alpha = 0.05$ using 5% labeled data.

Discussions: Table 1 summarizes the experimental results for CDSC-AL and other two semi-supervised methods when there is 5% labeled data in each incoming data chunk. As shown in Table 1, CDSC-AL provides better performance on most data streams except for Shuttle. In Figure 1, the Nemenyi post-hoc test indicates that CDSC-AL achieves statistically comparable performance than OReSSL method while showing statistically better performance than LNP method using 5% of labeled data.

(a) BA (b) F_{macro}

Figure 2: Comparison of CDSC-AL against semi-supervised methods with the Nemenyi test with $\alpha = 0.05$ using 15% labeled data.

Table 2: Comparison with semi-supervised approaches using 15% labeled data.

Datasets	Metric	LNP	OReSSL	CDSC-AL
Synthetic-1	*BA*	0.8939	0.9307	**0.9495**
	F_{macro}	0.8126	0.9318	**0.9537**
Synthetic-2	*BA*	0.8416	0.8542	**0.8549**
	F_{macro}	0.6598	0.7937	**0.8287**
Sea	*BA*	0.5418	0.8596	**0.9713**
	F_{macro}	0.6271	0.8619	**0.9754**
KDD cup 99	*BA*	0.5528	0.7458	**0.8366**
	F_{macro}	0.5739	0.7535	**0.7923**
Forest covtype	*BA*	0.5420	0.7512	**0.8612**
	F_{macro}	0.5438	0.7485	**0.8442**
Gas Sensor Drift	*BA*	0.6756	0.9202	**0.9291**
	F_{macro}	0.6858	0.9238	**0.9299**
Shuttle	*BA*	0.4259	0.5004	**0.5015**
	F_{macro}	0.4256	0.5193	**0.5205**
MNIST	*BA*	0.7874	0.9451	**0.9772**
	F_{macro}	0.7955	0.9479	**0.9773**
CIFAR-10	*BA*	0.4251	0.6596	**0.8051**
	F_{macro}	0.4264	0.6426	**0.8208**

Discussions: Table 2 presents the experimental results of CDSC-AL and other two semi-supervised methods with 15% labeled data in each incoming data chunk. In Table 2, we can observe that CDSC-AL outperforms the OReSSL and LNP methods on all data streams in terms of BA and F_{macro}. From Figure 2, the CD diagram reveals that CDSC-AL provides statistically better performance than OReSSL and LNP method using 15% of labeled data.

Table 3: Comparison with semi-supervised approaches using 20% labeled data.

Datasets	Metric	LNP	OReSSL	CDSC-AL
Synthetic-1	*BA*	0.9043	0.9333	**0.9634**
	F_{macro}	0.8343	0.9428	**0.9684**
Synthetic-2	*BA*	0.8562	**0.8634**	0.8616
	F_{macro}	0.6862	0.7988	**0.8367**
Sea	*BA*	0.5620	0.8970	**0.9715**
	F_{macro}	0.6497	0.9145	**0.9758**
KDD cup 99	*BA*	0.6088	0.7645	**0.8369**
	F_{macro}	0.6121	0.7812	**0.7927**
Forest covtype	*BA*	0.5692	0.7898	**0.8809**
	F_{macro}	0.5735	0.7834	**0.8668**
Gas Sensor Drift	*BA*	0.6791	0.9210	**0.9293**
	F_{macro}	0.6892	0.9248	**0.9301**
Shuttle	*BA*	0.4436	0.5197	**0.5223**
	F_{macro}	0.4643	0.5361	**0.5455**
MNIST	*BA*	0.8148	0.9612	**0.9816**
	F_{macro}	0.8249	0.9646	**0.9824**
CIFAR-10	*BA*	0.4291	0.6872	**0.8208**
	F_{macro}	0.4381	0.6778	**0.8263**

(a) BA (b) F_{macro}

Figure 3: Comparison of CDSC-AL against semi-supervised methods with the Nemenyi test with $\alpha = 0.05$ using 20% labeled data.

Discussions: Table 3 shows the experimental results of CDSC-AL and two semi-supervised methods with 20% labeled data in each incoming data chunk. From Table 2, it is clear that CDSC-AL achieves better performance than two compared semi-supervised methods on all data streams for

both BA and F_{macro}. As shown in Figure 3, it is observed that CDSC-AL also shows statistically comparable performance than OReSSL method while showing statistically better performance than LNP method using 20% of labeled data.

Summary of Discussions: In summary, the performance of CDSC-AL method shows better improvement than the two compared semi-supervised methods on all data streams as the label proportion increases.

1.2 Comparison with Supervised Methods

In this subsection, we summarize the comparison results between CDSC-AL and supervised approaches using 5%, 15%, and 20% of labeled data in Tables 4, 5, and 6. The best results are in bold-face.

Table 4: Comparison with supervised approaches using 5% labeled data.

Datasets	Metric	LB	OBA	AHT	SAMkNN	CDSC-AL
Syn-1	*BA*	0.7910	0.6640	0.6354	0.6247	**0.9123**
	F_{macro}	0.7965	0.6675	0.6513	0.6313	**0.9044**
Syn-2	*BA*	0.7124	0.7204	0.6926	0.6784	**0.8297**
	F_{macro}	0.7218	0.7219	0.6977	0.6864	**0.8013**
Sea	*BA*	0.8204	0.7498	0.7493	0.7205	**0.9554**
	F_{macro}	0.8227	0.7501	0.7505	0.7345	**0.9601**
KDD 99	*BA*	0.7585	0.7812	**0.8541**	0.7495	0.8219
	F_{macro}	0.7564	0.7798	**0.8012**	0.7682	0.7829
covtype	*BA*	**0.8888**	0.8707	0.8612	0.8545	0.8256
	F_{macro}	**0.8901**	0.8709	0.8688	0.8588	0.8066
GSD	*BA*	0.7185	0.6345	0.6111	0.6357	**0.8866**
	F_{macro}	0.7199	0.6361	0.6188	0.6412	**0.8934**
Shuttle	*BA*	**0.4789**	0.4477	0.4508	0.4424	0.4459
	F_{macro}	**0.5187**	0.5112	0.4987	0.4894	0.4457
MNIST	*BA*	0.8909	0.8498	0.8393	0.8549	**0.9307**
	F_{macro}	0.8946	0.8501	0.8412	0.8596	**0.9319**
CIFAR10	*BA*	0.7199	0.6208	0.7366	0.6218	**0.7625**
	F_{macro}	0.7208	0.6325	0.7381	0.6295	**0.7619**

Table 5: Comparison with supervised approaches using 15% labeled data.

Datasets	Metric	LB	OBA	AHT	SAMkNN	CDSC-AL
Syn-1	*BA*	0.7910	0.6640	0.6354	0.6247	**0.9195**
	F_{macro}	0.7965	0.6675	0.6513	0.6313	**0.9137**
Syn-2	*BA*	0.7124	0.7204	0.6926	0.6784	**0.8549**
	F_{macro}	0.7218	0.7219	0.6977	0.6864	**0.8616**
Sea	*BA*	0.8204	0.7498	0.7493	0.7205	**0.9713**
	F_{macro}	0.8227	0.7501	0.7505	0.7345	**0.9729**
KDD 99	*BA*	0.7585	0.7812	**0.8541**	0.7495	0.8366
	F_{macro}	0.7564	0.7798	**0.8012**	0.7682	0.7923
covtype	*BA*	**0.8888**	0.8707	0.8612	0.8545	0.8616
	F_{macro}	**0.8901**	0.8709	0.8688	0.8588	0.8442
GSD	*BA*	0.7185	0.6345	0.6111	0.6357	**0.9291**
	F_{macro}	0.7199	0.6361	0.6188	0.6412	**0.9299**
Shuttle	*BA*	0.4789	0.4477	0.4508	0.4424	**0.5015**
	F_{macro}	0.5187	0.5112	0.4987	0.4894	**0.5205**
MNIST	*BA*	0.8909	0.8498	0.8393	0.8549	**0.9772**
	F_{macro}	0.8946	0.8501	0.8412	0.8596	**0.9773**
CIFAR10	*BA*	0.7199	0.6208	0.7366	0.6218	**0.8051**
	F_{macro}	0.7208	0.6325	0.7381	0.6295	**0.8083**

Summary of Discussions: As mentioned in the paper, we increased the label proportion for CDSC-AL method up to 20% and presented the results in Tables 4, 5, and 6. As shown in Tables 4, 5, and 6, the performance of CDSC-AL method improves as the proportion of labeled data increases. More importantly, with 20% labeled data, CDSC-AL achieves the best performance on all benchmark data streams except for the KDD cup 99 and Forest covtype. For KDD cup 99 and Forest covtype, CDSC-AL shows comparable performance.

Table 6: Comparison with supervised approaches using 20% labeled data.

Datasets	Metric	LB	OBA	AHT	SAMkNN	CDSC-AL
Syn-1	*BA*	0.7910	0.6640	0.6354	0.6247	**0.9234**
	F_{macro}	0.7965	0.6675	0.6513	0.6313	**0.9184**
Syn-2	*BA*	0.7124	0.7204	0.6926	0.6784	**0.8616**
	F_{macro}	0.7218	0.7219	0.6977	0.6864	**0.8367**
Sea	*BA*	0.8204	0.7498	0.7493	0.7205	**0.9715**
	F_{macro}	0.8227	0.7501	0.7505	0.7345	**0.9758**
KDD 99	*BA*	0.7585	0.7812	**0.8541**	0.7495	0.8469
	F_{macro}	0.7564	0.7798	**0.8012**	0.7682	0.7927
covtype	*BA*	**0.8888**	0.8707	0.8612	0.8545	0.8809
	F_{macro}	**0.8901**	0.8709	0.8688	0.8588	0.8668
GSD	*BA*	0.7185	0.6345	0.6111	0.6357	**0.9293**
	F_{macro}	0.7199	0.6361	0.6188	0.6412	**0.9301**
Shuttle	*BA*	0.4789	0.4477	0.4508	0.4424	**0.5223**
	F_{macro}	0.5187	0.5112	0.4987	0.4894	**0.5445**
MNIST	*BA*	0.8909	0.8498	0.8393	0.8549	**0.9816**
	F_{macro}	0.8946	0.8501	0.8412	0.8596	**0.9824**
CIFAR10	*BA*	0.7199	0.6208	0.7366	0.6218	**0.8208**
	F_{macro}	0.7208	0.6325	0.7381	0.6295	**0.8263**