

TABLE I  
AVERAGE PREQUENTIAL ACCURACY (STANDARD DEVIATION) ACROSS 30 RUNS FOR THE SYNTHETIC DATA STREAMS WITH GRADUAL DRIFTS

Data Stream	Average Prequential Accuracy (Standard Deviation) Across 30 Runs For The Synthetic Data Streams							Friedman Test p-value	Nemenyi Post Hoc Test p-value					
	CDCMS	HTNB	OzaBag	DP	OAUE	RCD	DDD		CDCMS vs HTNB	CDCMS vs OzaBag	CDCMS vs DP	CDCMS vs OAUE	CDCMS vs RCD	CDCMS vs DDD
Sine1	87.737% (6.109%)	67.745% (22.178%)	75.283% (23.252%)	<b>88.3%</b> (5.958%)	85.92% (9.998%)	87.806% (6.026%)	<b>87.786%</b> (7.341%)	<2.2E-16	2.3E-08	0.00034	0.00074	0.17789	0.91726	0.13530
Sine2	<b>92.692%</b> (7.666%)	65.555% (26.393%)	74.39% (26.552%)	<b>91.689%</b> (7.935%)	89.968% (12.008%)	90.39% (8.643%)	<b>91.632%</b> (9.236%)	<2.2E-16	6.3E-14	2.9E-14	0.06265	1.6E-11	2.2E-06	0.13530
Agr1	<b>89.53%</b> (8.576%)	74.003% (11.758%)	71.998% (12.03%)	85.268% (10.152%)	<b>89.707%</b> (9.193%)	85.634% (8.302%)	87.293% (10.867%)	<2.2E-16	1.7E-12	7.6E-14	2.9E-07	0.8962	0.0020	0.3575
Agr2	<b>78.494%</b> (11.311%)	67.301% (12.327%)	68.304% (12.175%)	71.877% (9.767%)	<b>80.458%</b> (9.153%)	73.954% (10.229%)	74.09% (11.955%)	<2.2E-16	5.2E-14	6.5E-12	1.6E-06	0.55289	0.10088	0.10088
Agr3	<b>82.895%</b> (9.108%)	70.068% (8.178%)	70.295% (8.731%)	75.833% (7.616%)	<b>83.413%</b> (9.116%)	78.58% (9.037%)	77.759% (10.457%)	<2.2E-16	8.2E-14	1.4E-13	2.7E-05	0.553	0.074	0.026
Agr4	<b>81.187%</b> (11.45%)	69.749% (14.202%)	67.592% (14.072%)	75.46% (12.894%)	<b>82.505%</b> (10.119%)	77.783% (11.093%)	77.647% (13.051%)	<2.2E-16	1.6E-11	4.9E-14	1.6E-06	0.55289	0.22920	0.03731
SEA1	<b>87.591%</b> (1.488%)	86.117% (2.239%)	86.581% (2.777%)	86.879% (1.709%)	87.168% (2.505%)	86.71% (1.855%)	<b>87.525%</b> (1.652%)	<2.2E-16	7.5E-14	6.8E-14	2.0E-05	0.03110	4.6E-10	0.98114
SEA2	<b>86.841%</b> (1.652%)	85.036% (3.392%)	85.733% (3.62%)	86.296% (1.854%)	<b>86.713%</b> (2.579%)	86.033% (1.965%)	<b>86.738%</b> (1.964%)	<2.2E-16	6.7E-14	1.0E-13	0.00015	0.59364	7.3E-09	0.96311
STA1	<b>98.268%</b> (4.193%)	86.462% (14.325%)	86.782% (13.774%)	<b>98.183%</b> (4.306%)	96.898% (7.258%)	97.783% (4.91%)	<b>98.132%</b> (4.466%)	<2.2E-16	7.0E-14	4.7E-13	0.4720	5.7E-07	0.0040	1.0000
STA2	<b>98.155%</b> (4.623%)	78.25% (16.809%)	76.511% (16.635%)	<b>98.079%</b> (4.65%)	96.842% (7.484%)	97.713% (5.402%)	97.949% (5.028%)	<2.2E-16	7.0E-14	4.9E-14	0.5936	3.7E-11	4.2E-06	0.0144

Lime or light grey in the column of Data Stream means that synthetic data stream consists of recurring concepts. Cells with bold text represent the best accuracy (or approaches have no significant difference with best one) in data stream. Coloured Nemenyi p-value cells represent that there is significant difference ( $p\text{-value} \leq 0.05$ ) and the proposed approach performed better (lime or light grey) / worse (orange or dark grey) based on the mean and standard deviation of its prequential accuracy. Nemenyi p-value cells in white represent no significant difference.

TABLE II  
AVERAGE PREQUENTIAL ACCURACY (STANDARD DEVIATION) ACROSS 30 RUNS FOR THE SYNTHETIC DATA STREAMS WITH ABRUPT DRIFTS

Data Stream	Average Prequential Accuracy (Standard Deviation) Across 30 Runs For The Synthetic Data Streams							Friedman Test p-value	Nemenyi Post Hoc Test p-value					
	CDCMS	HTNB	OzaBag	DP	OAUE	RCD	DDD		CDCMS vs HTNB	CDCMS vs OzaBag	CDCMS vs DP	CDCMS vs OAUE	CDCMS vs RCD	CDCMS vs DDD
Sine1	<b>90.248%</b> (1.586%)	74.35% (18.25%)	78.58% (20.238%)	89.374% (1.877%)	86.846% (9.164%)	<b>89.527%</b> (1.827%)	88.54% (6.961%)	<2.2E-16	6.3E-14	4.9E-14	0.00122	2.4E-11	0.28928	8.6E-05
Sine2	<b>95.894%</b> (3.055%)	65.559% (27.081%)	76.702% (25.508%)	91.557% (7.497%)	91.593% (9.918%)	<b>94.061%</b> (4.032%)	93.755% (7.22%)	<2.2E-16	6.3E-14	4.9E-14	1.6E-11	1.6E-06	0.22920	0.03731
Agr1	<b>90.31%</b> (8.367%)	76.283% (12.768%)	76.282% (10.802%)	87.209% (8.58%)	<b>90.666%</b> (8.733%)	86.706% (8.735%)	88.805% (9.617%)	<2.2E-16	8.2E-14	1.7E-13	0.00620	0.59364	1.6E-06	0.59364
Agr2	<b>79.528%</b> (10.862%)	67.575% (12.765%)	68.871% (12.383%)	72.668% (10.878%)	<b>81.807%</b> (8.609%)	74.711% (10.684%)	75.038% (12.379%)	<2.2E-16	4.9E-14	1.6E-11	1.6E-06	0.55289	0.03110	0.25817
Agr3	<b>83.888%</b> (9.128%)	69.972% (8.402%)	70.404% (8.892%)	76.95% (8.428%)	<b>84.499%</b> (8.921%)	77.986% (9.576%)	79.789% (10.644%)	<2.2E-16	8.2E-14	1.7E-13	3.0E-06	0.55289	0.02581	0.22920
Agr4	<b>80.775%</b> (12.21%)	71.867% (15.6%)	70.141% (14.376%)	78.063% (12.118%)	<b>82.984%</b> (10.584%)	77.679% (11.969%)	78.324% (13.253%)	<2.2E-16	3.7E-11	5.2E-14	0.15553	0.47195	8.6E-05	0.02581
SEA1	<b>87.569%</b> (1.528%)	86.136% (2.467%)	86.58% (2.974%)	86.434% (1.893%)	<b>87.378%</b> (2.551%)	86.527% (1.633%)	<b>87.538%</b> (1.701%)	<2.2E-16	6.7E-14	8.6E-05	1.0E-13	0.1555	1.6E-08	1.0000
SEA2	<b>86.935%</b> (1.761%)	85.05% (3.605%)	85.773% (3.856%)	85.976% (2.514%)	<b>86.953%</b> (2.667%)	85.846% (2.105%)	<b>86.884%</b> (2.08%)	<2.2E-16	6.6E-14	4.2E-12	0.00122	0.99994	3.3E-08	0.71190
STA1	<b>99.936%</b> (0.223%)	90.324% (11.997%)	90.142% (11.639%)	99.904% (0.291%)	98.129% (5.867%)	<b>99.945%</b> (0.203%)	98.964% (3.343%)	<2.2E-16	1.0E-11	5.0E-14	0.5529	1.6E-06	0.5529	0.0062
STA2	<b>99.953%</b> (0.156%)	86.898% (15.764%)	87.342% (16.241%)	99.893% (0.3%)	98.004% (6.281%)	<b>99.946%</b> (0.161%)	99.024% (3.2%)	<2.2E-16	6.6E-14	5.0E-14	0.0077	2.4E-11	0.6340	2.2E-06

Lime or light grey in the column of Data Stream means that synthetic data stream consists of recurring concepts. Cells with bold text represent the best accuracy (or approaches have no significant difference with best one) in data stream. Coloured Nemenyi p-value cells represent that there is significant difference ( $p\text{-value} \leq 0.05$ ) and the proposed approach performed better (lime or light grey) / worse (orange or dark grey) based on the mean and standard deviation of its prequential accuracy. Nemenyi p-value cells in white represent no significant difference.

TABLE III  
AVERAGE PREQUENTIAL ACCURACY (STANDARD DEVIATION) ACROSS 30 RUNS FOR THE REAL WORLD DATA STREAMS

Data Stream	Average Prequential Accuracy (Standard Deviation) Across 30 Runs For The Real World Data Streams							Friedman Test p-value	Nemenyi Post Hoc Test p-value					
	CDCMS	HTNB	OzaBag	DP	OAUE	RCD	DDD		CDCMS vs HTNB	CDCMS vs OzaBag	CDCMS vs DP	CDCMS vs OAUE	CDCMS vs RCD	CDCMS vs DDD
KDD Cup 99	<b>99.738%</b> (0.191%)	99.65% (0.291%)	<b>99.728%</b> (0.278%)	99.65% (0.291%)	99.663% (1.073%)	99.65% (0.291%)	99.717% (0.277%)	<2.2E-16	4.9E-14	0.4325	4.9E-14	1.6E-06	4.9E-14	0.0117
Power Supply	<b>16.247%</b> (4.115%)	14.833% (3.359%)	14.907% (3.28%)	14.637% (3.671%)	<b>16.237%</b> (4.523%)	13.84% (3.226%)	14.838% (3.503%)	<2.2E-16	8.0E-07	0.01172	1.2E-13	0.99983	8.1E-14	2.2E-06
Sensor	<b>89.38%</b> (8.867%)	56.283% (17.35%)	71.126% (15.96%)	83.504% (16.628%)	<b>92.332%</b> (7.306%)	53.402% (18.793%)	85.838% (12.99%)	<2.2E-16	1.6E-11	1.6E-06	0.0062	0.5529	4.9E-14	0.5529

Cells with bold text represent the best accuracy (or approaches have no significant difference with best one) in data stream. Coloured Nemenyi p-value cells represent that there is significant difference ( $p\text{-value} \leq 0.05$ ) and the proposed approach performed better (lime or light grey) / worse (orange or dark grey) based on the mean and standard deviation of its prequential accuracy. Nemenyi p-value cells in white represent no significant difference.