

# “Differential Evolution with Duplication Analysis for Feature Selection in Classification”

## *Online Supplementary Materials*

Peng Wang, *Graduate Student Member, IEEE*, Bing Xue, *Senior Member, IEEE*, Jing Liang, *Senior Member, IEEE*, and Mengjie Zhang, *Fellow, IEEE*

TABLE I: *Ac* results of the three methods

Dataset	VS-CCPSO	NDEDA	<i>r</i> -NDEDA
Zoo	<b>90.43</b> ±1.31↑	90.17±0.59↑	89.78±0.24
WBCD	<b>94.76</b> ±1.11↑	94.15±0.02o	94.15±0.00
Ionosphere	86.95±3.48↓	88.15±1.87o	<b>88.68</b> ±1.74
Movement	73.98±2.25↓	<b>77.39</b> ±1.55o	76.99±1.69
Hillvally	54.93±1.84o	<b>55.38</b> ±1.29o	55.37±1.18
Musk1	97.95±0.56o	<b>98.08</b> ±0.57↑	97.81±0.54
Multiple	96.49±0.38o	96.40±0.35o	<b>96.50</b> ±0.41
Arrhythmia	59.85±2.77↓	62.14±1.92o	<b>62.38</b> ±2.43
CNAE	84.59±1.87o	<b>87.28</b> ±1.53↑	84.13±2.13
QSAR	93.87±0.35↑	<b>93.93</b> ±0.26↑	93.16±0.40
AD	97.36±0.50↑	<b>97.69</b> ±0.28↑	96.86±0.30
SRBCT	89.60±7.05↓	98.95±1.59o	<b>99.18</b> ±1.37
Leukemia	<b>91.67</b> ±5.26↑	89.31±3.96o	86.71±5.86
Leukemia1	67.27±4.61↓	78.70±6.42o	<b>81.47</b> ±7.04
9Tumor	37.41±7.98↓	49.52±6.72o	<b>50.04</b> ±5.99
DLBCL	94.58±3.90↓	<b>98.30</b> ±2.70o	98.08±2.35
Leukemia2	92.27±4.41o	<b>94.18</b> ±2.78↑	92.49±3.12
11Tumor	77.99±3.62o	79.46±3.36o	<b>79.81</b> ±3.81
Rank	2.50	1.69	1.81

TABLE II: *Size* results of the three methods

Dataset	VS-CCPSO	NDEDA	<i>r</i> -NDEDA
Zoo	<b>5.6</b> ±1.2o	5.8±0.2o	5.7±0.1
WBCD	8.6±2.1↓	<b>4.1</b> ±0.2o	<b>4.1</b> ±0.2
Ionosphere	4.5±1.1↓	4.2±0.3↓	<b>4.0</b> ±0.3
Movement	<b>20.7</b> ±4.7↑	23.7±3.8o	24.4±4.2
Hillvally	<b>20.0</b> ±5.9↑	24.8±3.9o	25.9±3.0
Musk1	<b>40.9</b> ±5.3↑	43.5±5.6o	44.6±4.5
Multiple	<b>78.0</b> ±6.9o	81.4±6.2o	81.7±8.2
Arrhythmia	<b>25.9</b> ±5.8o	42.3±8.8↓	26.0±4.5
CNAE	146.4±14.9↓	290.2±22.8↓	<b>47.7</b> ±2.7
QSAR	226.6±38.2↓	273.9±17.9↓	<b>24.7</b> ±3.2
AD	325.8±35.0↓	419.9±22.0↓	<b>27.5</b> ±6.2
SRBCT	<b>40.6</b> ±7.1↑	391.4±37.9↓	205.8±25.2
Leukemia	<b>159.2</b> ±56.9↑	833.4±66.3o	840.9±57.7
Leukemia1	<b>264.0</b> ±100.5↑	979.7±52.0o	981.2±75.5
9Tumor	<b>671.8</b> ±163.8↑	1,431.3±97.2o	1,433.0±100.4
DLBCL	<b>555.9</b> ±270.5↑	1,320.1±62.8o	1,325.6±67.5
Leukemia2	<b>589.3</b> ±174.7↑	2,083.2±129.0o	2,047.6±126.2
11Tumor	<b>1,511.6</b> ±465.1↑	3,223.1±233.6↓	3,121.3±142.0
Rank	1.39	2.47	2.14

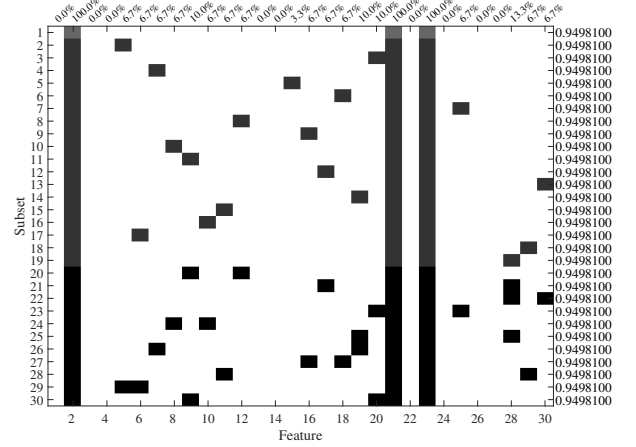


Fig. 1. Frequency matrix on the WBCD dataset. ‘Feature’ means one of the original features in the WBCD dataset, and ‘Subset’ means one of the obtained feature subsets from NDEDA. The training accuracy using each feature subset is shown on the right side, and the selected frequency of each feature is shown on the upper side.

TABLE III: Average training time of eight algorithms (unit: minute).

Dataset	<i>r3pso</i>	NCDE	NSDE	NShDE	LBPADe	SBDA	VS-CCPSO	NDEDA	<i>r</i> -NDEDA
Zoo	<b>0.1</b>	<b>0.1</b>	<b>0.1</b>	<b>0.1</b>	0.2	0.7	0.8	0.4	0.6
WBCD	2.1	2.1	2.0	2.1	<b>0.9</b>	2.1	1.5	2.7	1.5
Ionosphere	1.1	1.0	1.0	1.1	<b>0.8</b>	2.1	1.6	1.7	1.6
Movement	3.3	3.2	3.1	3.3	<b>3.0</b>	6.4	4.4	10.0	9.3
Hillvally	29.0	27.7	26.8	28.2	6.8	13.5	<b>5.2</b>	17.1	12.7
Musk1	164.8	150.4	147.6	150.4	33.5	72.4	<b>11.3</b>	70.6	66.1
Multiple	259.4	238.8	234.6	243.0	107.5	246.9	<b>20.0</b>	247.8	193.5
Arrhythmia	19.9	18.6	18.2	20.2	22.7	38.7	<b>15.7</b>	125.4	111.5
CNAE	198.8	187.7	178.5	190.5	137.0	315.6	<b>22.1</b>	490.7	161.4
QSAR	3,824.5	1,312.4	1,630.0	1,256.1	340.5	2,457.3	<b>97.2</b>	1,718.6	241.8
AD	4,895.3	4,245.6	3,891.7	4,447.6	1,081.2	1,346.0	<b>145.5</b>	2,452.6	308.5
SRBCT	4.0	5.6	<b>3.8</b>	13.7	60.2	39.3	16.0	709.9	643.7
Leukemia	7.6	12.4	<b>6.6</b>	31.4	121.8	69.4	15.7	1,466.4	767.5
Leukemia1	16.5	11.1	<b>6.6</b>	28.3	635.3	67.6	15.9	1,519.7	825.7
DLBCL	12.7	19.9	<b>11.7</b>	56.5	156.5	88.8	18.9	1,985.8	977.8
9Tumor	13.1	10.6	<b>5.6</b>	29.8	554.9	69.4	17.5	1,597.8	1,254.7
Leukemia2	39.4	25.5	<b>14.0</b>	74.7	956.9	175.8	19.7	2,820.6	2,463.5
11Tumor	290.3	82.8	80.0	135.8	1,069.7	322.3	<b>24.4</b>	2,399.5	2,132.8

TABLE IV: Average results of *Ac* (%) on test sets

Dataset	NDEDA- <i>l</i>	NDEDA- <i>g</i>	NDEDA- <i>lg</i>	NDEDA- <i>lgs</i>	NDEDA- <i>sr</i>	NDEDA
Zoo	90.70±2.27o	<b>91.40</b> ±1.97o	89.84±2.98o	90.65±2.24o	90.69±0.89o	90.17±0.59
WBCD	<b>94.18</b> ±0.21o	94.15±0.20o	93.95±0.51↓	94.07±0.33o	94.15±0.0o	94.15±0.02
Ionosphere	83.85±3.28↓	81.92±3.36↓	84.38±3.88↓	86.83±3.32o	<b>88.36</b> ±2.18o	88.15±1.87
Movement	75.81±1.65↓	76.18±2.38↓	76.16±2.57↓	76.89±1.64o	76.68±1.42o	<b>77.39</b> ±1.55
Hillvally	53.72±1.52↓	54.76±1.46o	54.67±1.61o	<b>55.83</b> ±1.15o	55.31±0.91o	55.38±1.29
Musk1	97.12±0.80↓	96.93±0.75↓	97.49±0.66↓	97.51±0.52↓	97.67±0.47↓	<b>98.08</b> ±0.58
Multiple	96.56±0.37o	96.56±0.43o	96.58±0.42o	96.43±0.34o	<b>96.65</b> ±0.32o	96.40±0.35
Arrhythmia	60.38±2.96↓	60.53±2.83↓	62.35±2.25o	61.98±2.96o	<b>62.43</b> ±2.93o	62.14±1.92
CNAE	83.11±1.85↓	82.74±2.33↓	83.56±2.19↓	84.97±1.56↓	86.11±1.97o	<b>87.28</b> ±1.53
QSAR	93.54±0.21↓	93.48±0.26↓	93.76±0.22↓	93.80±0.29o	93.88±0.39o	<b>93.93</b> ±0.26
AD	97.18±0.42↓	97.18±0.41↓	97.34±0.44↓	97.45±0.43o	97.44±0.38o	<b>97.69</b> ±0.28
SRBCT	95.52±3.22↓	94.88±3.66↓	97.40±2.30o	98.68±1.97o	98.11±2.40o	<b>98.95</b> ±1.59
Leukemia	86.75±1.88↓	84.88±3.52↓	89.09±4.62o	<b>90.60</b> ±5.05o	90.35±5.84o	89.31±3.96
Leukemia1	76.99±3.48o	75.73±5.69o	80.77±3.72↑	<b>81.11</b> ±7.94↑	80.68±6.25↑	78.70±6.42
9Tumor	44.83±5.97↓	44.25±5.08↓	48.37±4.49o	48.78±7.20o	48.92±6.91o	<b>49.52</b> ±6.72
DLBCL	97.50±1.75o	97.45±1.86o	97.18±2.78↓	97.43±3.52o	98.02±2.54o	<b>98.30</b> ±2.70
Leukemia2	90.35±1.06↓	90.04±1.62↓	92.18±1.32↓	92.80±3.20↓	94.14±2.91o	<b>94.18</b> ±2.78
11Tumor	75.64±2.55↓	74.67±3.50↓	80.29±3.75o	<b>80.56</b> ±3.64o	80.21±3.13o	79.46±3.36
Sum	0/13/5	0/12/6	1/9/8	1/3/14	1/1/16	N/A

TABLE V: Average number of selected features (*Size*)

Dataset	NDEDA- <i>l</i>	NDEDA- <i>g</i>	NDEDA- <i>lg</i>	NDEDA- <i>lgs</i>	NDEDA- <i>sr</i>	NDEDA
Zoo	6.5±1.3o	6.4±1.8o	6.2±1.5o	7.0±1.0↓	6.7±0.2o	<b>5.8</b> ±0.3
WBCD	7.6±2.3↓	5.5±1.5↓	4.6±0.9o	5.6±0.8↓	4.3±0.0o	<b>4.1</b> ±0.3
Ionosphere	7.0±2.2↓	9.7±2.9↓	6.2±2.2↓	5.3±1.9o	4.4±1.2o	<b>4.2</b> ±0.4
Movement	30.9±3.4o	33.7±4.4↓	28.4±4.1↓	28.8±4.9↓	28.0±4.7↓	<b>23.7</b> ±3.8
Hillvally	31.4±5.3↓	33.3±4.3↓	25.8±3.9o	25.2±3.3o	25.3±4.6o	<b>24.8</b> ±3.9
Musk1	61.5±6.6↓	65.4±5.8↓	53.6±7.0↓	51.3±5.9↓	53.1±5.1↓	<b>43.5</b> ±5.6
Multiple	97.8±5.9↓	102.3±6.4↓	92.8±6.9↓	92.3±6.9↓	91.8±6.4↓	<b>81.4</b> ±6.2
Arrhythmia	88.4±13.1↓	93.3±8.1↓	67.0±10.4↓	53.2±12.3↓	48.9±7.7o	<b>42.3</b> ±8.8
CNAE	344.3±17.9↓	352.3±12.8↓	315.3±28.7↓	311.4±30.3↓	302.0±18.8o	<b>290.2</b> ±22.8
QSAR	385.8±21.0↓	403.7±16.3↓	335.3±14.4↓	313.7±23.0↓	310.2±21.7↓	<b>273.9</b> ±17.9
AD	568.7±43.0↓	615.7±22.2↓	465.2±58.9↓	<b>394.9</b> ±52.3o	462.5±44.7↓	419.9±22.0
SRBCT	830.3±72.8↓	902.4±23.1↓	505.1±87.9↓	<b>374.1</b> ±61.0↑	455.3±47.4o	391.4±37.9
Leukemia	1,818.5±90.5↓	2,020.2±39.5↓	988.3±194.2↓	<b>696.3</b> ±97.3↑	898.4±89.8o	833.4±66.3
Leukemia1	1,811.5±124.2↓	2,077.8±34.4↓	1,416.5±58.3↓	1,131.6±80.3↓	1,082.1±92.2↓	<b>979.7</b> ±52.0
9Tumor	2,090.5±175.3↓	2,259.0±44.3↓	1,915.1±71.2↓	1,591.5±108.2↓	1,541.6±89.3↓	<b>1,431.3</b> ±97.2
DLBCL	2,574.0±194.5↓	2,505.4±190.1↓	1,422.3±239.3↓	<b>1,126.7</b> ±149.4↑	1,537.8±172.0↓	1,320.1±62.8
Leukemia2	3,605.3±263.7↓	4,438.7±55.5↓	2,898.8±117.6↓	2,325.2±116.1↓	2,194.3±104.3↓	<b>2,083.2</b> ±129.0
11Tumor	4,706.2±196.1↓	4,975.5±61.9↓	4,226.2±140.2↓	3,660.4±190.5↓	<b>3,167.7</b> ±189.2o	3,223.1±233.6
Sum	0/16/2	0/17/1	0/15/3	3/12/3	0/9/9	N/A

TABLE VI: Average results of  $F1$  score (%) on the test sets

Dataset	$r3pso$	NCDE	NSDE	NShDE	LBPADe	SBDA	VS-CCPSO	NDEDA
Zoo	88.14±4.73o	87.87±1.60o	88.23±0.41o	88.24±0.41o	87.23±1.75↓	90.22±2.70↑	<b>88.51</b> ±1.56o	88.00±0.84
WBCD	92.67±1.36↓	93.76±0.87↓	93.71±1.10↓	93.48±1.30↓	94.10±0.15o	94.31±0.50o	<b>94.77</b> ±1.11↑	94.15±0.02
Ionosphere	78.97±2.17↓	87.13±3.01o	87.29±3.15o	86.77±3.59o	84.54±3.01↓	80.13±2.24↓	86.41±3.79o	<b>87.79</b> ±2.04
Movement	74.98±2.22↓	76.41±1.69↓	76.07±1.67↓	76.67±1.52↓	77.08±1.17o	76.48±1.60↓	73.27±2.55↓	<b>77.52</b> ±1.65
Hillvally	52.65±1.87↓	54.47±1.43↓	53.85±1.17↓	54.33±1.40↓	54.48±1.15↓	53.41±1.95↓	53.38±2.13↓	<b>55.34</b> ±1.29
Musk1	96.14±0.59↓	97.45±0.42↓	97.65±0.39↓	97.62±0.41↓	97.61±0.53↓	96.46±0.54↓	97.95±0.56o	<b>98.08</b> ±0.58
Multiple	96.68±0.38↑	96.65±0.28↑	96.61±0.37↑	96.70±0.31↑	96.56±0.39o	<b>96.73</b> ±0.38↑	96.48±0.39o	96.39±0.35
Arrhythmia	49.78±3.02↓	54.07±2.73o	56.57±2.28o	55.25±2.83o	55.08±2.41o	<b>59.58</b> ±1.83↑	49.99±3.77↓	55.54±2.64
CNAE	77.23±5.23↓	85.14±2.29↓	86.49±1.66↓	86.62±1.33↓	84.74±1.59↓	81.52±2.20↓	75.34±1.92↓	<b>87.76</b> ±1.39
QSAR	91.88±0.45↓	92.21±0.36↓	92.34±0.39↓	92.25±0.28↓	92.14±0.41↓	<b>93.56</b> ±0.32↑	92.62±0.49o	92.70±0.35
AD	96.69±0.45↓	97.12±0.42↓	97.05±0.49↓	96.99±0.44↓	97.21±0.37↓	96.91±0.47↓	97.32±0.50↓	<b>97.63</b> ±0.29
SRBCT	92.82±2.01↓	95.70±1.47↓	96.48±1.47↓	96.11±1.31↓	96.39±1.57↓	94.00±2.88↓	89.87±7.36↓	<b>99.02</b> ±1.48
Leukemia	80.54±5.33↓	86.23±2.70↓	87.19±2.60o	87.03±2.48↓	86.51±3.62↓	85.61±4.40↓	<b>91.44</b> ±5.52↑	89.01±4.18
Leukemia1	73.97±6.07↓	78.05±6.28o	<b>81.03</b> ±5.48o	78.31±4.71o	76.83±4.96o	75.91±7.06o	65.67±5.25↓	77.80±7.24
9Tumor	35.95±6.34↓	38.65±6.19↓	37.94±6.34↓	39.57±5.56↓	39.66±6.49↓	<b>45.93</b> ±7.30o	36.74±8.65↓	45.32±7.01
DLBCL	97.17±1.45↓	97.63±1.40o	97.29±1.49o	97.74±1.32o	97.31±1.03↓	98.47±2.01o	94.39±4.09↓	<b>98.31</b> ±2.69
Leukemia2	90.27±1.05↓	90.62±0.93↓	90.71±0.86↓	90.70±1.18↓	91.03±0.86↓	90.15±1.69↓	92.35±4.35↓	<b>94.29</b> ±2.71
11Tumor	74.49±2.89↓	75.11±2.55↓	75.49±2.74↓	75.02±2.84↓	76.22±3.16↓	76.23±3.18↓	77.56±3.86↓	<b>79.49</b> ±3.49
Rank	7.22 (1/16/1)	4.89 (1/12/5)	4.00 (1/11/6)	4.06 (1/12/5)	4.50 (0/13/5)	4.17 (4/10/4)	4.89 (2/11/5)	2.28