Class Overlap Matters: Revisiting SOUP for Multi-Class Imbalanced Medical Data

Table a1 to Table a9 present the performance results of SOUP method and the three integrated strategies across ten multi-class medical datasets, evaluated using various classifiers and multi-class assessment metrics.

Table a1
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the AvAcc [%] Metric with KNN as the Base Classifier.

Dataset	**SOUP + WOA**	**SOUP + GA**	**SOUP + PSO**	**SOUP**
cleveland	**69.27**	68.65	64.4	52.07
dermatology	**96.31**	95.58	93.46	89.39
dna	**79**	78.48	73.42	75.64
hungarian	**76.32**	75.73	73.81	58.49
Laryngeal	**93.69**	91.92	69.68	78.94
vertebra-column	**87.43**	86.71	75.98	86.19
burczynski	**74.83**	74.13	73.32	68.86
chiaretti	**85.56**	84.51	85	65.61
khan	87.79	**88.48**	87.76	83.65
sorlie	**86.04**	85.87	84.51	80.36

Table a2
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the CBA [%] Metric with KNN as the Base Classifier

Dataset	**SOUP + WOA**	**SOUP + GA**	**SOUP + PSO**	**SOUP**
cleveland	**25**	24.65	22.26	16.72
dermatology	80.9	**83.05**	76.65	65.14
dna	**59.9**	59.62	54.26	53.92
hungarian	24.11	**26.18**	22.93	17.54
Laryngeal	**87.45**	84.79	59.24	65.04
vertebra-column	**72.48**	71.12	58.08	70.29
burczynski	**57.59**	56.08	55.73	50.52
chiaretti	**43.61**	43.16	41.19	21.05
khan	**72.49**	71.45	69.63	59.98
sorlie	**61**	60.85	58.07	51.42

Table a3
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the macro-F1 [%] Metric with KNN as the Base Classifier

Dataset	**SOUP + WOA**	**SOUP + GA**	**SOUP + PSO**	**SOUP**
cleveland	**27.4**	27.2	24.82	24.76
dermatology	83.48	**87.13**	81.59	71.85
dna	**74.33**	73.29	67.26	72.55
hungarian	27.99	**29.64**	27.18	26.55
Laryngeal	**90.74**	88.53	63.87	71.69
vertebra-column	**78.79**	77.41	64.29	77.85
burczynski	**66.95**	66.91	65.81	61.13
chiaretti	**58.93**	58.2	56.12	44.2
khan	79.91	**82.38**	81.12	75.4
sorlie	**76.09**	75.94	73.78	69.69

Table a4
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the AvAcc [%] Metric with MLP as the Base Classifier.

Dataset	SOUP + WOA	SOUP + GA	SOUP + PSO	SOUP
cleveland	**74.75**	73.25	73.41	54.52
dermatology	**98.91**	98.74	97.83	98.23
dna	**96.18**	94.89	91.03	95
hungarian	**81.41**	81.11	80.13	60.61
Laryngeal	**83.2**	71.4	66.51	78.88
vertebra-column	**87.99**	87.31	76.84	84.39
burczynski	59.39	66.74	**67.24**	63.58
chiaretti	59.81	66.02	**67.74**	51.67
khan	**98.84**	98.73	97.94	95.51
sorlie	**91.21**	90.52	89.86	88.05

Table a5
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the CBA [%] Metric with MLP as the Base Classifier.

Dataset	SOUP + WOA	SOUP + GA	SOUP + PSO	SOUP
cleveland	**27.83**	26.55	25.51	18.37
dermatology	**95.25**	94.21	90.88	92.95
dna	**92.59**	90.09	84.56	88.74
hungarian	**25.72**	25.5	24.83	18.5
Laryngeal	**72.55**	50.86	48.33	62.39
vertebra-column	**72.37**	72.31	54.76	68.3
burczynski	34.11	**41.29**	40.16	36.09
chiaretti	**9.96**	9.91	9.15	6.28
khan	93.36	**96.01**	93.34	86.56
sorlie	**73.36**	71.33	70.65	66.62

Table a6
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the macro-F1 [%] Metric with MLP as the Base Classifier.

Dataset	SOUP + WOA	SOUP + GA	SOUP + PSO	SOUP
cleveland	**31.99**	31.94	31.16	26.46
dermatology	**96.49**	95.88	93.21	94.77
dna	**93.77**	91.81	86.06	92.26
hungarian	32.87	**34**	32.5	26.41
Laryngeal	**77.87**	63.79	58.19	71.54
vertebra-column	**76.74**	75.45	61.89	73.57
burczynski	53.24	55.41	**55.63**	52.3
chiaretti	**28.82**	24.81	27.17	21.23
khan	96.72	**97.9**	96.66	92.88
sorlie	**80.66**	79.2	77.91	76.18

Table a7
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the AvAcc [%] Metric with NB as the Base Classifier.

Dataset	SOUP + WOA	SOUP + GA	SOUP + PSO	SOUP
cleveland	60.86	60.6	**64.56**	64.34
dermatology	**97.09**	96.01	95.2	96.5
dna	**95.31**	91.73	84.57	95.22
hungarian	58.25	48.89	50.95	**70.79**
Laryngeal	92.67	**94.69**	75.99	86.7
vertebra-column	**87.35**	86.82	78.01	85.16
burczynski	**80.87**	80.82	80.62	78.84
chiaretti	**85.28**	84.99	84.63	74.12
khan	**81.59**	80.18	81.56	79.72
sorlie	86.12	**86.42**	85.39	84.43

Table a8
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the CBA [%] Metric with NB as the Base Classifier.

Dataset	**SOUP + WOA**	**SOUP + GA**	**SOUP + PSO**	**SOUP**
cleveland	23.05	22.02	22.32	**23.99**
dermatology	**81.46**	80.13	77.44	81.41
dna	**88.93**	80.49	68.42	88.77
hungarian	18.46	14.23	14.77	**24.09**
Laryngeal	85.12	**89.11**	63.71	78.16
vertebra-column	**71.55**	70.92	60.37	67.61
burczynski	**60.84**	60.04	60.1	60.82
chiaretti	**26.97**	**26.97**	26.87	20.57
khan	**55.58**	52.44	54.9	54.83
sorlie	55.37	55.32	53.84	**58.52**

Table a9
Comparison between SOUP and SOUP Combined with Various Feature Selection Strategies, Evaluated by the macro-F1 [%] Metric with NB as the Base Classifier.

Dataset	**SOUP + WOA**	**SOUP + GA**	**SOUP + PSO**	**SOUP**
cleveland	**33.04**	31.54	30.65	31.78
dermatology	**90.39**	88.56	86.41	90.27
dna	**92.63**	87.69	80.37	92.54
hungarian	31.09	27.81	28.72	**36.01**
Laryngeal	89.13	**92.16**	71.38	81.42
vertebra-column	**77.6**	77.21	66.2	75.75
burczynski	**72.44**	72.43	72.07	68.83
chiaretti	**42.31**	42.1	41.6	35.44
khan	73.77	72.69	**74.27**	73.08
sorlie	72.13	**72.3**	70.88	71.16