

## Original article: Classification of Parkinson's Disease from Voice - Analysis of Data Selection Bias

Authors: Alexander Brenner, Catharina Marie van Alen, Lucas Plagwitz and Julian Varghese

### Supplement

**Table S1.** Overview of the samples. All subsets are derived from the mPower dataset. Stratified sampling was applied to achieve the same gender and age distribution between the classes PD and HC, while also balancing the number of records representing the respective classes.

Subsets	Class	#Records	#Participants	Age
Complete	PD	9929	524 (201 female, 322 male, 1 unknown)	60.6 ( $\pm$ 10.7)
	HC	15631	2566 (408 female, 2151 male, 7 unknown)	32.9 ( $\pm$ 13.2)
Stratified sampling (keeping 50%)	PD	7111	500 (196 female, 303 male, unknown)	60.4 (10.7)
	HC	5669	1029 (306 female, 716 male, 7 unknown)	39.3 (16.7)
Stratified sampling (keeping 20%)	PD	2886	392 (158 female, 233 male, 1 unknown)	60.6 ( $\pm$ 11.1)
	HC	2226	530 (178 female, 345 male, 7 unknown)	45.7 ( $\pm$ 17.2)
Stratified sampling (completed at 3%)	PD	444	163 (74 female, 89 male)	62.1 ( $\pm$ 8.9)
	HC	444	141 (47 female, 94 male)	59.4 ( $\pm$ 9.1)
Random sampling (keeping 50%)	PD	5032	454 (178 female, 275 male)	60.9 ( $\pm$ 10.2)
	HC	7748	2087 (343 female, 1741 male)	33.1 ( $\pm$ 13.4)
Random sampling (keeping 20%)	PD	1991	345 (134 female, 211 male)	61.5 ( $\pm$ 9.9)
	HC	3121	1322 (219 female, 1102 male)	34.0 ( $\pm$ 13.9)
Random sampling (keeping 3%)	PD	333	160 (62 female, 98 male)	61.2 ( $\pm$ 9.4)
	HC	555	379 (60 female, 318 male)	36.0 ( $\pm$ 15.5)

**Table S2.** Classification performance on different mPower subsets. We report classification performance for each combination of training set and feature extraction method using the SVM classifier. Results are given as mean ( $\pm$  std) over the 5-fold CV test sets.

Setting	Accuracy	F1	ROC AUC	Specificity	Sensitivity
Surfboard					
Complete set, CV	0.8317 (0.0065)	0.7869 (0.0088)	0.8259 (0.0074)	0.8518 (0.0080)	0.8001 (0.0143)
Complete set, grouped CV	0.7129 (0.0150)	0.6114 (0.0457)	0.6904 (0.0150)	0.7786 (0.0431)	0.6023 (0.0682)
Stratified sampling (50%)	0.6135 (0.0282)	0.6437 (0.0454)	0.6127 (0.0225)	0.5927 (0.0443)	0.6327 (0.0509)
Stratified sampling (20%)	0.5620 (0.0529)	0.6047 (0.0368)	0.5601 (0.0558)	0.5283 (0.0828)	0.5920 (0.0527)
Stratified sampling (3%)	0.5117 (0.0373)	0.5267 (0.0531)	0.5109 (0.0398)	0.4728 (0.1086)	0.5491 (0.0715)
Random sampling (50%)	0.7019 (0.0204)	0.6123 (0.0365)	0.6895 (0.0228)	0.7628 (0.0241)	0.6163 (0.0544)
Random sampling (20%)	0.6779 (0.0448)	0.5733 (0.0705)	0.6619 (0.0488)	0.7466 (0.0140)	0.5773 (0.1032)
Random sampling (3%)	0.6636 (0.0252)	0.5358 (0.0447)	0.6397 (0.0266)	0.7471 (0.0614)	0.5324 (0.0566)
ComParE					
Complete set, CV	0.8241 (0.0018)	0.7764 (0.0026)	0.8172 (0.0022)	0.8483 (0.0036)	0.7862 (0.0060)
Complete set, grouped CV	0.7145 (0.0175)	0.6114 (0.0434)	0.6922 (0.0214)	0.7851 (0.0452)	0.5993 (0.0794)
Stratified sampling (50%)	0.6314 (0.0230)	0.6633 (0.0414)	0.6288 (0.0216)	0.6006 (0.0250)	0.6570 (0.0368)
Stratified sampling (20%)	0.5712 (0.0557)	0.6219 (0.0446)	0.5652 (0.0584)	0.5069 (0.0739)	0.6234 (0.0525)
Stratified sampling (3%)	0.5151 (0.0304)	0.5369 (0.0482)	0.5186 (0.0396)	0.4673 (0.0733)	0.5700 (0.0887)
Random sampling (50%)	0.7035 (0.0211)	0.6103 (0.0426)	0.6903 (0.0258)	0.7715 (0.0339)	0.6090 (0.0600)
Random sampling (20%)	0.6741 (0.0247)	0.5628 (0.0514)	0.6557 (0.0294)	0.7549 (0.0390)	0.5565 (0.0717)
Random sampling (3%)	0.6522 (0.0406)	0.4761 (0.0420)	0.6116 (0.0313)	0.7862 (0.0760)	0.4371 (0.0871)
AVEC					
Complete set, CV	0.8314 (0.0047)	0.7849 (0.0059)	0.8242 (0.0049)	0.8565 (0.0047)	0.7918 (0.0066)
Complete set, grouped CV	0.7096 (0.0202)	0.5997 (0.0406)	0.6848 (0.0199)	0.7905 (0.0441)	0.5791 (0.0713)
Stratified sampling (50%)	0.6204 (0.0194)	0.6509 (0.0369)	0.6185 (0.0179)	0.5967 (0.0191)	0.6404 (0.0433)
Stratified sampling (20%)	0.5653 (0.0638)	0.6147 (0.0516)	0.5594 (0.0662)	0.5058 (0.0766)	0.6131 (0.0609)
Stratified sampling (3%)	0.5235 (0.0462)	0.5593 (0.0365)	0.5281 (0.0544)	0.4467 (0.1020)	0.6095 (0.0809)
Random sampling (50%)	0.6999 (0.0256)	0.5986 (0.0374)	0.6826 (0.0259)	0.7795 (0.0239)	0.5858 (0.0592)
Random sampling (20%)	0.6710 (0.0309)	0.5527 (0.0587)	0.6506 (0.0348)	0.7607 (0.0417)	0.5406 (0.0817)
Random sampling (3%)	0.6522 (0.0394)	0.4887 (0.0494)	0.6171 (0.0299)	0.7740 (0.0805)	0.4603 (0.0861)
eGeMAPS					
Complete set, CV	0.7650 (0.0067)	0.7103 (0.0092)	0.7608 (0.0076)	0.7799 (0.0057)	0.7417 (0.0129)
Complete set, grouped CV	0.6543 (0.0302)	0.5414 (0.0314)	0.6335 (0.0264)	0.7254 (0.0423)	0.5416 (0.0688)
Stratified sampling (50%)	0.5799 (0.0223)	0.6045 (0.0432)	0.5769 (0.0206)	0.5713 (0.0499)	0.5825 (0.0587)
Stratified sampling (20%)	0.5261 (0.0559)	0.5666 (0.0492)	0.5218 (0.0574)	0.4953 (0.0637)	0.5483 (0.0555)
Stratified sampling (3%)	0.4973 (0.0309)	0.5117 (0.0531)	0.5042 (0.0314)	0.4696 (0.0675)	0.5389 (0.1059)
Random sampling (50%)	0.6514 (0.0186)	0.5472 (0.0466)	0.6341 (0.0193)	0.7176 (0.0153)	0.5505 (0.0364)
Random sampling (20%)	0.6364 (0.0296)	0.5397 (0.0367)	0.6242 (0.0274)	0.6871 (0.0258)	0.5612 (0.0486)
Random sampling (3%)	0.6226 (0.0252)	0.4871 (0.0563)	0.5993 (0.0278)	0.7036 (0.0626)	0.4951 (0.0824)

**Table S3.** Classification performance on different mPower subsets. We report classification performance for each combination of training set and feature extraction method using the CatBoost classifier. Results are given as mean ( $\pm$  std) over the 5-fold CV test sets.

Setting	Accuracy	F1	ROC AUC	Specificity	Sensitivity
<b>Surfboard</b>					
Complete set, CV	0.8462 (0.0045)	0.8063 (0.0067)	0.8422 (0.0057)	0.8599 (0.0022)	0.8246 (0.0111)
Complete set, grouped CV	0.7087 (0.0202)	0.6088 (0.0409)	0.6897 (0.0192)	0.7763 (0.0488)	0.6030 (0.0697)
Stratified sampling (50%)	0.6076 (0.0219)	0.6456 (0.0424)	0.6031 (0.0144)	0.5584 (0.0467)	0.6479 (0.0481)
Stratified sampling (20%)	0.5660 (0.0565)	0.6254 (0.0430)	0.5561 (0.0606)	0.4717 (0.0734)	0.6404 (0.0571)
Stratified sampling (3%)	0.4983 (0.0421)	0.5303 (0.0524)	0.4952 (0.0421)	0.4204 (0.1227)	0.5700 (0.0603)
Random sampling (50%)	0.7032 (0.0252)	0.6103 (0.0476)	0.6894 (0.0296)	0.7689 (0.0291)	0.6099 (0.0636)
Random sampling (20%)	0.6889 (0.0339)	0.5780 (0.0622)	0.6709 (0.0387)	0.7725 (0.0356)	0.5693 (0.1000)
Random sampling (3%)	0.6779 (0.0411)	0.5003 (0.0511)	0.6317 (0.0399)	0.8245 (0.0676)	0.4390 (0.0444)
<b>ComParE</b>					
Complete set, CV	0.8430 (0.0034)	0.8022 (0.0039)	0.8387 (0.0032)	0.8580 (0.0043)	0.8193 (0.0036)
Complete set, grouped CV	0.7236 (0.0163)	0.6271 (0.0419)	0.7040 (0.0180)	0.7898 (0.0430)	0.6183 (0.0632)
Stratified sampling (50%)	0.6349 (0.0214)	0.6725 (0.0368)	0.6311 (0.0167)	0.5835 (0.0238)	0.6787 (0.0435)
Stratified sampling (20%)	0.5582 (0.0543)	0.6244 (0.0443)	0.5461 (0.0580)	0.4415 (0.0634)	0.6506 (0.0648)
Stratified sampling (3%)	0.5328 (0.0354)	0.5566 (0.0315)	0.5396 (0.0405)	0.4866 (0.0718)	0.5927 (0.0824)
Random sampling (50%)	0.7180 (0.0247)	0.6295 (0.0425)	0.7040 (0.0270)	0.7809 (0.0362)	0.6271 (0.0524)
Random sampling (20%)	0.7033 (0.0172)	0.5972 (0.0383)	0.6841 (0.0200)	0.7861 (0.0338)	0.5822 (0.0618)
Random sampling (3%)	0.6849 (0.0368)	0.5049 (0.0203)	0.6373 (0.0231)	0.8357 (0.0539)	0.4389 (0.0534)
<b>AVEC</b>					
Complete set, CV	0.8413 (0.0040)	0.7997 (0.0046)	0.8366 (0.0039)	0.8578 (0.0061)	0.8154 (0.0062)
Complete set, grouped CV	0.7105 (0.0228)	0.6097 (0.0401)	0.6899 (0.0176)	0.7804 (0.0461)	0.5994 (0.0498)
Stratified sampling (50%)	0.6216 (0.0121)	0.6592 (0.0335)	0.6175 (0.0067)	0.5720 (0.0386)	0.6629 (0.0501)
Stratified sampling (20%)	0.5634 (0.0575)	0.6256 (0.0443)	0.5522 (0.0615)	0.4596 (0.0745)	0.6448 (0.0567)
Stratified sampling (3%)	0.5114 (0.0350)	0.5344 (0.0420)	0.5197 (0.0427)	0.4680 (0.0676)	0.5714 (0.1067)
Random sampling (50%)	0.7108 (0.0284)	0.6188 (0.0441)	0.6960 (0.0308)	0.7769 (0.0277)	0.6151 (0.0652)
Random sampling (20%)	0.7059 (0.0221)	0.5998 (0.0504)	0.6874 (0.0288)	0.7883 (0.0403)	0.5866 (0.0830)
Random sampling (3%)	0.6908 (0.0416)	0.5212 (0.0381)	0.6474 (0.0329)	0.8342 (0.0568)	0.4607 (0.0678)
<b>eGeMAPS</b>					
Complete set, CV	0.7998 (0.0047)	0.7482 (0.0055)	0.7935 (0.0046)	0.8214 (0.0051)	0.7656 (0.0048)
Complete set, grouped CV	0.6721 (0.0205)	0.5579 (0.0341)	0.6494 (0.0147)	0.7493 (0.0460)	0.5494 (0.0634)
Stratified sampling (50%)	0.5911 (0.0122)	0.6223 (0.0300)	0.5868 (0.0152)	0.5647 (0.0505)	0.6089 (0.0382)
Stratified sampling (20%)	0.5257 (0.0454)	0.5828 (0.0292)	0.5177 (0.0498)	0.4499 (0.0736)	0.5854 (0.0400)
Stratified sampling (3%)	0.5170 (0.0364)	0.5329 (0.0509)	0.5231 (0.0363)	0.4852 (0.0630)	0.5610 (0.0971)
Random sampling (50%)	0.6680 (0.0213)	0.5608 (0.0415)	0.6490 (0.0208)	0.7444 (0.0161)	0.5537 (0.0372)
Random sampling (20%)	0.6696 (0.0323)	0.5565 (0.0401)	0.6492 (0.0309)	0.7520 (0.0207)	0.5464 (0.0585)
Random sampling (3%)	0.6581 (0.0314)	0.4947 (0.0518)	0.6211 (0.0325)	0.7817 (0.0647)	0.4606 (0.0692)