remote sensing

MDPI

Supplementary Material:

Dual Activation Function based Extreme Learning Machine (ELM) for Estimating Grapevine Berry Yield and Quality

Matthew Maimaitiyiming[1], Vasit Sagan [1,*], Paheding Sidike[1] and Misha T. Kwasniewski[2]

[1] Department of Earth and Atmospheric Sciences, Saint Louis University, Saint Louis, MO 63108, USA; maimaitiyiming@slu.edu, sidike.paheding@slu.edu

[2] Grape and Wine Institute, University of Missouri, 221 Eckles Hall, Columbia, MO 65211, USA; kwasniewskim@missouri.edu

* Correspondence: Vasit.Sagan@slu.edu

Table S1 shows mean R^2 and standard deviation of prediction models in validation. To explore the effects of training data size, we used one-way analysis of variance (ANOVA) [1] on R^2 values of prediction models in validation generated with different data splits (50-50%, 60-40%, 70-30%, 80-20% and 90-10%). When ANOVA showed significant ($p<0.05$) variable effects, Tukey's honest significant difference (HSD) test was performed to further determine which data split was different from each other [2]. The results of ANOVA and Tukey's test were presented in Table S2-5. All data analyses were performed using the stats package in R [3]

Table 1. Mean R^2±standard deviation of model runs with 10 repetitions.

Berry yield and quality parameters	Data split	MLR	PLSR	RFR	WRELM	WRELM-TanhRe
Yield (kg/vine)	50-50%	0.115±0.102	0.292±0.063	0.294±0.125	0.337±0.072	0.450±0.058
	60-40%	0.163±0.117	0.280±0.090	0.308±0.073	0.286±0.125	0.290±0.058
	70-30%	0.151±0.128	0.273±0.106	0.250±0.120	0.330±0.095	0.316±0.130
	80-20%	0.187±0.104	0.470±0.085	0.622±0.067	0.614±0.049	0.680±0.034
	90-10%	0.171±0.152	0.292±0.162	0.202±0.162	0.283±0.145	0.333±0.110
TSS (°Brix)	50-50%	0.053±0.087	0.167±0.110	0.232±0.081	0.117±0.109	0.380±0.083
	60-40%	0.120±0.087	0.206±0.063	0.204±0.115	0.203±0.090	0.249±0.056
	70-30%	0.068±0.092	0.169±0.104	0.176±0.111	0.166±0.127	0.171±0.140
	80-20%	0.418±0.091	0.430±0.160	0.445±0.094	0.346±0.114	0.520±0.052
	90-10%	0.255±0.184	0.158±0.181	0.230±0.169	0.287±0.161	0.237±0.184
TA (g tartaric acid L^{-1})	50-50%	0.212±0.098	0.198±0.127	0.300±0.084	0.270±0.118	0.484±0.103
	60-40%	0.210±0.143	0.243±0.112	0.336±0.090	0.351±0.103	0.327±0.124
	70-30%	0.217±0.137	0.350±0.069	0.353±0.126	0.225±0.168	0.378±0.112
	80-20%	0.316±0.120	0.401±0.074	0.515±0.090	0.542±0.034	0.530±0.060
	90-10%	0.203±0.131	0.231±0.122	0.341±0.163	0.317±0.140	0.260±0.141
IMAD (TSS/TA)	50-50%	0.185±0.084	0.298±0.134	0.372±0.047	0.267±0.097	0.447±0.041
	60-40%	0.295±0.146	0.310±0.061	0.379±0.021	0.284±0.113	0.330±0.058

70-30%	0.197±0.091	0.332±0.086	0.389±0.071	0.340±0.087	0.316±0.130
80-20%	0.540±0.037	0.640±0.019	0.628±0.050	0.643±0.034	0.650±0.039
90-10%	0.085±0.111	0.161±0.121	0.294±0.135	0.209±0.135	0.227±0.193

Table S2. Tukey's test and ANOVA results of different data splits in MLR models.

	p-value				
Data split	50-50%	60-40%	70-30%	80-20%	90-10%
Yield(kg/vine)					
50-50%	1.000	0.905	0.963	0.087	0.840
60-40%		1.000	1.000	0.428	1.000
70-30%			1.000	0.313	0.996
80-20%				1.000	0.523
90-10%					1.000
ANOVA	F-statistic = 1.843, p-value = 0.137				
TSS (°Brix)					
50-50%	1.000	0.704	0.999	**0.001**	**0.003**
60-40%		1.000	0.855	**0.001**	0.078
70-30%			1.000	**0.001**	**0.006**
80-20%				1.000	**0.016**
90-10%					1.000
ANOVA	F-statistic = 18.488, p-value = **< 0.001**				
TA (g tartaric acid L^{-1})					
50-50%	1.000	1.000	1.000	0.146	1.000
60-40%		1.000	1.000	0.137	1.000
70-30%			1.000	0.172	0.999
80-20%				1.000	0.106
90-10%					1.000
ANOVA	F-statistic = 2.285, p-value = 0.075				
IMAD (TSS/TA)					
50-50%	1.000	0.122	0.999	**0.001**	0.191
60-40%		1.000	0.210	**0.001**	**0.001**
70-30%			1.000	**0.001**	0.110
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 31.204, p-value = **< 0.001**				

Boldface = statistically significant values with ≤ 5% level of significance

Table S3. Tukey's test and ANOVA results of different data splits in PLRS models.

	p-value				
Data split	50-50%	60-40%	70-30%	80-20%	90-10%
Yield(kg/vine)					
50-50%	1.000	0.999	0.994	**0.003**	1.000
60-40%		1.000	1.000	**0.001**	0.999
70-30%			1.000	**0.001**	0.995
80-20%				1.000	**0.003**
90-10%					1.000
ANOVA	F-statistic = 6.664, p-value = < **0.001**				
TSS (°Brix)					
50-50%	1.000	0.964	1.000	**0.001**	1.000
60-40%		1.000	0.968	**0.003**	0.924
70-30%			1.000	**0.001**	1.000

80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 8.134, p-value = < **0.001**				
TA (g tartaric acid L^{-1})					
50-50%	1.000	0.876	**0.018**	**0.001**	0.954
60-40%		1.000	0.165	**0.008**	0.999
70-30%			1.000	0.734	1.000
80-20%				1.000	**0.004**
90-10%					1.000
ANOVA	F-statistic = 7.193, p-value = < **0.001**				
IMAD (TSS/TA)					
50-50%	1.000	0.998	0.920	**0.001**	**0.018**
60-40%		1.000	0.982	**0.001**	**0.008**
70-30%			1.000	**0.001**	**0.002**
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 36.440, p-value = **< 0.001**				

Boldface = statistically significant values with ≤ 5% level of significance

Table S4. Tukey's test and ANOVA results of different data splits in RFR models.

	p-value				
Data split	50-50%	60-40%	70-30%	80-20%	90-10%
Yield(kg/vine)					
50-50%	1.000	0.999	0.908	**0.001**	0.385
60-40%		1.000	0.787	**0.001**	0.250
70-30%			1.000	**0.001**	0.881
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 21.256, p-value = < **0.001**				
TSS (°Brix)					
50-50%	1.000	0.983	0.831	**0.001**	1.000
60-40%		1.000	0.986	**0.001**	0.988
70-30%			1.000	**0.001**	0.853
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 8.921, p-value = < **0.001**				
TA (g tartaric acid L^{-1})					
50-50%	1.000	0.953	0.841	**0.001**	0.929
60-40%		1.000	0.998	**0.007**	1.000
70-30%			1.000	**0.017**	0.999
80-20%				1.000	**0.009**
90-10%					1.000
ANOVA	F-statistic = 5.652, p-value = < **0.001**				
IMAD (TSS/TA)					
50-50%	1.000	1.000	0.990	**0.001**	0.151
60-40%		1.000	0.998	**0.001**	0.104
70-30%			1.000	**0.001**	0.054
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 30.021, p-value = **< 0.001**				

Boldface = statistically significant values with ≤ 5% level of significance

Table S4. Tukey's test and ANOVA results of different data splits in WRELM models.

	p-value				
Data split	50-50%	60-40%	70-30%	80-20%	90-10%
Yield(kg/vine)					
50-50%	1.000	0.799	1.000	**0.001**	0.772
60-40%		1.000	0.867	**0.001**	1.000
70-30%			1.000	**0.001**	0.844
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 20.297, p-value = < **0.001**				
TSS (°Brix)					
50-50%	1.000	0.519	0.896	**0.001**	**0.025**
60-40%		1.000	0.960	**0.035**	0.545
70-30%			1.000	**0.006**	0.192
80-20%				1.000	0.602
90-10%					1.000
ANOVA	F-statistic = 6.592, p-value = < **0.001**				
TA (g tartaric acid L^{-1})					
50-50%	1.000	0.582	0.920	**0.001**	0.914
60-40%		1.000	0.161	**0.005**	0.969
70-30%			1.000	**0.001**	0.457
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 10.766, p-value = < **0.001**				
IMAD (TSS/TA)					
50-50%	1.000	0.997	0.605	**0.001**	0.783
60-40%		1.000	0.807	**0.001**	0.577
0.870-30%			1.000	**0.001**	0.091
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 23.023, p-value = **< 0.001**				

Boldface = statistically significant values with ≤ 5% level of significance

Table S5. Tukey's test and ANOVA results of different data splits in WRELM-TanhRe models.

	p-value				
Data split	50-50%	60-40%	70-30%	80-20%	90-10%
Yield(kg/vine)					
50-50%	1.000	**0.003**	**0.012**	**0.001**	**0.047**
60-40%		1.000	0.985	**0.001**	0.936
70-30%			1.000	**0.001**	0.985
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 32.220, p-value = < **0.001**				
TSS (°Brix)					
50-50%	1.000	0.113	**0.002**	0.057	0.070
60-40%		1.000	0.579	**0.001**	0.999
70-30%			1.000	**0.001**	0.716
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 14.422 , p-value = < **0.001**				
TA (g tartaric acid L^{-1})					
50-50%	1.000	**0.022**	0.218	0.857	**0.001**

60-40%		1.000	0.845	**0.001**	0.663
70-30%			1.000	**0.023**	0.143
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 10.203, p-value = < **0.001**				
IMAD (TSS/TA)					
50-50%	1.000	0.137	0.075	**0.001**	**0.001**
60-40%		1.000	0.999	**0.001**	0.254
70-30%			1.000	**0.001**	0.396
80-20%				1.000	**0.001**
90-10%					1.000
ANOVA	F-statistic = 22.322, p-value = **< 0.001**				

Boldface = statistically significant values with ≤ 5% level of significance

References

1. Dixon, W.J.; Massey Jr, F.J. Introduction to statistical analysis, ed 3.; McGraw-Hill Book Co: New York, 1969; Vol. 61, pp. 1l.
2. Tukey, J.W. Exploratory data analysis; Reading, Mass.: 1977; Vol. 2.
3. R Core Team R: A language and environment for statistical computing https://www.R-project.org/. Vienna, Austria: R Foundation for Statistical Computing; 2014, 2015.