

Uncovering patterns in skin and gut microbiota of rainbow trout (*Oncorhynchus mykiss*): insights from machine learning and feeding trials with sustainable aquafeeds based on yellow mealworm (*Tenebrio molitor*)

Aquaculture International

Silvio Rizzi¹, Giulio Saroglia², Violeta Kalemi¹, Simona Rimoldi¹, Genciana Terova^{1*}

¹ Department of Biotechnology and Life Sciences, University of Insubria, Via Jean Henry Dunant 3, 21100 Varese, Italy

² Medical Devices Area, Institute of Digital Technologies for Personalized Healthcare (MeDiTech), University of Applied Sciences and Arts of Southern Switzerland, Via la Santa 1, 6962 Lugano, Switzerland

* Corresponding author: genciana.terova@uninsubria.it

Supplementary Table 1

Regression model performance on rainbow trout gut microbiota at the family level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.59	0.59	0.59	0.59	-0.07	0.58	0.57	0.57	0.57	0.59	-0.18
	MAE	262.22	261.14	262.22	262.22	262.22	237.76	240.54	240.62	240.59	262.22	458.69
	RMSE	373.25	373.62	373.25	373.25	373.25	422.81	425.54	425.69	425.62	373.25	621.55
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.12	0.12	0.12	0.12	0.12	-0.23	-0.29	-0.32	-0.31	0.12	-2.31
	MAE	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85	2720.90	2723.63	2722.38	2348.78	7758.63
	RMSE	3177.56	3183.32	3177.56	3177.55	3177.56	4023.39	4065.80	4068.46	4067.22	3177.56	8523.39
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85				2348.78	7777.34
	BM R ²	0.37	0.36	0.37	0.37	0.37	-0.09				0.37	-2.51

	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24

OPT1: optimization scenario 1 - no optimization (performed on the full dataset without hyperparameter tuning). **OPT2:** optimization scenario 2 - partial optimization (performed on the reduced dataset without hyperparameter tuning). **OPT3:** optimization scenario 3 - full optimization (performed on the reduced dataset with hyperparameter tuning). **AVG NCV:** root mean squared error of the model, averaged over nested cross-validation. **BM R²:** coefficient of determination of the highest-performing model returned by nested cross-validation. **BM MAE:** mean absolute error of the highest-performing model returned by nested cross-validation. **BM RMSE:** root mean squared error of the highest-performing model returned by nested cross-validation. The acronyms of the regression algorithms are provided in the manuscript.

Supplementary Table 2

Regression model performance on rainbow trout gut microbiota at the genus level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.64	0.64	0.64	0.64	-0.06	0.64	0.63	0.63	0.63	0.64	-0.14
	MAE	193.01	192.14	193.01	193.01	193.01	173.76	175.64	175.70	175.68	193.01	335.45
	RMSE	276.32	276.51	276.32	276.32	276.32	311.82	313.72	313.82	313.78	276.32	457.08
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.12	0.12	0.12	0.12	0.12	-0.23	-0.29	-0.32	-0.31	0.12	-2.31
	MAE	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85	2720.90	2723.63	2722.38	2348.78	7758.63
	RMSE	3177.56	3183.32	3177.56	3177.55	3177.56	4023.39	4065.80	4068.46	4067.22	3177.56	8523.39
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85				2348.78	7777.34
	BM R ²	0.37	0.36	0.37	0.37	0.37	-0.09				0.37	-2.51

	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24

OPT1: optimization scenario 1 - no optimization (performed on the full dataset without hyperparameter tuning). **OPT2:** optimization scenario 2 - partial optimization (performed on the reduced dataset without hyperparameter tuning). **OPT3:** optimization scenario 3 - full optimization (performed on the reduced dataset with hyperparameter tuning). **AVG NCV:** root mean squared error of the model, averaged over nested cross-validation. **BM R²:** coefficient of determination of the highest-performing model returned by nested cross-validation. **BM MAE:** mean absolute error of the highest-performing model returned by nested cross-validation. **BM RMSE:** root mean squared error of the highest-performing model returned by nested cross-validation. The acronyms of the regression algorithms are provided in the manuscript.

Supplementary Table 3

Regression model performance on rainbow trout gut microbiota at the species level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.68	0.68	0.68	0.68	-0.06	0.69	0.68	0.68	0.68	0.68	-0.11
	MAE	152.75	152.05	152.75	152.75	152.75	137.02	138.31	138.38	138.35	152.75	264.60
	RMSE	219.25	219.37	219.25	219.25	219.25	246.89	248.19	248.28	248.24	219.25	360.96
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.12	0.12	0.12	0.12	0.12	-0.23	-0.29	-0.32	-0.31	0.12	-2.31
	MAE	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85	2720.90	2723.63	2722.38	2348.78	7758.63
	RMSE	3177.56	3183.32	3177.56	3177.55	3177.56	4023.39	4065.80	4068.46	4067.22	3177.56	8523.39
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	2348.78	2349.19	2348.78	2348.79	2348.78	2666.85				2348.78	7777.34
	BM R ²	0.37	0.36	0.37	0.37	0.37	-0.09				0.37	-2.51

	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24

OPT1: optimization scenario 1 - no optimization (performed on the full dataset without hyperparameter tuning). **OPT2:** optimization scenario 2 - partial optimization (performed on the reduced dataset without hyperparameter tuning). **OPT3:** optimization scenario 3 - full optimization (performed on the reduced dataset with hyperparameter tuning). **AVG NCV:** root mean squared error of the model, averaged over nested cross-validation. **BM R²:** coefficient of determination of the highest-performing model returned by nested cross-validation. **BM MAE:** mean absolute error of the highest-performing model returned by nested cross-validation. **BM RMSE:** root mean squared error of the highest-performing model returned by nested cross-validation. The acronyms of the regression algorithms are provided in the manuscript.