

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 skin microbiota at the family level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	-0.12	-0.11	-0.12	-0.12	-0.27	-0.15	-0.10	-0.07	-0.08	-0.12	-0.40
	MAE	110.91	109.77	110.91	110.91	110.91	103.78	103.69	103.59	103.64	110.91	143.62
	RMSE	153.76	152.82	153.76	153.76	153.76	159.52	160.00	159.94	159.97	153.76	206.02
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.17	0.15	0.17	0.17	0.17	-0.72	-1.26	-1.30	-1.28	0.17	-1.14
	MAE	380.17	395.51	380.17	380.17	380.17	675.25	785.48	791.34	788.65	380.17	584.53
	RMSE	493.96	501.63	493.96	493.96	493.96	712.05	816.42	822.09	819.49	493.96	792.95
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	380.17	395.51	380.17	380.17	380.17	675.25				380.17	610.56
	BM R ²	0.38	0.36	0.38	0.38	0.38	-0.26				0.37	-1.26

	BM MAE	316.39	314.25	316.81	316.81	316.81	465.85	332.78	611.01
	BM RMSE	428.39	432.73	428.39	428.39	428.39	608.81	429.73	816.57

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 skin microbiota at the genus level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	-0.04	-0.03	-0.04	-0.04	-0.24	-0.07	-0.04	-0.03	-0.03	-0.04	-0.31
	MAE	62.58	61.93	62.58	62.58	62.58	57.91	58.29	58.30	58.29	62.58	78.76
	RMSE	87.36	86.82	87.36	87.36	87.36	90.34	90.99	91.03	91.01	87.36	144.09
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.14	0.13	0.14	0.14	0.14	-0.05	-0.72	-0.88	-0.83	0.14	-0.55
	MAE	144.58	149.95	144.58	144.58	144.58	242.55	294.03	298.10	296.30	144.58	219.79
	RMSE	188.54	191.19	188.54	188.54	188.54	259.77	306.13	310.08	308.33	188.54	296.31
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	144.58	149.95	144.58	144.58	144.58	242.55				144.58	230.40
	BM R ²	0.37	0.37	0.37	0.37	0.37	0.14				0.36	-0.71

	BM MAE	121.67	120.66	121.67	121.67	121.67	169.33	127.17	230.72
	BM RMSE	163.48	164.94	163.48	163.48	163.48	223.70	164.00	309.48

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 skin microbiota at the species level.

OPT1		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	-0.02	-0.01	-0.02	-0.02	-0.23	-0.06	-0.03	-0.01	-0.02	-0.02	-0.28
	MAE	47.42	46.95	47.42	47.42	47.42	43.65	44.02	44.01	44.01	47.42	58.29
	RMSE	66.17	65.76	66.17	66.17	66.17	68.59	69.16	69.20	69.18	66.17	84.97
OPT2		DTR	RFR	ETR	GBR	XGBR	SVRL	SVRP	SVRR	SVRS	KNR	MLPR
	R ²	0.11	0.11	0.11	0.11	0.11	-0.21	-0.42	-0.58	-0.51	0.11	-0.26
	MAE	92.05	95.16	92.05	92.05	92.05	155.34	184.96	187.61	186.42	92.05	125.35
	RMSE	119.11	120.75	119.11	119.11	119.11	166.80	193.42	196.08	194.86	119.11	174.31
OPT3		DTR	RFR	ETR	GBR	XGBR	SVRL				KNR	MLPR
	AVG NCV	92.05	95.16	92.05	92.05	92.05	155.34				92.05	130.16
	BM R ²	0.34	0.34	0.34	0.34	0.34	0.12				0.34	-0.36

	BM MAE	77.39	76.11	77.39	77.39	77.39	105.25	80.03	130.21
	BM RMSE	103.15	104.07	103.15	103.15	103.15	141.92	103.36	182.27

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.