

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

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

OPT1		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.33	0.67	0.58	0.33	0.67	0.58	0.83	0.58	0.75	0.25	0.75	0.67	0.67	0.58
	MCC	-0.35	0.35	0.17	-0.35	0.33	0.17	0.71	0.17	0.51	-0.58	0.51	0.35	0.35	0.17
OPT2		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.58	0.67	0.67	0.58	0.67	0.58	0.83	0.50	0.58	0.58	0.58	0.58	0.50	0.58
	MCC	0.17	0.33	0.33	0.17	0.33	0.17	0.71	0.00	0.30	0.17	0.17	0.17	0.00	0.17
OPT3		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK				KNC	LOGREG	MLPC
	AVG NCV	0.33	0.67	0.67	0.58	0.67	0.58	0.83	0.50				0.58	0.42	0.58
	BM ACC	0.92	1.00	1.00	1.00	0.83	1.00	0.83	0.83				1.00	0.83	0.83
	BM MCC	0.85	1.00	1.00	1.00	0.67	1.00	0.71	0.67				1.00	0.67	0.67

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:** average model accuracy during nested cross-validation. **BM ACC:** accuracy of the highest-performing model returned by nested cross-validation. **BM MCC:** Matthews correlation coefficient of the highest-performing model returned by nested cross-validation. The acronyms of the classification algorithms are provided in the manuscript.

Supplementary Table 2

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

OPT1		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.42	0.67	0.75	0.42	0.58	0.58	0.83	0.33	0.67	0.17	0.75	0.75	0.58	0.58
	MCC	-0.17	0.33	0.51	-0.17	0.17	0.17	0.71	-0.33	0.33	-0.67	0.51	0.51	0.19	0.19
OPT2		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.58	0.58	0.67	0.58	0.58	0.50	0.75	0.50	0.42	0.42	0.67	0.75	0.75	0.75
	MCC	0.17	0.17	0.33	0.17	0.17	0.00	0.58	0.00	-0.17	-0.17	0.33	0.51	0.51	0.51
OPT3		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK				KNC	LOGREG	MLPC
	AVG NCV	0.33	0.58	0.67	0.58	0.58	0.42	0.75	0.42				0.75	0.58	0.75
	BM ACC	0.92	1.00	1.00	1.00	0.92	1.00	0.75	1.00				0.83	0.92	1.00
	BM MCC	0.85	1.00	1.00	1.00	0.85	1.00	0.58	1.00				0.67	0.85	1.00

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:** average model accuracy during nested cross-validation. **BM ACC:** accuracy of the highest-performing model returned by nested cross-validation. **BM MCC:** Matthews correlation coefficient of the highest-performing model returned by nested cross-validation. The acronyms of the classification algorithms are provided in the manuscript.

Supplementary Table 3

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

OPT1		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.33	0.67	0.75	0.42	0.50	0.58	0.83	0.50	0.67	0.08	0.75	0.75	0.50	0.50
	MCC	-0.35	0.35	0.51	-0.17	0.00	0.17	0.71	0.00	0.33	-0.85	0.51	0.51	0.00	0.00
OPT2		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	SVC-PK	SVC-RK	SVC-SK	KNC	LOGREG	MLPC
	ACC	0.50	0.67	0.67	0.58	0.58	0.67	0.75	0.50	0.50	0.58	0.67	0.58	0.67	0.67
	MCC	0.00	0.33	0.33	0.17	0.17	0.33	0.58	0.00	0.00	0.17	0.33	0.17	0.33	0.33
OPT3		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK				KNC	LOGREG	MLPC
	AVG NCV	0.58	0.67	0.58	0.50	0.58	0.58	0.75	0.50				0.58	0.67	0.67
	BM ACC	0.92	1.00	1.00	1.00	0.92	1.00	0.75	0.83				1.00	0.83	0.83
	BM MCC	0.85	1.00	1.00	1.00	0.85	1.00	0.58	0.71				1.00	0.71	0.67

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:** average model accuracy during nested cross-validation. **BM ACC:** accuracy of the highest-performing model returned by nested cross-validation. **BM MCC:** Matthews correlation coefficient of the highest-performing model returned by nested cross-validation. The acronyms of the classification algorithms are provided in the manuscript.