ST 1. Non-default parameters adapted from the tuning process of the AutoML algorithm and applied for regression and classifications models using the GBM and DL algorithms in H_2O. GBM = Gradient Boosting Machine, DL = DeepLearning.

	Classification		Regression	
Parameter	GBM	DL	GBM	DL
nfolds	10			
keep_cross_validation_models	FALSE			
keep_cross_validation_predictions	TRUE			
score_tree_interval	5		5	
fold_assignment	"Modulo"			
ntrees			72	
Max_depth	9		3	
Min_rows	100		15	
overwrite_with_best_model		FALSE		FALSE
activation		"RectifierWithDropout"		"RectifierWithDropout"
hidden		C(50,50)		C(20)
epochs		802		1840.727
seed	16	163	767	31
epsilon		1e-09		1e-09
input_dropout_ratio		0.2		0.1
hidden_dropout_ratios				0.5
distribution	“multinomial”		"gaussian"	
Sample_rate	0.7		0.5	
Col_sample_rate			0.4	
col_sample_rate_per_tree	0.4		0.7	
histogram_type	"UniformAdaptive"		"UniformAdaptive"	
stopping_rounds		0		0
stopping_metric	“logloss”	“logloss”	"deviance"	
categorical_encoding	“Enum"	"OneHotInternal"	“Enum"	"OneHotInternal"
calibration_method	"PlattScaling"		"PlattScaling"	

ST 2. Validation results for sensitivity, precision, and F1-score in percentage for DL, using Herd 1 as validation herd. DL = DeepLearning, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	6	86	60	100	10	93	36	4
2.0	52	91	64	97	57	94	297	33
2.5	60	97	45	95	51	96	297	33
3.0	44	96	36	83	40	89	190	21
3.5	7	77	17	63	10	70	71	8
4.0	0	13	0	100	0	22	16	2
4.5	50	50	50	50	50	50	4	0
WAvg	47	91	47	91	45	90	911	

ST 3. Validation results for sensitivity, precision, and F1-score in percentage for DL, using Herd 2 as validation herd. DL = DeepLearning, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	0	40	0	0	0	0	10	2
2.0	30	100	74	100	43	100	106	23
2.5	89	99	48	94	63	97	154	33
3.0	33	100	45	88	38	94	105	22
3.5	43	85	41	87	42	86	60	13
4.0	8	71	100	100	15	83	24	5
4.5	0	0	0	0	0	0	10	2
WAvg	49	93	53	90	45	91	469	

ST 4. Validation results for sensitivity, precision, and F1-score in percentage for DL, using Herd 3 as validation herd. DL = DeepLearning, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	0	89	0	0	0	0	18	2
2.0	35	97	50	91	41	94	152	17
2.5	66	91	40	90	50	90	256	29
3.0	21	96	30	88	24	92	209	24
3.5	45	76	34	80	39	78	151	17
4.0	0	74	0	100	0	85	65	7
4.5	9	14	0	0	0	0	22	3
WAvg	38	87	33	84	34	85	873	

ST 5. Validation results for sensitivity, precision, and F1-score in percentage for GBM, using Herd 1 as validation herd. GBM = Gradient Boosting Machine, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	8	75	50	88	14	81	36	4
2.0	51	80	64	97	57	88	297	33
2.5	35	92	45	93	39	93	297	33
3.0	56	91	35	82	43	86	190	21
3.5	42	87	28	74	34	80	71	8
4.0	25	94	17	54	20	69	16	2
4.5	0	50	0	25	0	33	4	0
WAvg	44	87	47	90	44	88	911	

ST 6. Validation results for sensitivity, precision, and F1-score in percentage for GBM, using Herd 2 as validation herd. GBM = Gradient Boosting Machine, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	0	70	50	100	0	82	10	2
2.0	58	97	60	96	59	97	106	23
2.5	64	98	48	94	55	96	154	33
3.0	31	95	35	89	33	92	105	22
3.5	33	85	38	90	36	88	60	13
4.0	13	79	27	91	17	85	24	5
4.5	0	30	0	50	0	38	10	2
WAvg	46	93	45	92	44	92	469	

ST 7. Validation results for sensitivity, precision, and F1-score in percentage for GBM, using Herd 3 as validation herd. GBM = Gradient Boosting Machine, BCS = Body Condition Score, Exact = exact score, DV = deviation, WAvg = weighted average by frequency.

	Sensitivity, %		Precision, %		F1-Score, %		FREQ	%
BCS	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**	**Exact**	**0.5 DV**		
1.5	6	94	0	100	0	97	18	2
2.0	57	94	44	83	49	88	152	17
2.5	50	91	40	86	45	88	256	29
3.0	29	84	28	81	28	82	209	24
3.5	24	73	28	72	26	72	151	17
4.0	2	46	10	40	3	43	65	7
4.5	9	18	50	100	15	31	22	3
WAvg	36	81	33	79	33	79	873	

ST 8. Accuracy in percentage of the classification models using different combinations predictors for DL and GBM from the 7:3 random split. Exact scores are presented, and 0.5-unit DEV are in parentheses. DL= DeepLearning, GBM= Gradient Boosting Machine, FEAT=All features used for prediction, FIX= all fixed effects used for prediction, FEAT_rFIX= All features and fixed effect of evaluator and round of evaluation used for prediction.

	DL, %			GBM, %		
	FEAT	**FIX**	**FEAT_rFIX**	**FEAT**	**FIX**	**FEAT_rFIX**
Mean	40.6(85.6)	35.2(80.1)	48.9(93.8)	41.1(87.7)	32.7(78.5)	45.3(90.7)

ST 9. F1-score in percentage of the classification models using different combinations predictors for DL and GBM from the 7:3 random split. Exact scores are presented, and 0.5-unit DEV are in parentheses. DL= DeepLearning, GBM= Gradient Boosting Machine, FEAT=All features used for prediction, FIX= all fixed effects used for prediction, FEAT_rFIX=All features and fixed effect of evaluator and round of evaluation used for prediction.

	DL, %			GBM, %		
BCS	**FEAT**	**FIX**	**FEAT_rFIX**	**FEAT**	**FIX**	**FEAT_rFIX**
1.5	5(36)	0(0)	0(9)	0(38)	0(0)	5(70)
2.0	50(95)	40(88)	56(99)	55(92)	39(82)	60(95)
2.5	48(89)	43(87)	58(96)	45(92)	39(88)	49(94)
3.0	26(86)	21(85)	38(95)	32(87)	27(81)	37(90)
3.5	28(64)	17(43)	41(86)	28(77)	14(53)	33(81)
4.0	2(23)	0(0)	9(60)	20(67)	0(22)	15(73)
4.5	2(7)	0(0)	9(22)	12(44)	0(0)	13(49)
WAvg	37(81)	30(74)	46(90)	40(86)	30(74)	44(90)

ST 10. Validation results from regression models using the 7:3 random split. Accuracy results are presented in percentage and exact scores are presented, and 0.5-unit DEV are in parentheses. R^2 = R-square, RMSE = Root Mean Square Error, Exact = exact score, DEV = deviation, WAvg = weighted average by frequency, PLS = Partial Least Square, DL = DeepLearning, GBM = Gradient Boosting Machine, BCS = Body Condition Score, FEAT=All features used for prediction, FIX= all fixed effects used for prediction, FEAT_rFIX=All features and fixed effect of evaluator and round of evaluation used for prediction.

	PLS			DL			GBM		
BCS	**FEAT**	**FIX**	**FEAT_rFIX**	**FEAT**	**FIX**	**FEAT_rFIX**	**FEAT**	**FIX**	**FEAT_rFIX**
1.5	15(80)	0(20)	30(89)	16(83)	0(10)	19(89)	0(81)	0(6)	0(84)
2.0	42(92)	13(79)	49(97)	44(94)	9(82)	51(98)	43(93)	3(83)	53(97)
2.5	55(96)	56(99)	60(98)	60(96)	63(98)	68(98)	57(96)	69(100)	62(98)
3.0	45(96)	48(99)	55(98)	40(96)	47(99)	51(98)	41(96)	47(99)	49(97)
3.5	36(88)	10(64)	46(94)	33(84)	9(67)	43(92)	34(83)	3(63)	43(90)
4.0	19(71)	0(15)	22(88)	19(68)	0(12)	25(76)	5(65)	0(6)	16(78)
4.5	8(33)	0(0)	11(62)	8(41)	0(0)	16(65)	0(6)	0(0)	0(31)
WAvg	44(92)	33(82)	51(96)	44(92)	34(83)	53(96)	43(91)	34(83)	50(94)
R^2	0.55	0.19	0.67	0.54	0.19	0.67	0.51	0.18	0.63
RMSE	0.33	0.26	0.31	0.33	0.24	0.29	0.31	0.20	0.29