NATHALIA FARIAS DE SOUZA

BODY WEIGHT PREDICTION OF CROSSBRED BEEF CATTLE TROUGH THE IMAGE PROCESSING AND MACHINE LEARNING ALGORITHMS

Dissertation submitted to the Animal Science Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae.*

Adviser: Mario Luiz Chizzotti

VIÇOSA – MINAS GERAIS
2022

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

Souza, Nathalia Farias de, 1997-

S729b Body weight prediction of crossbred beef cattle trough the
2022 image processing and machine learning algorithms / Nathalia Farias de Souza. – Viçosa, MG, 2022.

1 dissertação eletrônica (43 f.): il.

Texto em inglês.

Inclui apêndice.

Orientador: Mário Luiz Chizzotti.

Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Zootecnia, 2022.

Referências bibliográficas: f. 24-27.

DOI: https://doi.org/10.47328/ufvbbt.2022.729

Modo de acesso: World Wide Web.

1. Bovinos de corte - Pesos e medidas. 2. Bovinos de corte - Sensoriamento remoto. 3. Aprendizado do computador. I. Chizzotti, Mário Luiz, 1980-. II. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. III. Título.

CDD 22. ed. 636.213

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

NATHALIA FARIAS DE SOUZA

BODY WEIGHT PREDICTION OF CROSSBRED BEEF CATTLE TROUGH THE IMAGE PROCESSING AND MACHINE LEARNING ALGORITHMS

Dissertation submitted to the Animal Science Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: August 26, 2022

Assent:

Nathalia Farias de Souza
Author

Mario Luiz Chizzotti
Adviser

To my great-grandmother, a surreally strong woman until the end of her life, I will never forget our moments, rest in peace.

ACKNOWLEDGMENTS

First of all, I am thankful to my family (Gleyce, Adriano, Adriano Jr, Carmem and Maria de Lourdes) for all the support during my entire life, this paper is much more yours than mine.
Also, I couldn’t thank enough my auntie Sherly Ramos for all the conversations and for giving me the love of my life my baby cousin Samuel.
To my dear friend Jonas Carneiro, for believing in and encouraging me to try this program selection in 2020, I am 100% sure that I would not be here without you.
To my soul sister (Mannu/Carol) and her family for letting me be part of their lives.
To the friends who become family (Amanda, Cris Luana, Isabela, Jenifer, Pamela, and Rizielly), this could be hardest without our happy moments, I am thankful to have you all in my life, was an honor to work with each one of you.
To the friends that I made during these 2 years (Aline, Ronaldo, and Samires).
To Ms. Érica Schultz for the support in this paper construction, share your knowledge and passion for livestock precision science, and above all for the care that you had with me during these couple of months, you are an inspiration for all young researchers.
To my mentor, Dr. Mario Luiz Chizzotti, for all the time and support during the execution of this experiment, my entire gratitude and admiration, hope that maybe one day become close to the professional that you are.
To all LCC and GEZP team for support (especially Maria and Lilian).
To friends that I probably forgot.
To my pets for being with me every day.
To each animal used in this experiment.
To the Federal University of Viçosa, especially to the Department of Animal Science, for the opportunity to accomplish my master’s degree.
To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), to granting the scholarship.

ABSTRACT

SOUZA, Nathalia Farias de, M.Sc., Universidade Federal de Viçosa, August, 2022. **Body weight prediction of crossbred beef cattle through the image processing and machine learning algorithms**. Adviser: Mario Luiz Chizzotti.

For the beef cattle system, one of the most valuable information is the body weight that can be linked to animal growth and performance. The bidimensional sensors area is the cheapest technology among all sensors used as a tool to extract information that can be applied to machine learning to predicts value phenotype. This study aimed to predict body weight using video image analysis with simple bidimensional equipment, from the dorsal view of crossbreed beef cattle (½ Angus x ½ Nellore) in a finishing system, applying different frame information and machine learning algorithms. The experimental procedures were performed at Federal University of Viçosa. A total of 40 crossbreed steers (½ Angus x ½ Nellore) were used, averaging 8 months of age at the beginning of the feedlot trial, and 291.7±23.8 kg and 517.42±54.8kg of initial and final body weight, respectively. The data collection occurred from September (12 Month) to December/2021 (15 month). Body weight (BW) was collected using an automatic Intergado company drink fountain/scale and the video images were collected using cameras Intelbras from the animals' dorsal view. Three approaches were tried for segmentation of the animals' dorsal images, however, their color characteristics did not allow do this automatically, so were used ImageJ software to manually do the delimitations, extracting 8 Shape Descriptors. For regression were used 6 machine learning algorithms, Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET), Multiple Linear Regression (MLR), Adaboost (ADAB) and Random Forest (RF) to construct predictive models, the dataset was split in 70:30 for training and test. The regularizations RIDGE and MLR without AGE as a predictor had similar performance. The AGE addition improved all algorithms, the best metrics results were for ENET and ENET using AGE for a dataset with 5 Frames information (5F) R^2=0.68 and 0.76, respectively. Thus, the use of bidimensional sensor in the dorsal view can predict the BW of crossbreed (½ Angus x ½ Nellore).

Keywords: Correlation. Image Processing Regularization

RESUMO

SOUZA, Nathalia Farias de, M.Sc., Universidade Federal de Viçosa, August, 2022. **Predição do peso corporal de bovinos cruzados a partir de processamento de imagens e algoritmos de aprendizado de máquina**. Adviser: Mario Luiz Chizzotti.

Para a bovinocultura de corte uma das informações mais valiosas é o peso corporal que está relacionado ao crescimento e desempenho animal. Sensores bidimensionais são as tecnologias mais baratas entre todos os sensores, e podem ser usados como uma ferramenta para extrair informações que associadas com o aprendizado de máquina predizem fenótipos de valor. Objetivou-se predizer o peso corporal (PC) utilizando análise de imagens de vídeo e equipamento bidimensional simples, a partir da visão dorsal de bovinos mestiços (½ angus x ½ nelore) em fase de terminação, aplicando diferentes informações de frames (1, 3 e 5), e algoritmo de aprendizado de máquina. Os procedimentos experimentais foram realizados na Universidade Federal de Viçosa, foram utilizados 40 animais mestiços (½ angus x ½ nelore), machos, com idade de 8 meses no início do confinamento e peso corporal inicial médio de 291,7±23,8 kg e finalizado com 517,42±54,8 kg. A coleta de dados ocorreu de setembro (12 meses) a dezembro/2021 (15 meses). O peso corporal (PC) foi coletado por meio de bebedouro/balança automática da empresa Intergado e as imagens de vídeo foram coletadas por meio de câmeras Intelbras a partir da visão dorsal do animal. Foram testadas três abordagens para segmentação dos animais com base na limiarização, porém, suas características de cor não permitiam fazer isso automaticamente, então foi utilizado o software ImageJ 1.35k, para fazer as delimitações manualmente, extraindo 8 Descritores de Forma. Para regressão foram avaliados seis algoritmos de aprendizado de máquina, Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET), Multiple Linear Regression (MLR), Adaboost (ADAB) e Random Forest (RF) para construir modelos preditivos, conjunto de dados foi dividido em 70:30 para treino e teste. As regularizações RIDGE e MLR sem AGE como preditor tiveram desempenho semelhante. A adição de AGE melhorou todos os algoritmos, os melhores resultados de métricas foram para ENET e ENET usando AGE para dataset com informação de 5 Frames (5F), R^2=0,68 e 0,76, respectivamente. Assim, o uso de sensor bidimensional na vista dorsal pode predizer o PC de mestiços (½ angus x ½ nelore).

Palavra-Chave: Correlação. Processamento de Imagem. Regularização.

LIST OF ACRONYMS AND ABBREVIATIONS

1F	1 Frame dataset
1F _A	1 Frame dataset plus AGE information
3F	3 Frames dataset
3F_A	3 Frames dataset plus AGE information
5F	5 Frames dataset
5F_A	5 Frames dataset plus AGE information
A	Area
ADAB	Adaboost
BW	Body Weight
ENET	Elastic Net
Fer	Feret
L	Length
LASSO	Least Absolute Shrinkage and Selection Operator
Ma	Major
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MFer	MinFeret
Mi	Minor
Per	Perimeter
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Square Error
W	Width

SUMMARY

1 Introduction

For the beef cattle system, one of the most valuable information is the body weight (BW) once it allows inferring the animal productivity and helps manage the entire process. This variable is related to the animal growth and performance (GOMES et al. 2016; ZHANG et al. 2019).

Normally to obtain this the body weight (BW) information 3 aspects within the system need to be considered. First, time and employees demand once is necessary more people to handle the animals to the scale for weight the cattle, also the number of animals will infer in the time to do it. Second, the cattle welfare with some studies shows that an unusual handle can cause some stress in the animals and affect their feeding and, consequentially, their performance. Third, the scale and the structure to traditionally weighting the animals can cost elevated values. Must be considering that the usual method for body weight is unviable active to perform every day, so in the farm this information is obtain each month (RUCHAY et al. 2020, RUCHAY et al. 2022).

The use of images in livestock farming and beef cattle production is already a reality, for example, images have been used for years in carcasses evaluation (CRAIGE et al. 2012; QIAO et al. 2021) and became more popular in animal science in recent years because of the easier access to these technologies by the population.

Bidimensional video images are the simplest ones to work due to cheap equipment such as cameras or smartphones. Also, for data collection is not necessary animal handling, which avoids animal stress and reduction in performance (MENESATTI et al. 2014; GJERGJI et al. 2020; WEBER et al. 2020), and for the producer, it requires less labor, with a shorter data and time collection interval. On the other hand, data analyses require a lot of computational work, which consists of three steps, collection of frames from videos, segmentation of the region of interest, and extraction of features based on pixel information. When associated with the features extracted from the images, the machine learning algorithm allows for predicting quantitative variables such as body weight, hot and cold carcass weight, for example (ZHANG et al. 2019; ARAÚJO et al. 2019). The most popular algorithms for prediction are regularizations: Least Absolute Shrinkage and Selection Operator, Ridge and Elastic Net that apply L1 and L2 penalization (GARCIA-NETO et al. 2021), and the ensembles: Adaboost and Random Forest (NANTRA et al. 2022).

Thus, this study aimed to predict body weight through video image analysis using simple bidimensional equipment, from the dorsal view of crossbreed beef cattle (½ Angus x

½ Nellore) in a finishing feedlot, applying different frames information, and machine learning algorithm.

2 Materials and Methods

2.1 General Information

The experimental procedures were approved by the Federal University of Viçosa, Committee of Ethics on Animal (CEUAP/UFV protocol nº. 023/2021). In the experiment 40 crossbreed steers (½ Angus x ½ Nellore) were used, averaging 8 months of age at the beginning of the trial, 291.7±23.8 kg and 517.42±54.8kg of initial and final body weight, respectively. The data collection occurred from September (12 Month) to December/2021 (15 Month)

2.2 Data Collection

2.2.1 Obtaining the videos and body weights

Before the animals the trial started, cameras Intelbras (VHD 3130 B G6) were placed at strategic points above Intergado drink fountain system to capture the dorsal view (Fig. 1). The animals were in four collective pens (with 10 animals/each). The animal identification and BW were collected from the Intergado software, which collected data through a radio frequency identification (RFID) antenna and automatic weighing scale placed in each pen. The videos were related to the identification and weight at the same hour. For the next analysis the Intel Core i7 8th Gen from DELL were used.

.

A B

Fig 1. Intergado drinking fountain system plus, intelbras camera for data collection (A) and dorsal view obtained (B)

2.2.2 Videos processing and image collection

The video segmentations were done by adaptation of a Python Code using Visual Studio Code software 1.69.2*,* the libraries required were *Datetime* (function *timedelta*), CV2, NumPy, and OS. Each video was collected 1, 3, and 5 Frames.

Frames are a sequence of images that construct the videos, through these images is possible to extract measures named shape descriptors, which were considered the independent variables in this study.

2.2.3 Frame processing

First approach (Rstudio)

RStudio version 1.4.1717 (R Core team, 2021) were used to perform a threshold segmentation using images in greyscale for contour detection, the packages image version 0.42.13 (Wickham et al. 2022) and dyplyr version 1.0.10 (Barthelme et al. 2022) were used.

Second approach (ImageJ)

The images were converted to 8-bit format (grayscale) to use the Threshold tool as a path to extract the Region of Interest (ROI) and shape descriptors automatically, the threshold interval was defined for each animal based on their dorsal color characteristics.

Third approach (ImageJ)

The last approach was a semi-automatic region of interest (ROI) segmentation done at

ImageJ software, the measures required were Area, Perimeter, Width, Length, Major, Minor, Feret diameter, and MinFeret. Then, using the tool polygon, boundaries in the animal's dorsal were manually performed. Thus, to minimize human errors, manual delimitation of frames was proposed to use averages of different numbers of frames (Total of 1,075 frames) information to compose the predictive datasets.

2.2.4 Shape Descriptors

The information extracted from ROI generated 8 shape descriptors (LEIBRANDT; PENNEC, 2015) that were based on mathematical equations in which pixels were applied, each information was described (Table 1).

Table 1. Features extracted from image with abbreviation and description.

Feature	Abbreviation	Description
Area (pixel²)	A	Sum of pixels inside the delimitation
Perimeter (pixel)	Per	The length of the outside boundary of the selection.
Width (pixel)	W	Width based on boundary rectangle
Length (pixel)	L	Length based on boundary rectangle
Major(pixel)	Ma	Length based on ellipse form
Minor(pixel)	Mi	Width based on ellipse form
Feret Diameter (pixel)	Fer	The longest distance between any two points along the selection boundary
MinFeret (pixel)	MFer	The shorter distance between any two points along the selection boundary

2.3 Statistical Analysis

The first step was to verify the datasets normality and outlier, shape descriptors that used pixels threshold information and coordinates were removed. All 3 datasets had each frame chosen by individual positive slope and $R^2 > 0.60$.

The Pearson correlation was performed for the features extracted and the BW with significance level $p < 0.05$ (DANCEY et al., 2006), processed in the Rstudio 1.4.1717 (R

Core team, 2021).

The final dataset was composed of 122 samples, meaning that 1 Frame (1F) dataset had information on BW, AGE, and shape descriptors, the 3 Frames (3F) dataset used mean of 326 frames to compose the 122-dataset sample, at least the mean of 529 frames information created the 5 Frames (5F) dataset. All datasets were performed first without AGE as the independent variable and then using this information for prediction.

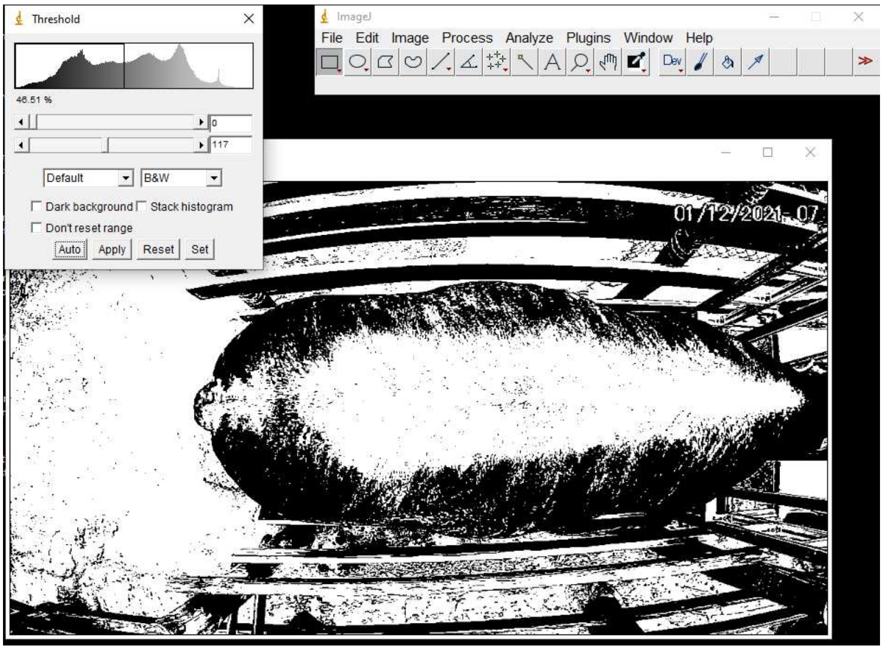
This study adopted machine learning algorithms: Ridge, Shrinkage Selection Absolute Operator (LASSO), Elastic Net (ENET), Multiple Linear Regression (MLR), Adaboost (ADAB), and Random Forest (RF) to construct predictive models. The datasets were submitted to normalization for scale of independent variables in the first four algorithms, then the observations were split into 70% for training and 30% test and standardized 5 K-fold for cross-validation were considered.

These analyses were performed using RStudio version 1.4.1717 (R Core team, 2021), Ridge, LASSO, ENET, ADAB, and RF were used the packages Caret, Glmnet, gbm, and randomForest. These packages allow the models to train (caret), fits the regularization model paths for regression, and execute the models. The models evaluation metrics: Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were extracted. Besides the observed and predictions values were plotted.

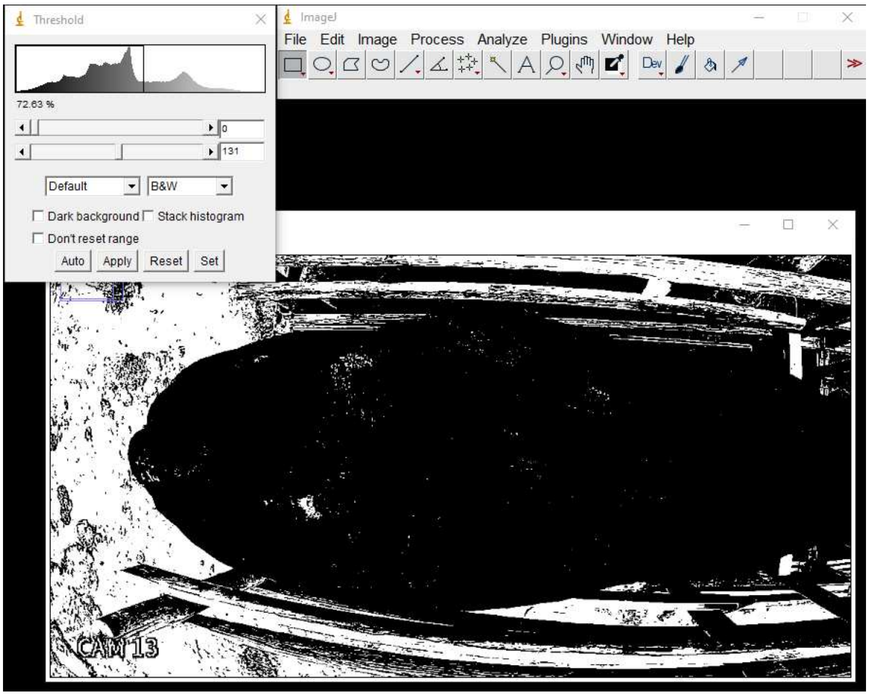
3 Results and Discussion

3.1 Image segmentation

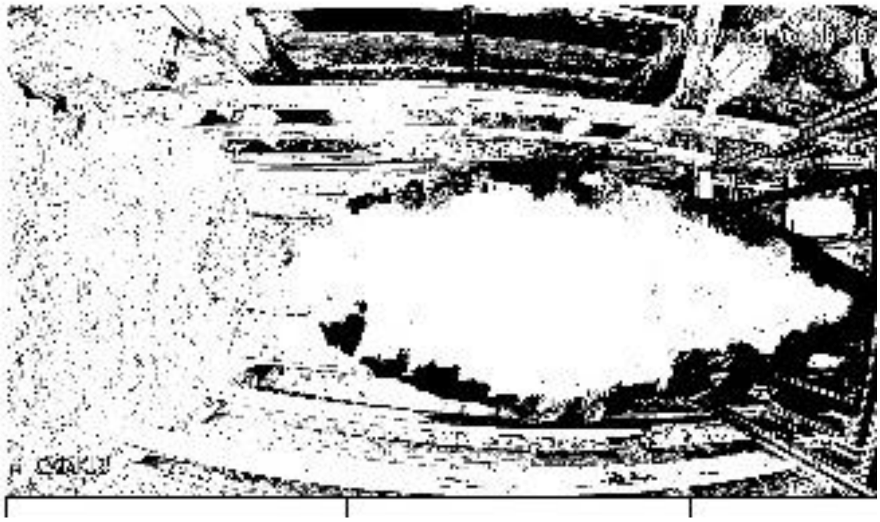
The automatic segmentation in both approaches was tried, nevertheless, animals' dorsal characteristics become unviable to detect the boundaries once pixels from the animal and the ground mixed, causing an overestimation or underestimation of shape descriptors. The environment is also a challenge, once there are changes in light or weather through the days and pens characteristics that influences the segmentation (Fig. 2).


Threshold
46.51 %
0
117
Default
B&W
Dark background
Stack histogram
Don't reset range
Auto
Apply
Reset
Set
ImageJ
File
Edit
Image
Process
Analyze
Plugins
Window
Help
01/12/2021 07

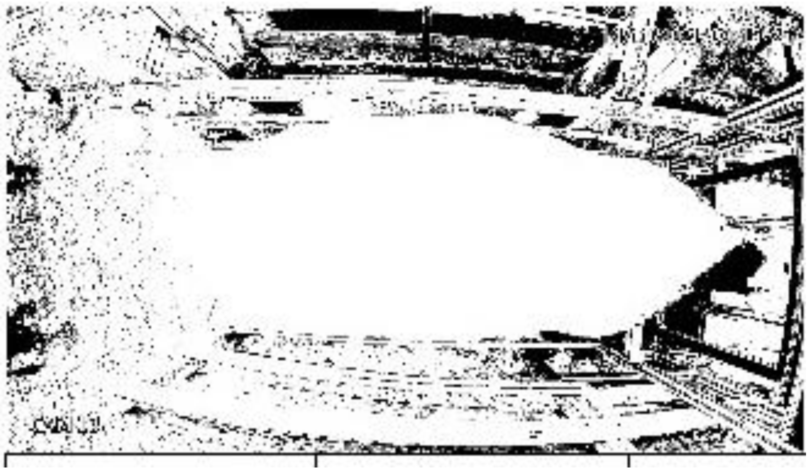
A


Threshold
72.63 %
0
131
Default
B&W
Dark background
Stack histogram
Don't reset range
Auto
Apply
Reset
Set
ImageJ
File
Edit
Image
Process
Analyze
Plugins
Window
Help
CAM 13

B



C

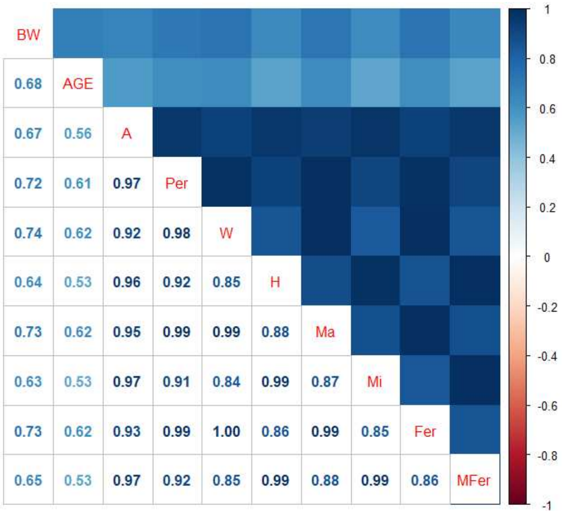


D

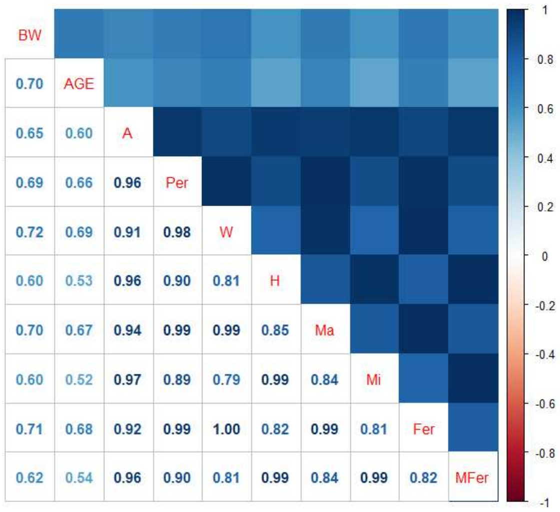
Fig 2. Segmentation using 8-bit images in crossbreed animal (½ Angus ½ Nellore), mixed fur color animal (A), regular color animal (B) at ImageJ program; mixed fur color animal (C), regular color animal (D) at Rstudio software.

The animals' dorsal boundaries were manually draw as the alternative path to the image segmentation. To extract the shape descriptors, ImageJ software just demands the delimitation, and it calculates all shape descriptors.

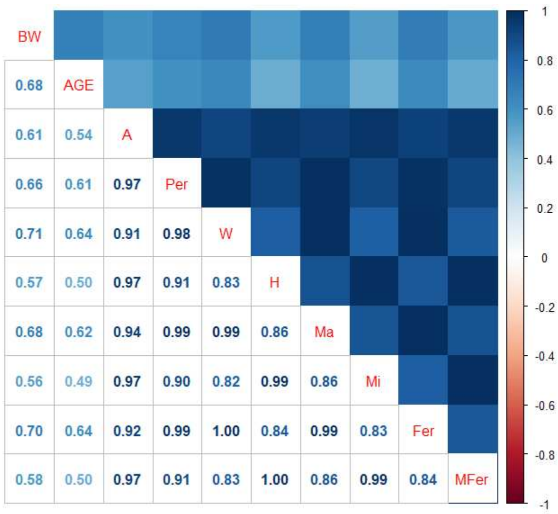
3.2 Correlations



1F



3F



5F

Fig 3. The correlation between the dependent variable (BW) and the independent variables for body weight was obtained through bidimensional images for each dataset. 1F: 1 Frame Information, 3F: 3 Frames Information, 5F: 5 Frames Information.

Moderate $|r>0.4|$ and strong $|r>0.7|$ correlations were observed in this study between the BW to the shape descriptors extracted from bidimensional images (Fig. 3). However, among the measures the correlation was very strong correlation $|r>0.89|$, this value means the occurrence of multicollinearity, which may be reflected in final regression, once this factor cause distortion and affect the models' predictions (YAKABU, 2010). The correlation between BW to A and lengths and widths were like those found by Gomes et al. (2016) evaluating Angus breed, for similar characteristics obtained by image in their study, seems, that the phenotype of crossbreed animals (½ Angus x ½ Nellore) is closest to Angus than the

Nellore measures collected. The descriptors chosen were based on notorious relationships with quantitative carcass characteristics and the animals' body weight (GOMES et al. 2016; ARAÚJO et al. 2019; ZHANG et al. 2019). Besides that, mass is mathematically more closely related to the object shape (GJERGJI et al. 2020).

3.3 Machine learning algorithm

The metrics for each algorithm were plotted in Table 2, models without AGE showed similar values of R^2 in dataset 1F less RF. To 3F and 5F datasets, LASSO and ENET had the highest coefficient of determination.

According to Hair et al. (2011) and Chicco et al. (2021), moderate and good values of predictions are $R^2 > 0.40$ and $R^2 > 0.75$, respectively, which means that all the models had at least a moderate performance.

A recent study that discusses the interpretation of evaluation metrics obtained in machine learning algorithms demonstrated that R^2 explained the quality better than even MAE and RMSE (CHICCO et al. 2021). On the other hand, Fernandes et al. (2020) reported that R^2 by itself does not measurements prediction accuracy once it does not take prediction bias into account, so using MAE or RMSE is necessary. So, this study adopted R^2, MAE. and RMSE to choose the best model.

Considering these metrics information, the best model without AGE as the independent variable (Table 2) was 5F dataset using ENET regularization that ranged R^2 = 0.68, MAE =30.70, and RMSE = 36.0. Datasets and models applying AGE as a predictor (Table 3), LASSO_A and ENET_A showed better results being R^2= 0.75 and 0.76, MAE=25.62 and 25.45, RMSE=31.11 and 30.96, respectively.

Table 2. Metrics for each model and dataset for BW prediction of crossbreed cattle using only shape descriptors as independent variable.

DATASET	METRIC	RIDGE	LASSO	ENET	MLR	ADAB	RF
	R^2	0.55	0.52	0.51	0.55	0.52	0.40
1F	MAE	36.47	35.99	36.20	33.69	35.84	34.44
	RMSE	44.15	44.92	45.45	40.41	43.83	44.72
	R^2	0.51	0.60	0.62	0.62	0.53	0.68
3 F	MAE	37.83	34.53	34.14	31.06	35.81	30.26
	RMSE	44.90	41.33	41.01	37.59	41.95	35.88
	R^2	0.54	0.65	0.68	0.61	0.50	0.48
5 F	MAE	35.13	31.62	30.72	33.73	34.34	34.36
	RMSE	40.75	36.00	34.98	41.14	42.27	41.50

1F: dataset using 1 Frame information; 3F: dataset using 3 Frames information; 5F: dataset using 5 Frames information; RIDGE: Ridge regularization; LASSO: Least Absolute Shrinkage Select Operator; ENET: Elastic Net; MLR: Multiple Linear Regression; ADAB: Adaboost; RF: Random Forest; R^2: Determinant Coefficient; MAE: Mean Absolute Error in Kg; RMSE: Root Mean Square in Kg.

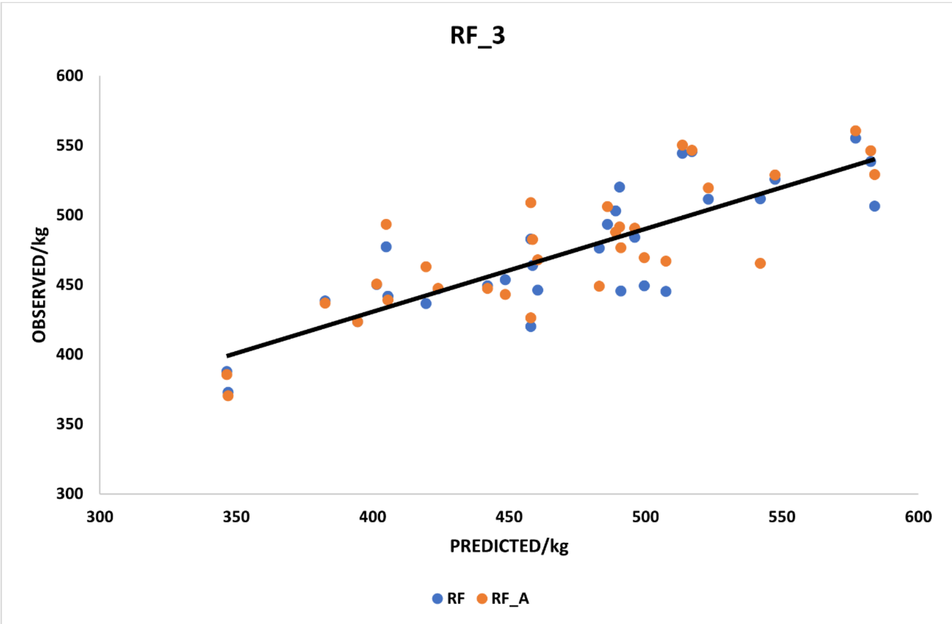
Table 3. Metrics for each model and dataset for BW prediction of crossbreed cattle using shape descriptors and AGE as independent variable.

DATASET	METRIC	RIDGE	LASSO	ENET	MLR	ADAB	RF
	R^2	0.64	0.62	0.62	0.59	0.62	0.47
1F_A	MAE	32.51	31.61	31.69	29.87	31.10	34.81
	RMSE	39.62	40.18	40.04	37.13	39.05	42.48
	R^2	0.62	0.69	0.69	0.67	0.66	0.60
3F_A	MAE	32.45	31.43	31.47	31.14	29.92	28.25
	RMSE	39.90	37.09	37.15	37.35	36.50	33.7
	R^2	0.66	0.75	0.76	0.68	0.46	0.50
5F_A	MAE	29.11	25.62	25.45	33.77	36.83	34.06
	RMSE	35.63	31.11	30.96	42.01	42.67	40.81

1F: dataset using 1 Frame information; 3F: dataset using 3 Frames information; 5F: dataset using 5 Frames information; 1F_A: dataset using 1 Frame information plus AGE; 3F_A: dataset using 3 Frames information plus AGE; 5F_A: dataset using 5 Frames information plus AGE; RIDGE: Ridge regularization; LASSO: Least Absolute Shrinkage Select Operator; ENET: Elastic Net; MLR: Multiple Linear Regression; ADAB: Adaboost; RF: Random Forest; R^2: Determinant Coefficient; MAE: Mean Absolute Error in Kg; RMSE: Root Mean Square in Kg.

The ensemble group did not demand normalization of the dataset, once that normalization is a preprocessing step that uses tree-based approach, like RF and ADAB, do not require, even so, other MLs based on matrix distance as the regularizations and MLR are necessary (VIVARACHO-PASCUAL et al. 2016).

Based on this information RF had the best result using only shape descriptors (R^2= 0.68, MAE= 30.26 and RMSE= 35.88) for dataset 3F (Table 2 and Fig 4), showing that AGE information improved metrics, by elevating R^2 and reducing the errors, for all models but for the ensemble group it was not enough to equate them to regularization best models. Denoting that these results were lower than the best performance in regularization without and using AGE as ENET and ENET_A (Table 2 and 3), respectively. Khan et al. (2014) findings showed RF presented inferior results for sheep BW predictions, compared to other algorithms such as MLR, for example. Demonstrating that in our study regularization approach was more effective to explain BW.



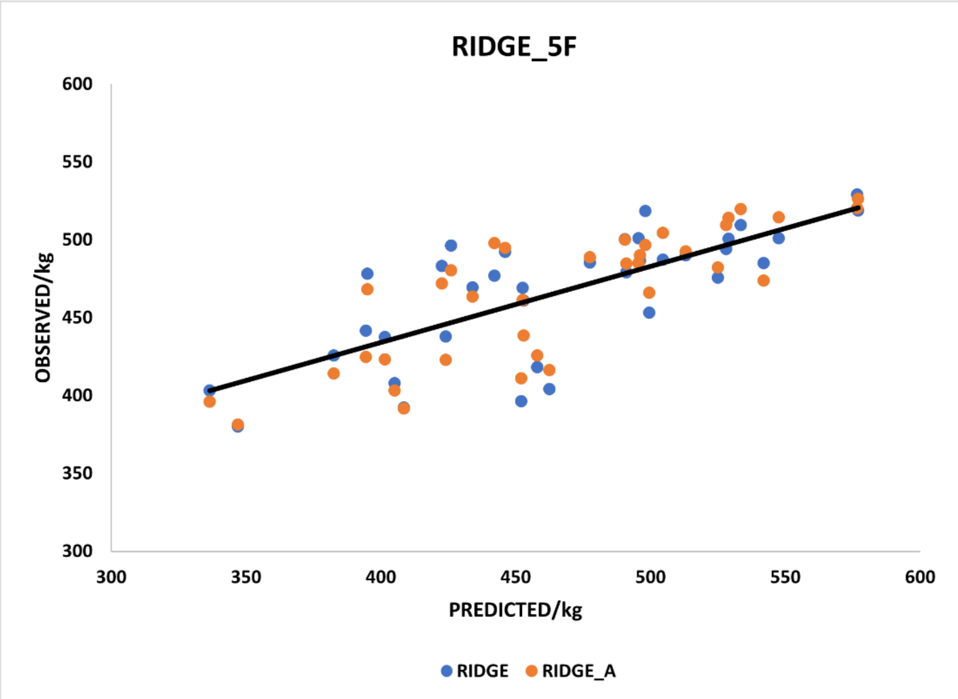
RF
R^2=0.68
MAE=30.26
RMSE= 35.88

RF_A
R^2=0.60
MAE=28.25
RMSE= 33.70

Fig 4. Comparison of the measured and estimated value of body weight for validation dataset of 5F dataset for RF with and without AGE as independent variable. RF: Random Forest; RF_A: Random Forest using AGE.

For the regularizations, RIDGE (Fig 5) had the lowest performance in each dataset and even with AGE inclusion. The normalization of datasets allowed this algorithm to keep the same lambda value (Appendix Table 4) for all processes, meaning that values above this could negatively affect the prediction model accuracy. These values were close to the findings by Çankaya et al. (2019) for body weight prediction in multicollinearity datasets.

According to Garcia-Neto et al. (2021), RIDGE improves the prediction error by shrinking large regression coefficients to reduce overfitting, but it does not perform descriptor selection, which coincides with the final regressions generated by RIDGE when compared to other regularizations. Therefore, RIDGE final equations do not help to make the model more interpretable, but we can assume that W and AGE had more contribution in the final equation, for RIDGE and RIDGE_A, respectively (Appendix Table 5).



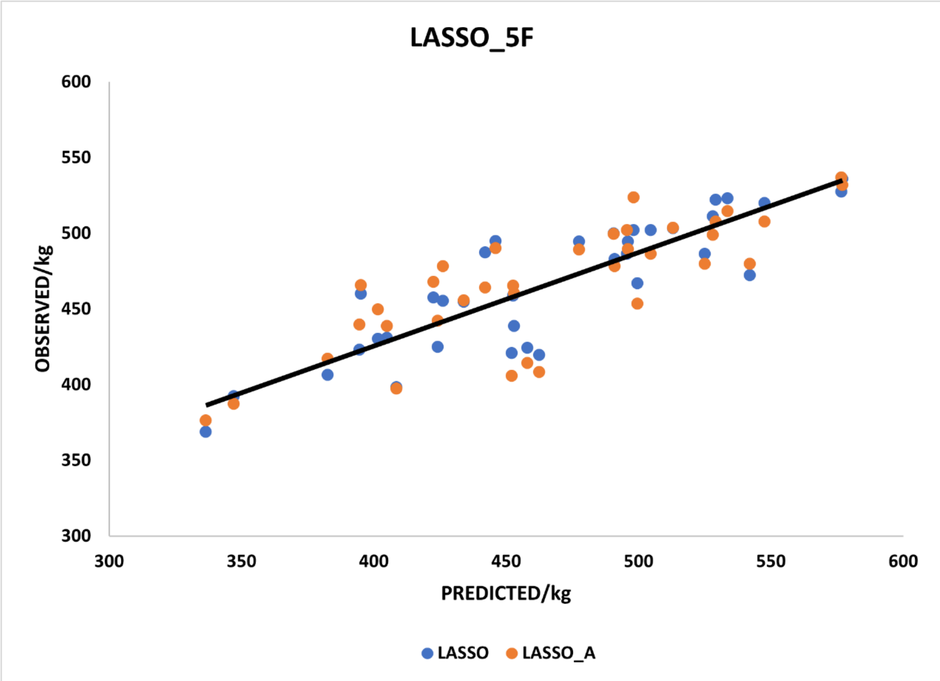
RIDGE
R^2=0.54
MAE=35.13
RMSE= 40.75

RIDGE_A
R^2=0.66
MAE=29.11
RMSE= 35.63

Fig 5 . Comparison of the measured and estimated value of body weight for validation dataset of 5F dataset for RIDGE with and without AGE as independent variable. RIDGE_A: Ridge using AGE.

MLR best performance was with 3F dataset and it had results close to 3F dataset for LASSO regularization. The difference between these models is the independent variable elimination (Appendix Table 5), that LASSO performs (Garcia-Neto et al. 2021).

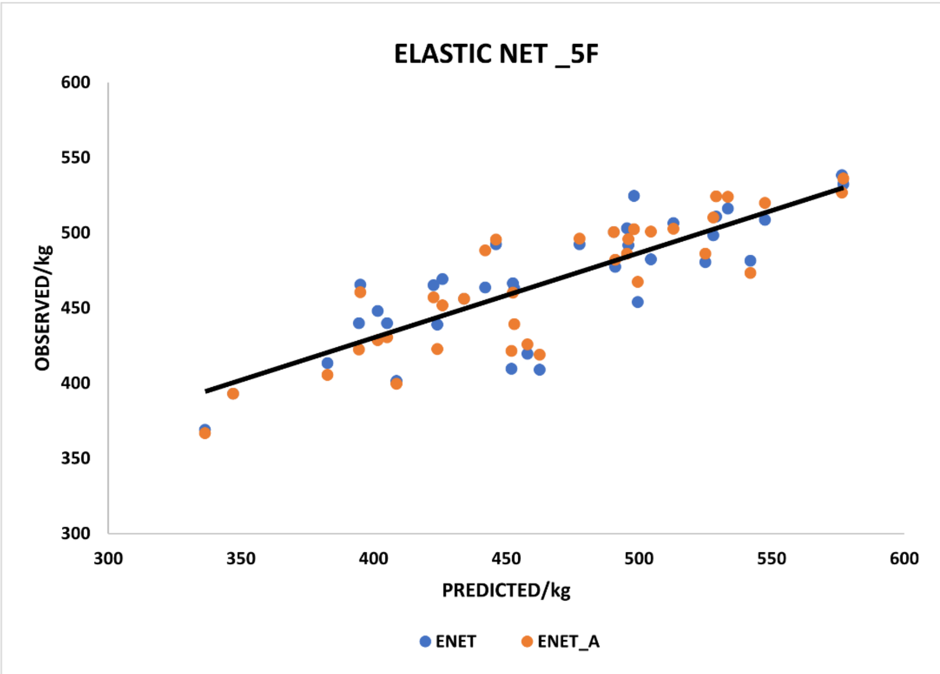
The regularizations using only shape descriptors as an independent variable, the results were from 5F dataset. However, LASSO and ENET (Fig 6 and 7) had equivalent results, but ENET was slightly superior. Both algorithms work with the L1 penalty, but ENET also uses penalty L2 (Ridge), so it appears as a solution to fill the gaps of other regularizations and comparing to LASSO. According to Garcia-Neto et al. (2021), the difference between these methods, is their alfa and lambda choices, while LASSO chooses one of them randomly, ENET chooses both simultaneous.



LASSO
R^2=0.65
MAE=31.62
RMSE= 36.00

LASSO_A
R^2=0.75
MAE=25.62
RMSE= 31.11

Fig 6. Comparison of the measured and estimated value of body weight for validation dataset of 5F dataset for LASSO with and without AGE as independent variable



ENET
R^2=0.68
MAE=30.72
RMSE= 34.98

ENET_A
R^2=0.76
MAE=25.45
RMSE= 30.96

Fig 7. Comparison of the measured and estimated value of body weight for validation dataset of 5F dataset for ENET with and without AGE as independent variable

The addition of AGE in the datasets helped to improve the performance of regularizations and further highlighted the superiority of the ENET algorithm, which reached R^2 >0.75.

The best result found among all the analyses, coincides to the findings by Jang et al. (2020) that improved the coefficient of determination using this variable as predictor. The linearity between measured and estimated body weight for both datasets can be seen in Fig 7.

The MAE and RMSE are the most common metrics to evaluate machine learning algorithms, both express the average of the prediction models error in unit of the variable of interest, in this paper in kilogram (kg), they are negatively oriented scores, which means that lower values are better. Based on the previously information, and the models' metric results for the best performance, using ENET without AGE as an independent variable, it's possible infer that for bidimensional sensors the values were greater (MAE: 30.72 and RMSE: 34.98 kg) when compare with papers like Weber et al. 2020b which had 41.56 kg of MAE and 50.94 kg of RMSE for BW prediction of Girolando cattle through images. Also, the values of metrics were inferior to those found by Weber et al. 2020a for Nellore cattle (MAE: 15.44 and 19.16 kg) without and using AGE as a predictor, although, its important notice that they extracted 30 and 50 measures, while we use only 8 features plus AGE, meaning that animals color characteristics could influence in features extracted from images.

In addition, it was possible to observe that the increase in the determination coefficients, implied in MAE and RMSE errors reductions, they decreased by approximately 5 kg (Table 2 and 3) for all models, representing 5.4% (MAE) and 6.6% (RMSE) of BW like found by Gomes et al. (2016) experiment for Angus breed and better than Jang et al. (2016) experiment using 3D images. Age is valuable information that helps management of animal production systems, once age along with other information identifying precocious or super precocious animals and avoiding those which end up presenting losses, besides age influences the growth intensity (HOZÁKOVÁ et al 2020). In addition, age allow to interpret the body development (ZHANG et al. 2018), or choosing animals that are suitable for slaughter, for example, keeping in mind that age affect in carcasses quality characteristics (KUSS et al. 2010).

For final equations, AGE had a significant value, but it was not predominant for LASSO and ENET in 5F_A dataset, since the shape descriptors demanded higher coefficients, thus considering that AGE did not mask shape descriptors extracted from images (Appendix Table 5). Both ENET and LASSO models are correlated by reduction of dimensionality and penalty that implies in final equations (GARCIA-NETO et al. 2021). The LASSO demanded fewer features than ENET, which used all shape descriptors (Appendix Table 3).

Considering that LASSO was the model that most eliminated features, width and length information were kept in the final model, which coincide to features present in studies that use images that used different number of lengths in bidimensional (WEBER et al. 2020) and even tridimensional images (GOMES et al. 2016; COMINOTTE et al. 2020) models that demand less information require less computational power and performing these processing faster than models that need more features for estimation.

For the beef cattle system, BW is one of the principal factors for feeding management. The nutrients necessary for growth and body weight gain are related to feeding and it represents about 70% of all cost for meat production (PACHECO et al. 2006).

Techniques based on body measurements *in loco* and through images (ASHWINI et al. 2019; WEBER et al. 2020) are widely used in studies that try to determine the relationship between BW and the structure of the animal species. However, the statistical approach used to estimate BW is also important, our study showed that for shape descriptors the regularizations gave better estimations.

It is important to keep in mind that experiments that used bidimensional sensors and ranged substantial or very good R^2 (>0.75) combine the dorsal view with the side view or in loco measures (NICHOLAS et al. 2018; ZHANG et al. 2018; NILCHUEN et al. 2021) or using tridimensional sensors as Kinect (GOMES et al. 2016; FERNANDES et al. 2019; JANG et al. 2020) that gives the volume information that is a very for BW prediction. In our experiment the animals were not handled, meaning that had no human contact for any of the data collection, besides our record system used the simplest and cheapest equipment.

4 Conclusion

For images segmentation, automatic processing cannot extract the ROI - animals dorsal - by edge detection or via threshold. For the treatments 5 Frames information plus ENET regularization shows substantial performance for BW prediction results. Thus, the use of bidimensional sensor in the dorsal view can predict BW of crossbreed (½ Angus x ½ Nellore) and the prediction power of Machine Learning Algorithm improves when age data is associated.

Funding

This study was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Reference

Ashwini, J. P., Sanjay, P., Amipara, G. J., Lunagariya, P. M., Parmar, D. J., & Rank, D. N. (2019). Prediction of body weight based on body measurements in crossbred cattle. *Int. J. Curr. Microbiol*. App. Sci, 8(03), 1597-1611.

Andrew, W., Greatwood, C., & Burghardt, T. (2019). Aerial Animal Biometrics: Individual Friesian Cattle Recovery and Visual Identification via an Autonomous UAV with Onboard Deep Inference. *IEEE International Conference on Intelligent Robots and Systems*, 237–243.

Andrew, W., Greatwood, C., & Burghardt, T. (2020). Fusing Animal Biometrics with Autonomous Robotics: Drone-based Search and Individual ID of Friesian Cattle (Extended Abstract). *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision Workshops*, WACVW 2020, 38–43.

Bahlo, C., Dahlhaus, P., Thompson, H., & Trotter, M. (2019). The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Computers and electronics in agriculture*, 156, 459-466.

Banhazi, T. M., Lehr, H., Black, J. L., Crabtree, H., Schofield, P., Tscharke, M., & Berckmans, D. (2012). Precision livestock farming: an international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering,* 5(3), 1-9.

Barthelme, S., Tschumperle, D., Wijffels, J., Assemlal H. E., Ochi, S. Rstudio. (2022). dyplyr version 1.0.10.

Çankaya, S., Eker, S., & Abacı, S. H. (2019). Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis. *Turkish Journal of Agriculture-Food Science and Technology*, 7(8), 1166-1172.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623.

Cominotte, A., Fernandes, A. F. A., Dorea, J. R. R., Rosa, G. J. M., Ladeira, M. M., van Cleef, E. H. C. B., ... & Neto, O. M. (2020). Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. *Livestock Science*, 232, 103904.

Craigie, C. R., Navajas, E. A., Purchas, R. W., Maltin, C. A., Bünger, L., Hoskin, S. O., ... & Roehe, R. (2012). A review of the development and use of video image analysis (VIA) for beef carcass evaluation as an alternative to the current EUROP system and other subjective systems. Meat science, 92(4), 307-318.

Dancey, Christine.; Reidy, John. (2006). Estatística Sem Matemática para Psicologia: Usando SPSS para Windows. Porto Alegre, Artmed.

Fernandes, A. F., Dórea, J. R., Fitzgerald, R., Herring, W., & Rosa, G. J. (2019). A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. *Journal of animal science*, 97(1), 496-508.

Fernandes, A. F. A., Dórea, J. R. R., & Rosa, G. J. D. M. (2020). Image analysis and computer vision applications in animal sciences: an overview. *Frontiers in Veterinary Science*, 7, 551269.

García-Nieto, P. J., García-Gonzalo, E., & Paredes-Sánchez, J. P. (2021). Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques. *Neural Computing and Applications*, 33(24), 17131-17145.

Gardenier, J., Underwood, J., & Clark, C. (2018). Object Detection for Cattle Gait Tracking. Proceedings - IEEE International Conference on Robotics and Automation, 2206–2213.

Gjergji, M., de Moraes Weber, V., Silva, L. O. C., da Costa Gomes, R., De Araújo, T. L. A. C., Pistori, H., & Alvarez, M. (2020). Deep learning techniques for beef cattle body weight prediction. *In 2020 International Joint Conference on Neural Networks (IJCNN*) (pp. 1-8). IEEE.

Gomes, R. A., Monteiro, G. R., Assis, G. J. F., Busato, K. C., Ladeira, M. M., & Chizzotti, M. L. (2016). Estimating body weight and body composition of beef cattle trough digital image analysis. *Journal of Animal Science*, *94*(12), 5414-5422.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, 19(2), 139-152.

Halachmi, I., Klopčič, M., Polak, P., Roberts, D. J., & Bewley, J. M. (2013). Automatic assessment of dairy cattle body condition score using thermal imaging. *Computers and Electronics in Agriculture*, 99, 35–40.

Halachmi, I., Polak, P., Roberts, D. J., & Klopcic, M. (2008). Cow body shape and automation of condition scoring. *Journal of Dairy Science*, 91(11), 4444–4451.

Hozáková, K., Vavrišínová, K., Neirurerová, P., & Bujko, J. (2020). Growth of beef cattle as prediction for meat production: A review. *Acta Fytotechnica et Zootechnica*, 23(2), 58-69.

Jang, D. H., Kim, C., Ko, Y. G., & Kim, Y. H. (2020). Estimation of body weight for Korean cattle using three-dimensional image. *Journal of Biosystems Engineering*, 45(4), 325-332.

Khan, M. A., Tariq, M. M., Eyduran, E., Tatlıyer, A., Rafeeq, M., Abbas, F., & Javed, K. (2014). Estimating body weight from several body measurements in Harnai sheepwithout multicollinearity problem. *The Journal of Animal & Plant Sciences*, 24(1)

Kuss, F., López, J., Restle, J., Barcellos, J. O. J., Moletta, J. L., & Paula-Leite, M. C. D. (2010). Meat quality of non-castrate or castrated males feedlot finished and slaughtered at 16 or 26 months of age. *Brazilian Journal of Animal Science*, 39, 924-931.

Leibrandt, S., & Le Pennec, J. L. (2015). Towards fast and routine analyses of volcanic ash morphometry for eruption surveillance applications. *Journal of Volcanology and Geothermal Research,* 297, 11-27.

Menesatti, P., Costa, C., Antonucci, F., Steri, R., Pallottino, F., & Catillo, G. (2014). A low-cost stereovision system to estimate size and weight of live sheep. *Computers and Electronics in Agricultu*re, 103, 33-38.

Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. *Remote Sensing*, 14(15), 3547.

Nicolas, F. F. C., Saludes, R. B., Relativo, P. L. P., & Saludes, T. A. (2018). Estimating live weight of Philippine dairy buffaloes (Bubalus bubalis) using digital image analysis. *Philippine Journal of Veterinary and Animal Sciences*, 44(2), 129-138.

Nilchuen, P., Yaigate, T., & Sumon, W. (2021). Body measurements of beef cows by using mobile phone application and prediction of body weight with regression model. *Songklanakarin Journal of Science & Technology*, *43*(6).

Pacheco, P. S., Restle, J., Vaz, F. N., Freitas, A. K., Pádua, J. T., Neumann, M.; Arboitte, M. Z. (2006). Economical evaluation of feedlot finished steers and young steers from different genetic groups. *Brazilian Journal of Animal Science*.35: 309-320

Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., & Sukkarieh, S. (2021). Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. *Computers and Electronics in Agriculture*, 185, 106143.

R Core team (2021). R: A language and environment for statistical computing.

Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V., Dzhulamanov, K., Kalschikov, V., & Guo, H. (2022). Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows. Computers and Electronics in Agriculture, 195, 106837.

Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V., & Miroshnikov, S. (2020). Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery. Computers and Electronics in Agriculture, 179, 105821.

Salau, J., Haas, J. H., Junge, W., & Thaller, G. (2017). A multi-Kinect cow scanning system: Calculating linear traits from manually marked recordings of Holstein-Friesian dairy cows. *Biosystems Engineering*, 157, 92–98.

Suryawanshi, K. R., Redpath, S. M., Bhatnagar, Y. V., Ramakrishnan, U., Chaturvedi, V., Smout, S. C., & Mishra, C. (2017). Impact of wild prey availability on livestock predation by snow leopards. *Royal Society Open Science*, 4(6), 170026.

Tullo, E., Finzi, A., & Guarino, M. (2019). Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Science of the total environment*, 650, 2751-2760.

Vivaracho-Pascual, C., Simon-Hurtado, A., Manso-Martinez, E., & Pascual-Gaspar, J. M. (2016). Client threshold prediction in biometric signature recognition by means of Multiple Linear Regression and its use for score normalization. *Pattern Recognition*, 55, 1-13.

Wickham , H., François, R., Henry, L., Müller, K.,, RStudio. (2022). Image version .42.13

Weber, V. A. M., de Lima Weber, F., da Silva Oliveira, A., Astolfi, G., Menezes, G. V., de Andrade Porto, J. V., ... & Pistori, H. (2020). Cattle weight estimation using active contour models and regression trees Bagging. *Computers and Electronics in Agriculture*, 179, 105804.a

Weber, V. A. D. M., Weber, F. D. L., Gomes, R. D. C., Oliveira Junior, A. D. S., Menezes, G. V., Abreu, U. G. P. D., ... & Pistori, H. (2020). Prediction of Girolando cattle weight by means of body measurements extracted from images. Revista Brasileira de Zootecnia, 49.b

Yakubu, A. (2010). Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani cows. *Journal of Central European Agriculture.*

Zhang, A. L. N., Wu, B. P., Jiang, C. X. H., Xuan, D. C. Z., Ma, E. Y. H., & Zhang, F. Y. A. (2018). Development and validation of a visual image analysis for monitoring the body size of sheep. *Journal of Applied Animal Research*, 46(1), 1004-1015.

Zhang, W., Wu, C., Li, Y., Wang, L., & Samui, P. (2019). Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *15*(1), 27-40.

Appendix

TABLES

Table 1. Statistical description of animal characteristics in the dataset 1F.

	Mean	SD	Max	Min
Body Weight (BW/kg)	467.8	± 61,71	606.5	336.5
Area (A/pixel²)	273145	± 84361,55	611812	148813
Perimeter (Per/pixel)	2181	±307,44	1611	3162
Width(W/pixel)	917.9	± 123,21	688.0	1259.0
Length (L/pixel)	391.5	± 61,54	588.0	280.0
Major (Ma/pixel)	909.9	±131,39	1308.0	663.3
Minor (Mi/pixel)	374.6	±60,85	595.5	268.7
Fer Diameter (Fer/pixel)	921.3	±124,22	1269.7	688.7
MinFeret (MFer/pixel)	386.1	± 60,30	586.4	278.2

Table 2. Statistical description of animal characteristics in the dataset 3F.

	Mean	SD	Max	Min
Body Weight (BW/kg)	467.8	± 61,71	606.5	336.5
Area (A/pixel²)	265285	±71903.99	567744	148813
Perimeter (Per/pixel)	2167.56	±277.14	3046.93	1611.38
Width(W/pixel)	917.32	±114.59	1223.0	684.0
Length (L/pixel)	385.20	±54.83	600.0	277.0
Major (Ma/pixel)	904.12	±119.11	1265.70	659.62
Minor (Mi/pixel)	367.70	±54.08	588.44	267.87
Fer Diameter (Fer/pixel)	920.74	±115.34	1237.19	684.26
MinFeret (MFer/pixel)	380.04	±54.07	591.19	277.0

Table 3. Statistical description of animal characteristics in the dataset 5F.

	Mean	SD	Max	Min
Body Weight (BW/kg)	467.8	± 61,71	606.5	336.5
Area (A/pixel²)	265363.18	±76477.99	570045	148499
Perimeter (Per/pixel)	2164.55	±285.90	3082.728	1611.379

Width (W/pixel)	915.92	±117.52	1252.0	678.0
Length (L/pixel)	385.63	±57.39	602.0	277.0
Major (Ma/pixel)	902.98	±122.84	1289.931	659.615
Minor (Mi/pixel)	368.30	±56.81	592.057	267.03
Fer Diameter (Fer/pixel)	919.40	±118.29	1270.623	679.559
MinFeret (MFer/pixel)	380.63	±56.86	600.102	277.0

Table 4. Lambda values obtained for Ridge model.

	1 FRAME	**3 FRAMES**	**5 FRAMES**
RIDGE	3.16	3.16	3.16
RIDGE_A	3.16	3.16	3.16

Table 5. Final equation for each model and dataset for BW prediction of crossbreed cattle.

	1F DATASET
	FEATURES
RIDGE	Y=468.52+A*(5.33)+Per*(-1.89)+W*(13.69)+L*(-1.13)+ Ma*(9.91)+ Mi*(8.16)+ Fer*(9.67)+ MFer*(1.91)
LASSO	Y=468.52+A*(-182.35)+Per*(-151.03)+ W*(149.04)+ L*(1.86)+ Ma*(149.16)+ Mi*(144.92)+ Fer*(-55.68)
ENET	Y=468.52+A*(-292.37)+Per*(-136.49)+ W*(223.65)+ L*(13.80)+ Ma*(200.92)+ Mi*(207.67)+ Fer*(-139.66)+ MFer*(-18.30)
MLR	Y= 469.260+ A*(-60.758)+Per*(-110.22)+ W*(205.09)+ L*(5.45)+ Ma*(71.66)+ Mi*(55.55)+ Fer*(-126.54)+ MFer*(14.14)
	FEATURES + AGE
RIDGE	Y=468.52+AGE*(19.42)+A*(1.95)+Per*(-2.28)+ W*(10.94)+ L*(0.72)+ Ma*(5.38)+ Mi*(5.28)+ Fer*(8.59)+ MFer*(2.97)
LASSO	Y=468.52+AGE*(18.61)+A*(-227.33)+Per*(-118.28)+ W*(92.25)+ L*(4.70)+ Ma*(136.63)+ Mi*(156.61)
ENET	Y=468.52+AGE*(19.77)+A*(-44.93)+Per*(-96.53)+ W*(69.77)+ L*(6.98)+ Ma*(32.30)+ Mi*(55.66)+ Fer*(17.45)+ MFer*(-2.83)
MLR	Y= 461.307 + AGE*(22.75) + A*(-10.34)+ Per*(-110.23) + W*(-3.78)+ L*(-15.30)+Ma*(87.45)+Mi*(-8.94) + Fer*(42.24)+ MFer*(56.25)
	3F DATASET
	FEATURES
RIDGE	Y=463.43+A*(2.97)+Per*(-9.08)+W*(17.87)+L*(-0.59)+ Ma*(4.31)+ Mi*(6.73)+ Fer*(12.31)+ MFer*(9.51)

LASSO	Y=463.43+A*(0.77)+ Per*(-150.98)+ W*(142.68)+ Ma*(2.07)+Mi*(46.16) +MFer*(9.97)
ENET	Y=463.43+A*(7.85)+Per*(-185.32)+ W*(145.75)+ Ma*(23.77)+ Mi*(35.68)+ Fer*(0.02) + MFer*(23.46)
MLR	Y= 472.57 + A*(-92.50) +Per *(-229.29)+ W* (384.24)+L*(31.83) + Ma*(123.53) +Mi*(62.33)+Fer*(-256.77)+ MFer*(26.07)
FEATURES + AGE	
RIDGE	Y=463.43+AGE*(20.36)+A*(1.24)+Per*(-9.55)+ W*(11.78)+ L*(0.76)+ Ma*2.18)+ Mi*(5.90)+ Fer*(7.78)+ MFer*(10.40)
LASSO	Y=463.43+AGE*(19.10)+Per*(-160.57) +W*(100.32)+ Ma*(40.99)+ Mi*(12.32)+ Fer*(0.02)+ MFer*(43.92)
ENET	Y=463.43+AGE*(19.22)+ Per*(166.16)+ W*(88.76)+ Ma*(44.10)+ Mi*(10.12)+ Fer*(13.21)+ MFer*(46.94)
MLR	Y= 467.87+ AGE*(22.59)+A*(-42.05)+ Per*(-233.54) + W*(87.67)+ L*(-23.72)+ Ma*(145.58)+ Mi*(29.96)+ Fer*(-6.44) +MFer*(80.95)
5F DATASET	
FEATURES	
RIDGE	Y=466.32+ A*(-0.86)+ Per*(-4.14)+ W*(19.51)+ L*(-3.74)+ Ma*(3.54)+ Mi*(2.07)+ Fer*(13.68)+ MFer*(10.24)
LASSO	Y=466.32+Per*(-96.49)+W*(105.22)+L*(-1.44)+MFer*(36.91)
ENET	Y=466.32+A*(-25.67)+ Per*(-120.65)+ W*(79.29)+ L*(-33.62)+ Ma*(18.53)+ Mi*(24.92)+ Fer*(39.22)+MFer*(63.74)
MLR	Y= 459.829+ A*(-102.46)+ Per*(-325.11)+ W*(88.99)+L*(-6.11)+ Ma*(163.70)+ Mi*(72.55)+ Fer*(91.10) + MFer*(68.19)
FEATURES + AGE	
RIDGE	Y=466.32+ AGE*(18.24)+ A*(-0.43)+ Per*(-4.80)+ W*(13.48)+ L*(-3.51)+ Ma*(3.91)+ Mi*(2.34)+ Fer*(8.75)+ MFer*(9.24)
LASSO	Y=466.32+ AGE*(17.34)+ Per*(-104.77)+ W*(78.33)+ L*(-18.52)+ Ma*(24.11)+ MFer*(53.95)
ENET	Y=466.32+AGE*(17.65)+A*(-27.09)+Per*(-108.85)+W*(56.73)+L*(-33.43)+ Ma*(45.30)+ Mi*(13.19)+ Fer*(17.52)+ MFer*(70.46)
MLR	Y= 462.47+ AGE*(19.55)+ A*(-69.43)+Per(-261.84)+ W*(-61.41) + L*(-54.12) + Ma*(88.92)+ Mi*(57.01)+ Fer*(230.54) + MFer*(109.56)

A: Area; Per: Perimeter; W: Width; L: Length; Ma: Major; Mi: Minor; Fer: Feret diameter; MFer: MinFeret; LASSO: Least Absolute Shrinkage Select Operator; ENET: Elastic Net; RIDGE: Ridge; MLR: Multiple Linear Regression;

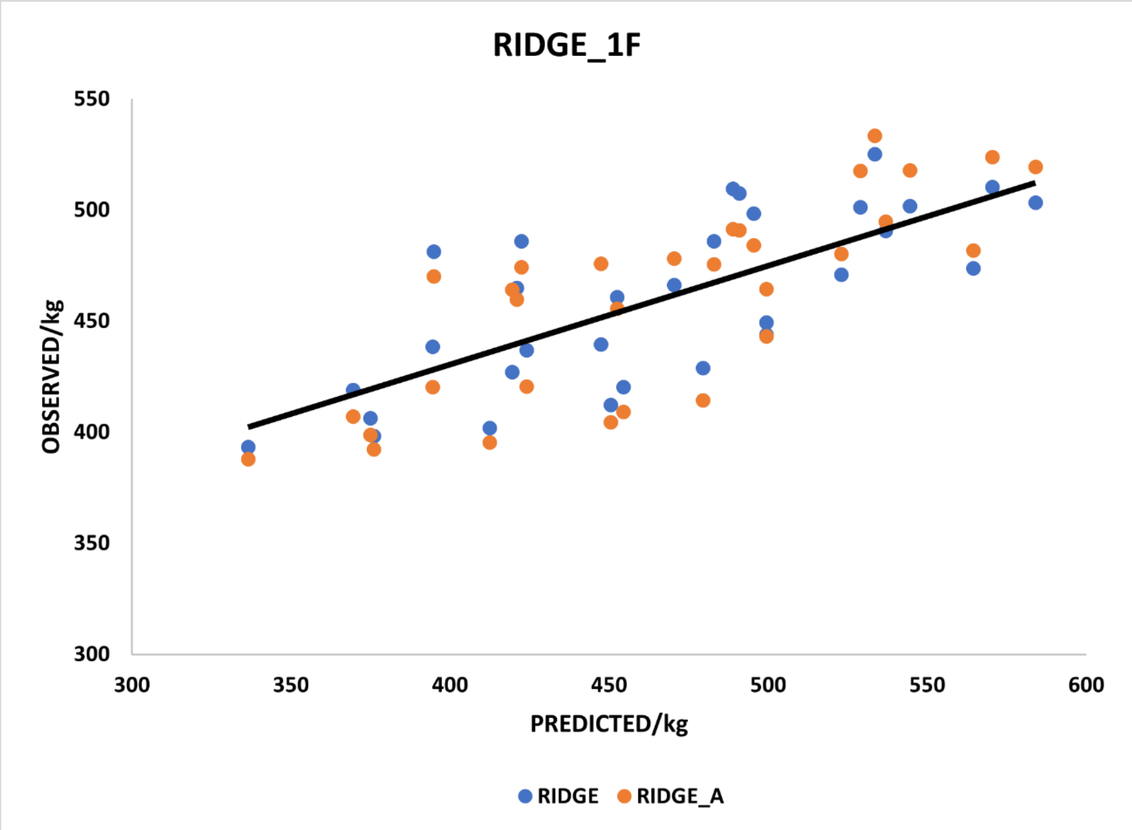
Table 6. Variable information relevance from the most to least important, for ensemble group.

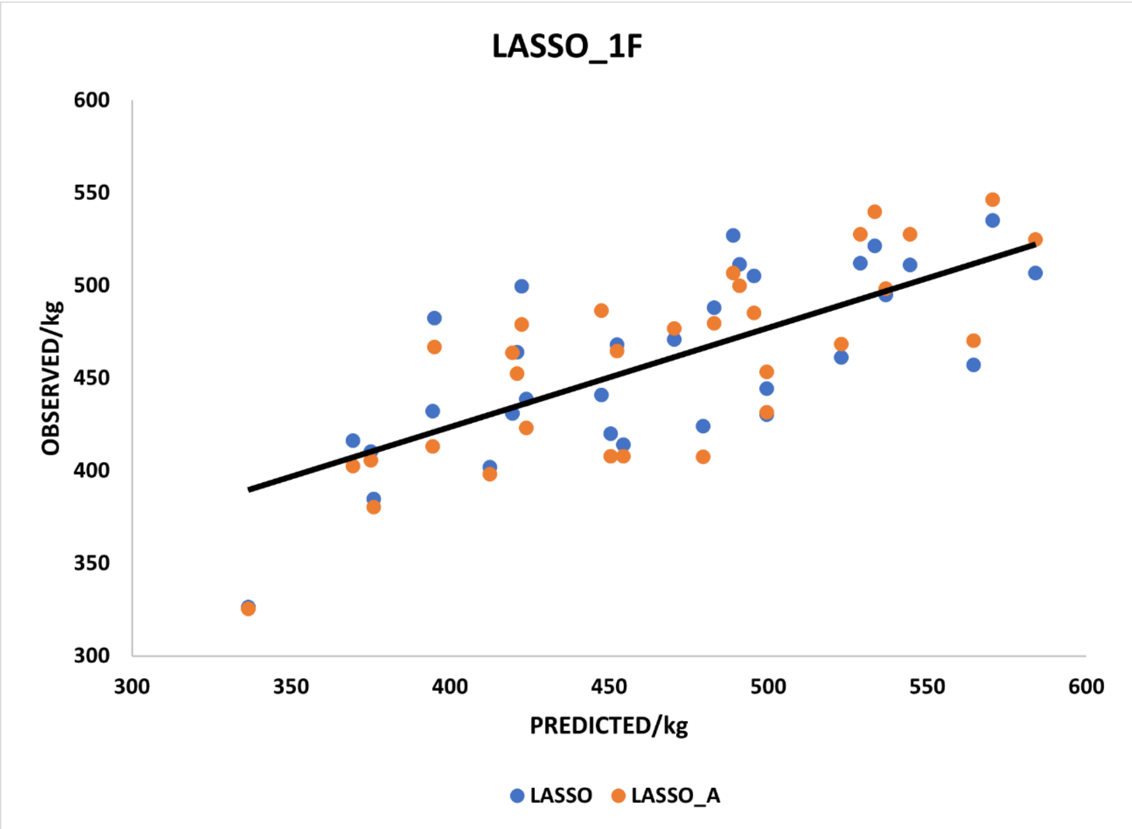
DATASET	RF	ADAB
	Without BW	
1F	A = 11.55| Per = 9.32| W = 10.69| L = 10.00| Ma = 10.02| Mi= 9.46| Fer=11.95| MFer= 9.46	W=20.90| Per=19.26| A=16.45| Mi=13.34| Ma=12.75| MFer=6.82| L=5.95| Fer=4.52
3F	A=11.20| Per=10.56| W=14.27| L=7.21| Ma=11.76| Mi=7.95| Fer=12.80| MFer=9.65	W=38.30| A=11.92| Per=10.80| Ma=10.79| L=10.49| Mi=9.30| Fer= 4.97| MFer=3.44
5F	A=11.74| Per=7.68| W=13.83| L=8.39| Ma=10.66| Mi=7.37| Fer=12.14| MFer=8.45	W=31.88| Ma=15.51| A=13.55| Per=13.15| L=10.14| Fer=7.23| Mi=5.41| MFer=3.14
	With BW	
1F	AGE=7.60|A=10.45|Per=9.29|W=11.49|L= 5.42| Ma=9.11|Mi=6.87| Fer=11.12| MFer=6.52	AGE=23.01| A= 5.98| W=15.69| Per = 14.46| Ma = 8.75| Mi=7.80| MFer =6.24 |L= 4.62| Fer = 3.25
3F	AGE=15.84| A=10.61| Per=7.58| W=11.77| L=8.74| Ma=8.12| Mi=10.36| Fer=9.56| MFer=9.74	AGE=26.48 | A=22.34| W=12.72| Mi=9.51| MFer=7.89| Per=7.82| Fer=5.21| Ma=4.16| L=3.87
5F	AGE=16.88| A=12.36| Per=6.78| W=13.22| L=10.82| Ma=9.04| Mi=11.15| Fer=10.03| MFer=12.17	W=39.97| AGE=23.52 |Ma=9.75| L=8.44| Mi=6.25| MFer=3.94| A=3.58| Fer=2.53| Per=2.02

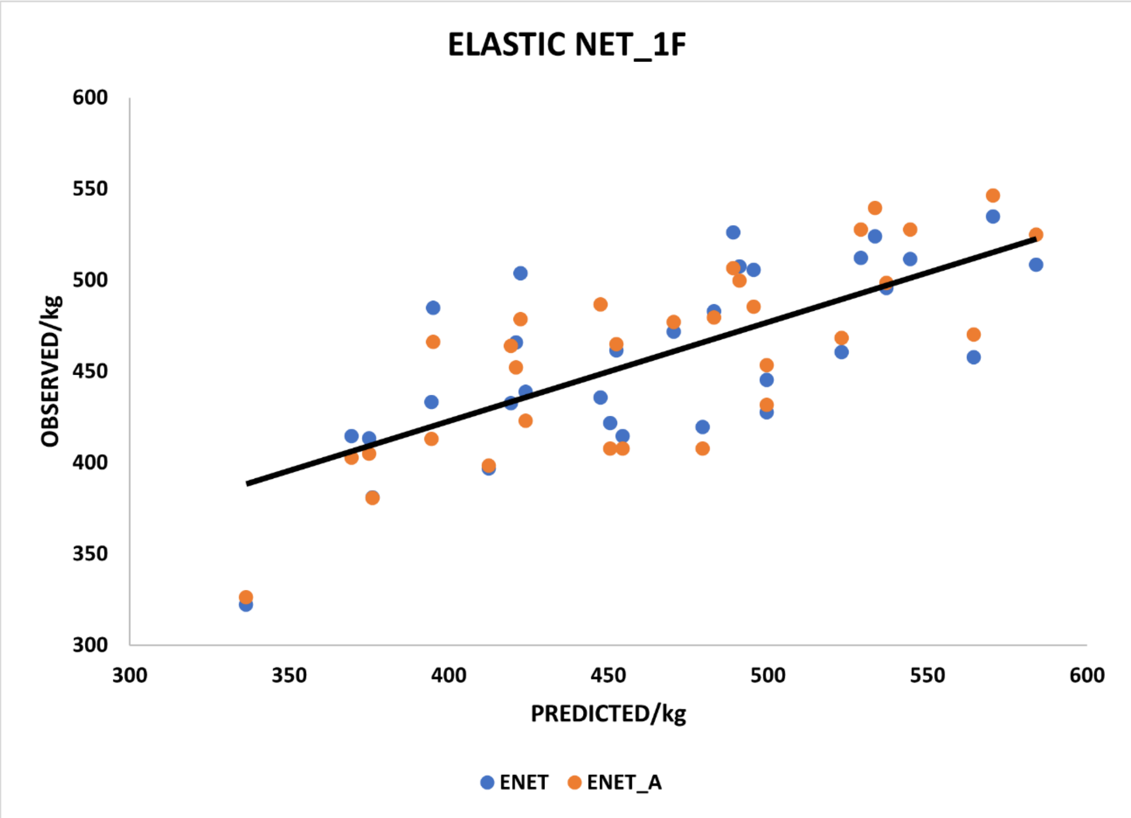
RF: Random Forest; ADAB: Adaboost; A: Area; Per: Perimeter; W: Width; L: Length; Ma: Major; Mi: Minor; Fer: Feret diameter; MFer: MinFeret; 1F: 1 Frame Information, 3F: 3 Frames Information, 5F: 5 Frames Information.

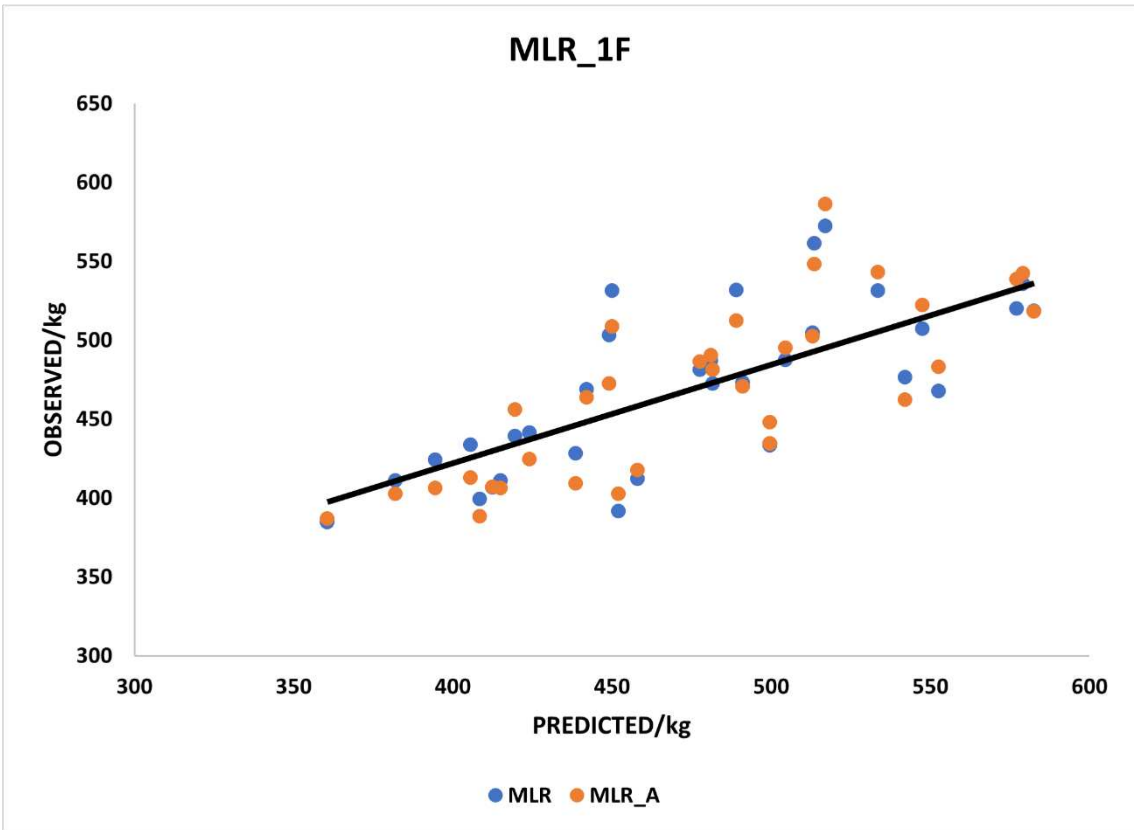
FIGURES

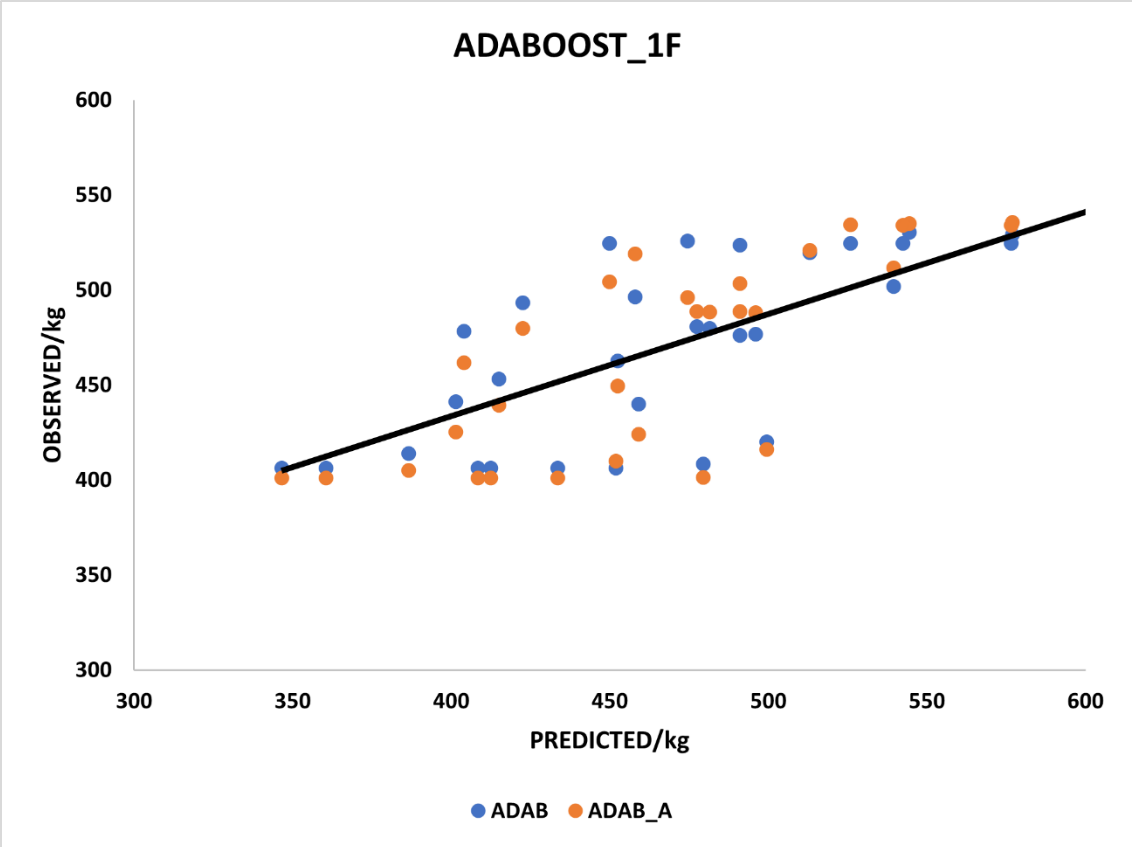
GRAPHICS






ELASTIC NET_1F
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
ENET
ENET_A


MLR_1F
650
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
MLR
MLR_A



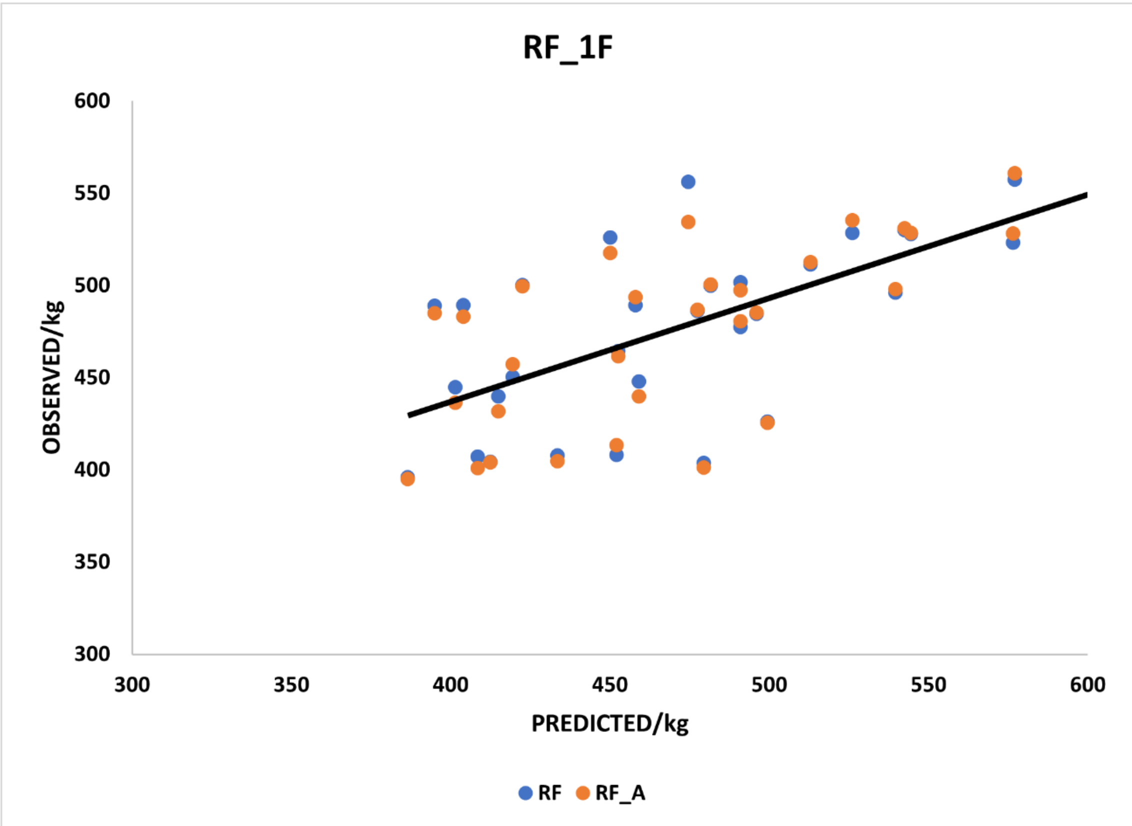
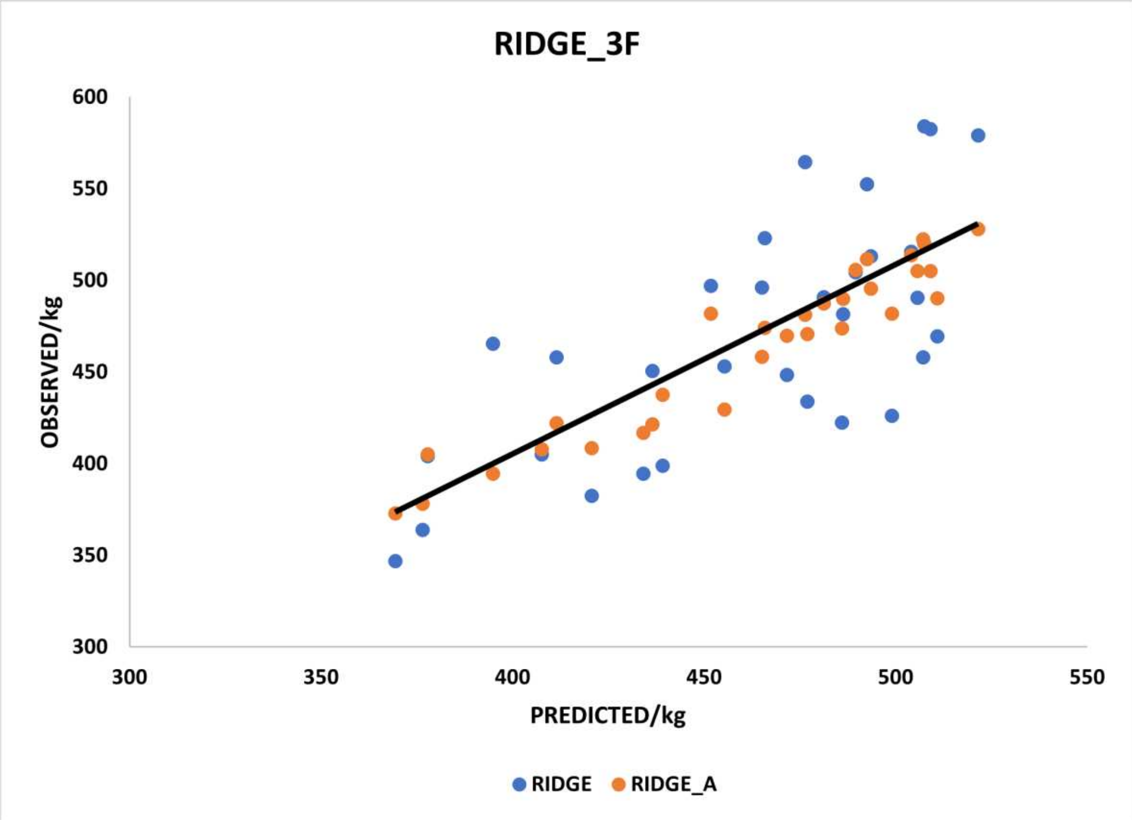
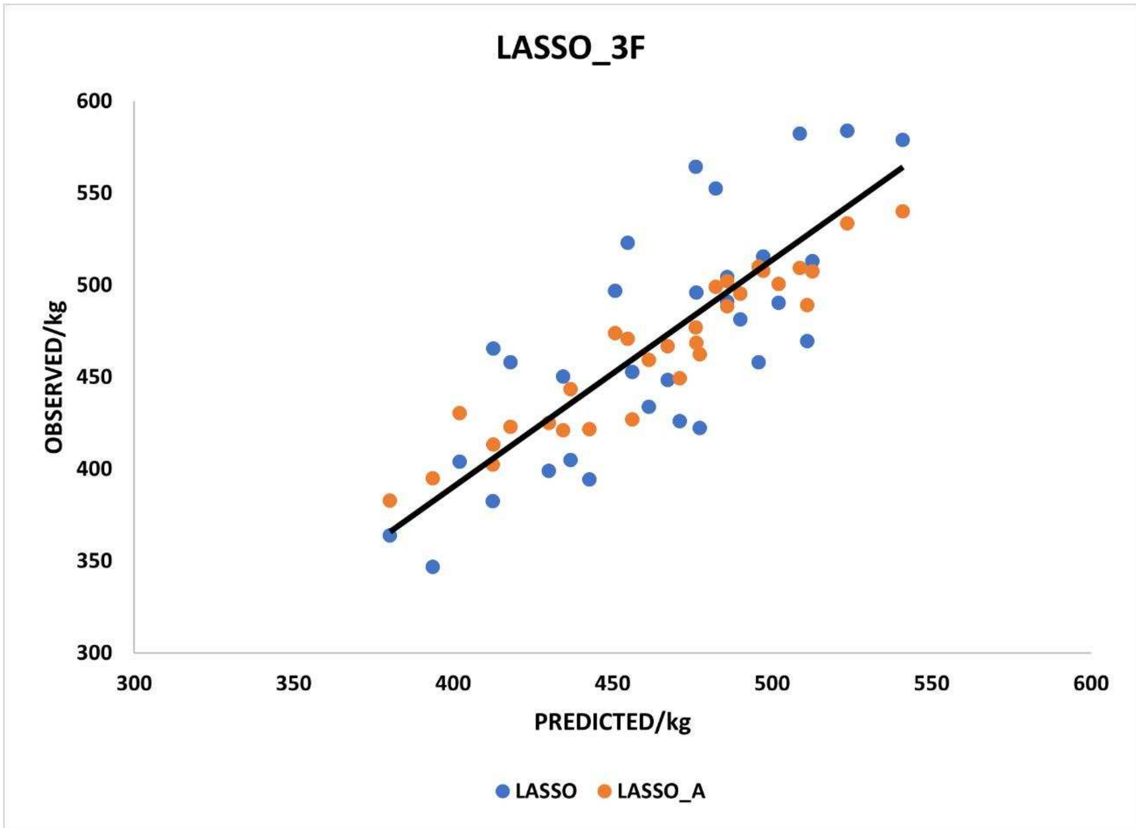
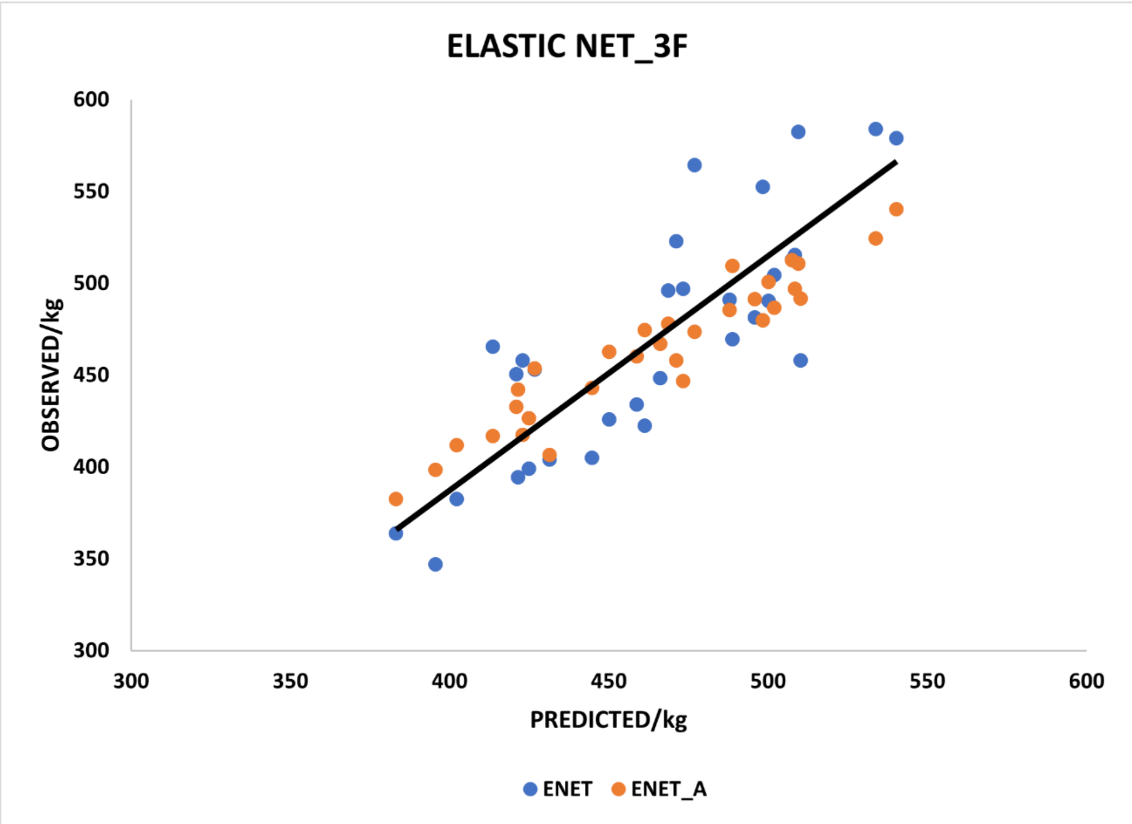


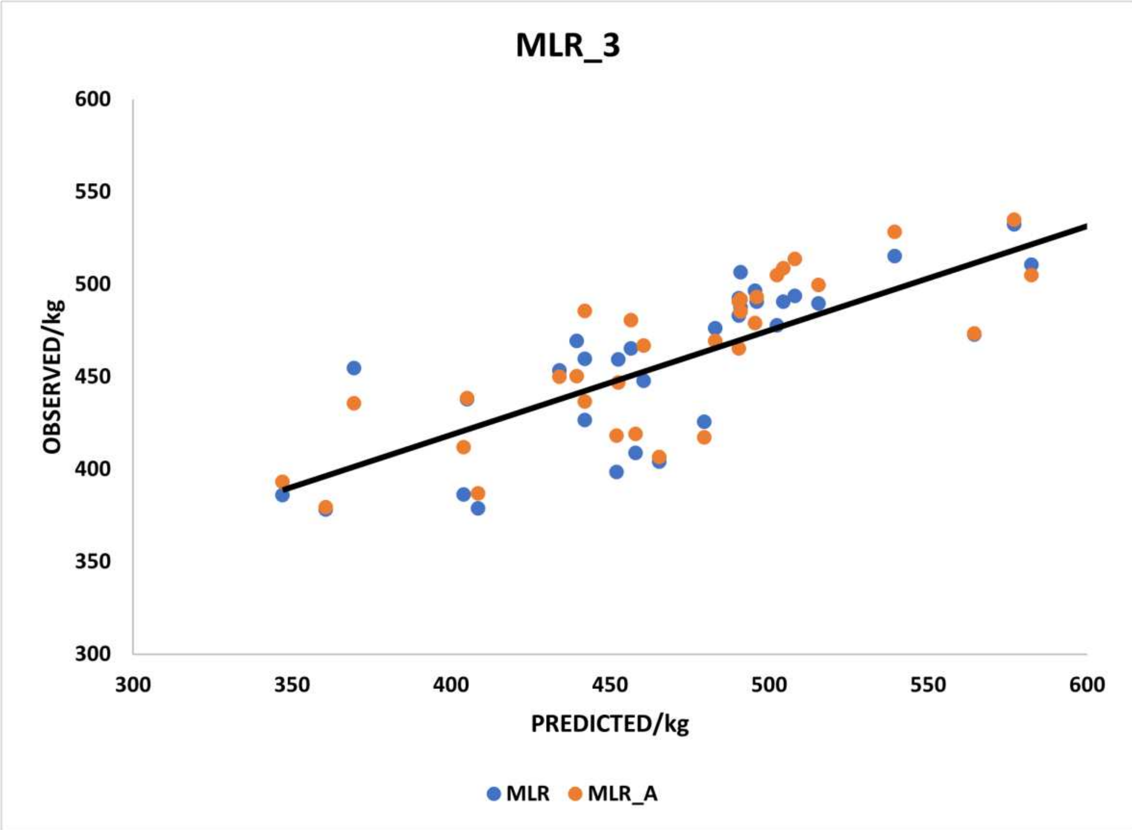
Fig 1. Observed vs. Predicted BW for each algorithm without and with AGE as independent variable and metrics results for 1F dataset. RIDGE: ridge using only shape descriptors; LASSO: LASSO: using only shape descriptors; ENET: ENET using only shape descriptors; MLR: MLR using only shape descriptors; ADAB:

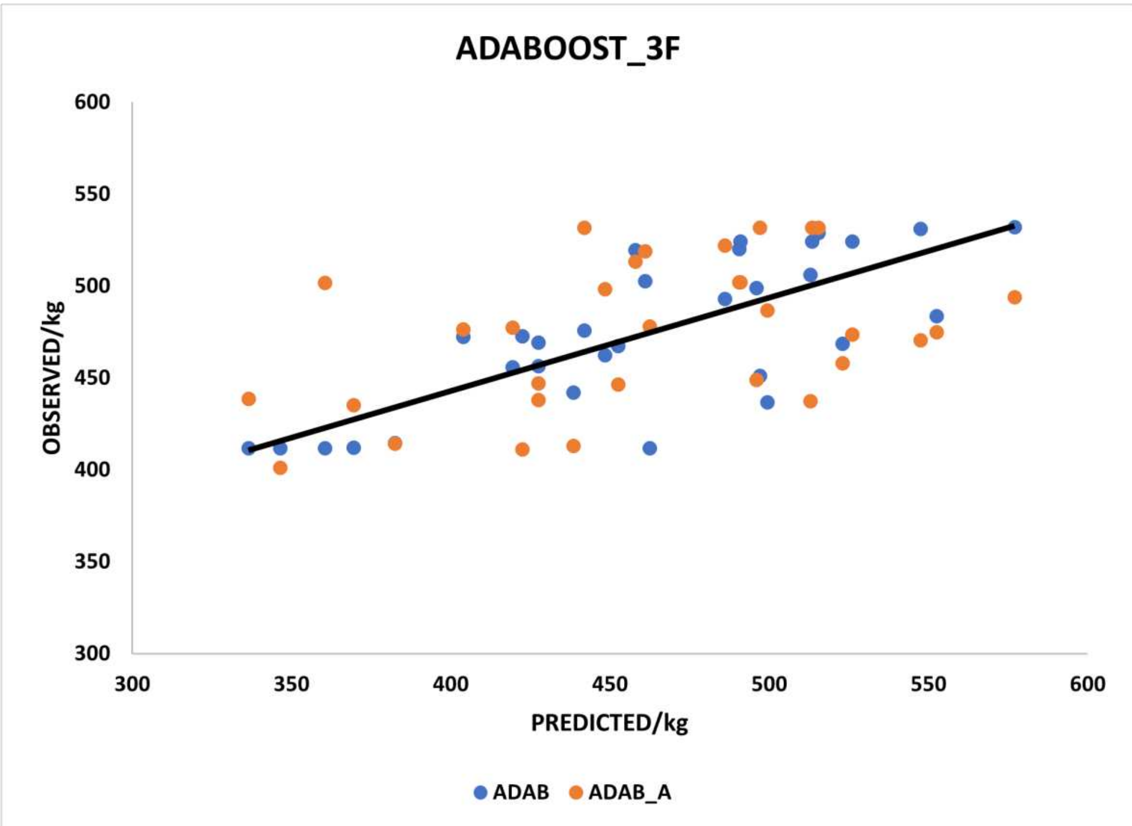
ADAB using only shape descriptors; RF: RF using only shape descriptors; RIDGE_A: ridge using AGE; LASSO_A: LASSO using AGE; ENET_A: ENET using AGE; MLR_A: MLR using AGE; ADAB_A: ADAB using AGE; RF_A: RF using AGE.




LASSO_3F
OBSERVED/kg
PREDICTED/kg
300
350
400
450
500
550
600
LASSO
LASSO_A


ELASTIC NET_3F
OBSERVED/kg
PREDICTED/kg
300
350
400
450
500
550
600
ENET
ENET_A


MLR_3
OBSERVED/kg
600
550
500
450
400
350
300
300
350
400
450
500
550
600
PREDICTED/kg
MLR
MLR_A


ADABOOST_3F
OBSERVED/kg
600
550
500
450
400
350
300
300
350
400
450
500
550
600
PREDICTED/kg
ADAB
ADAB_A

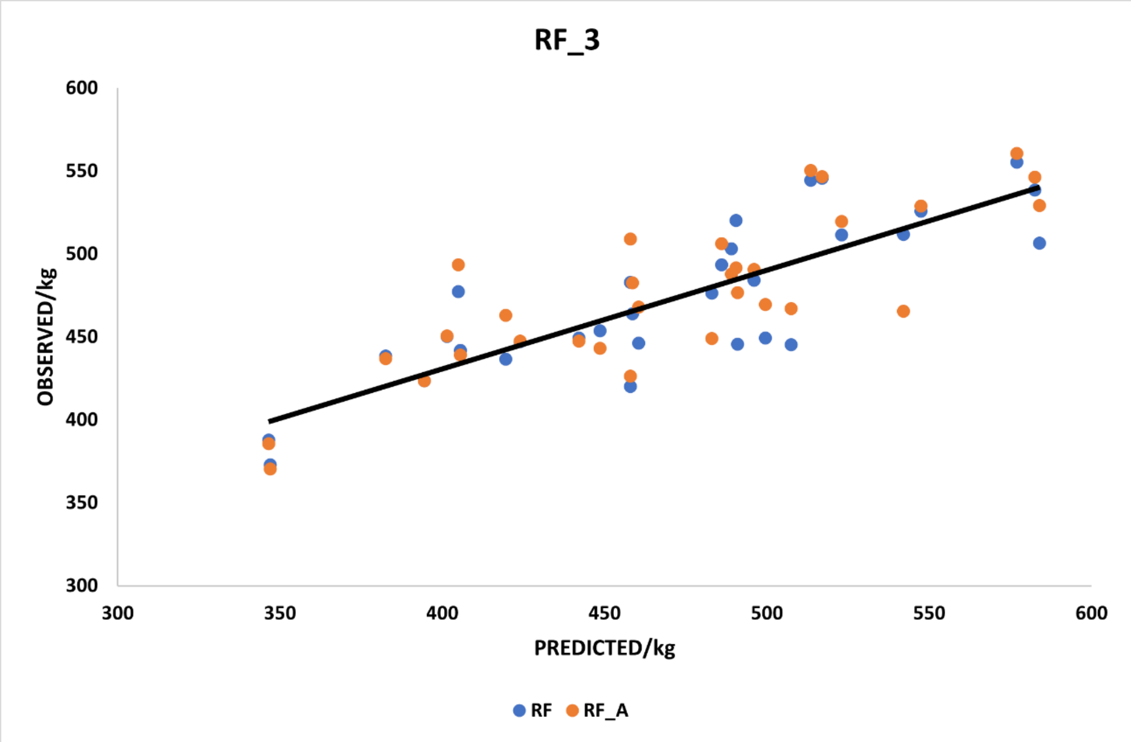
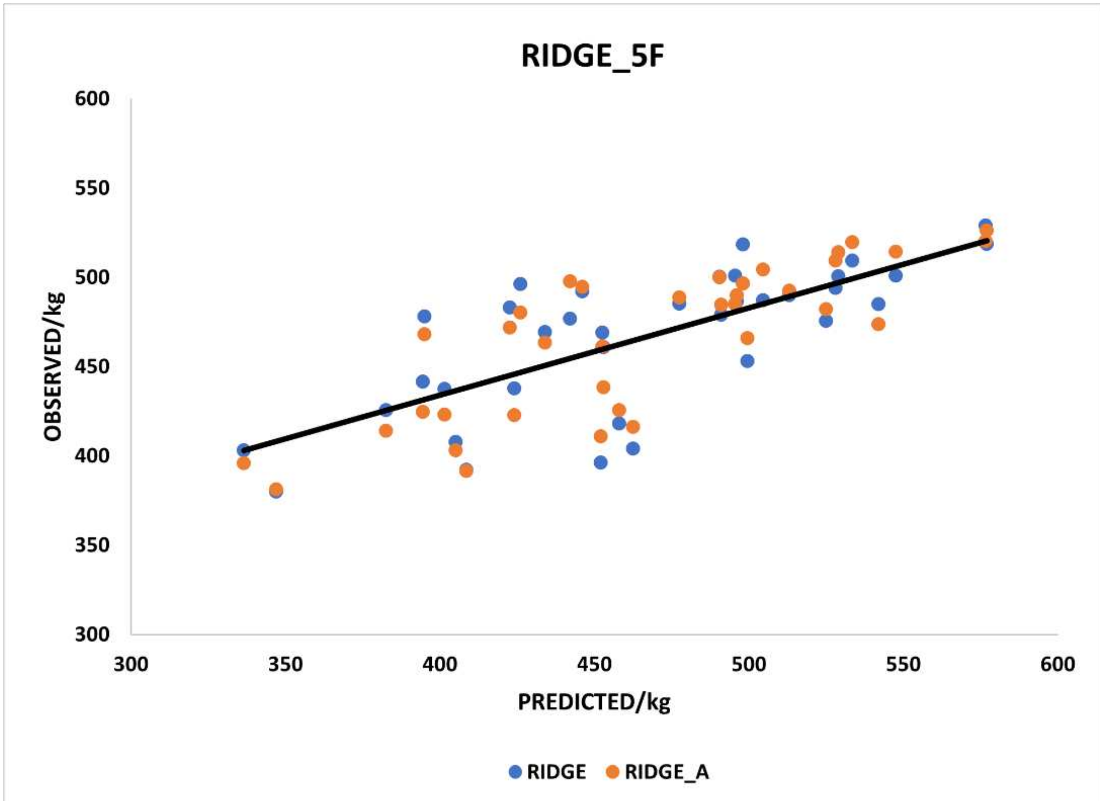
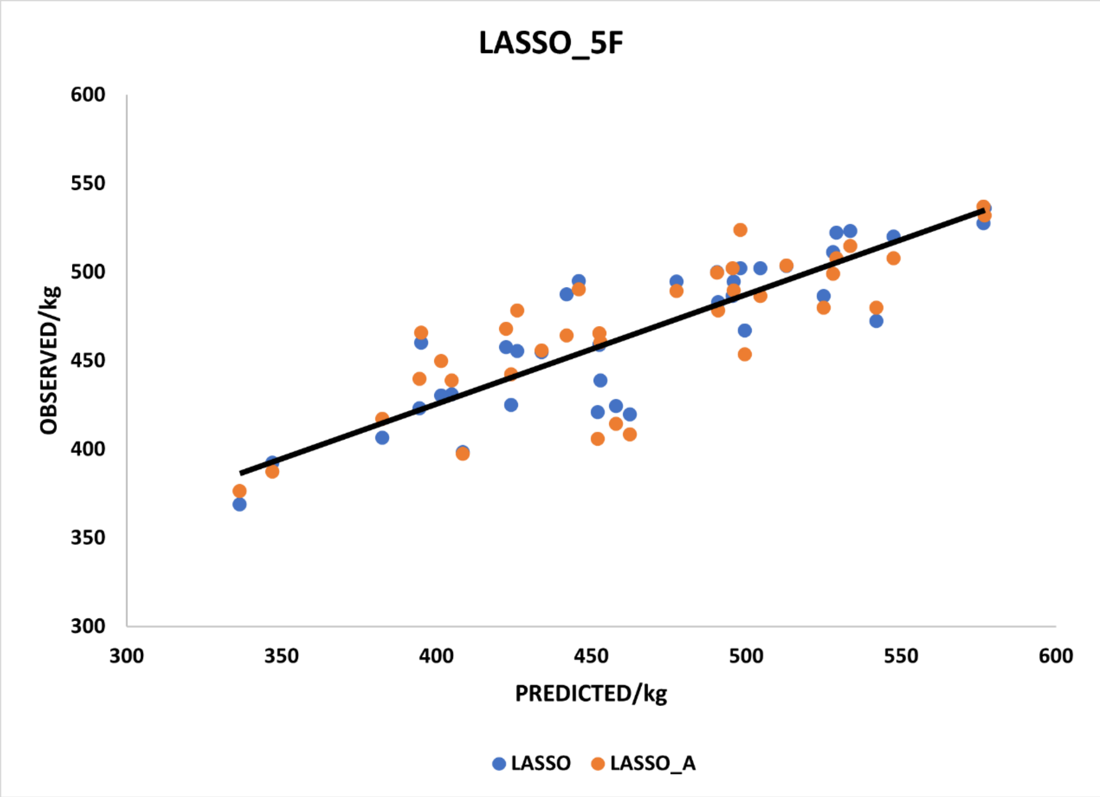
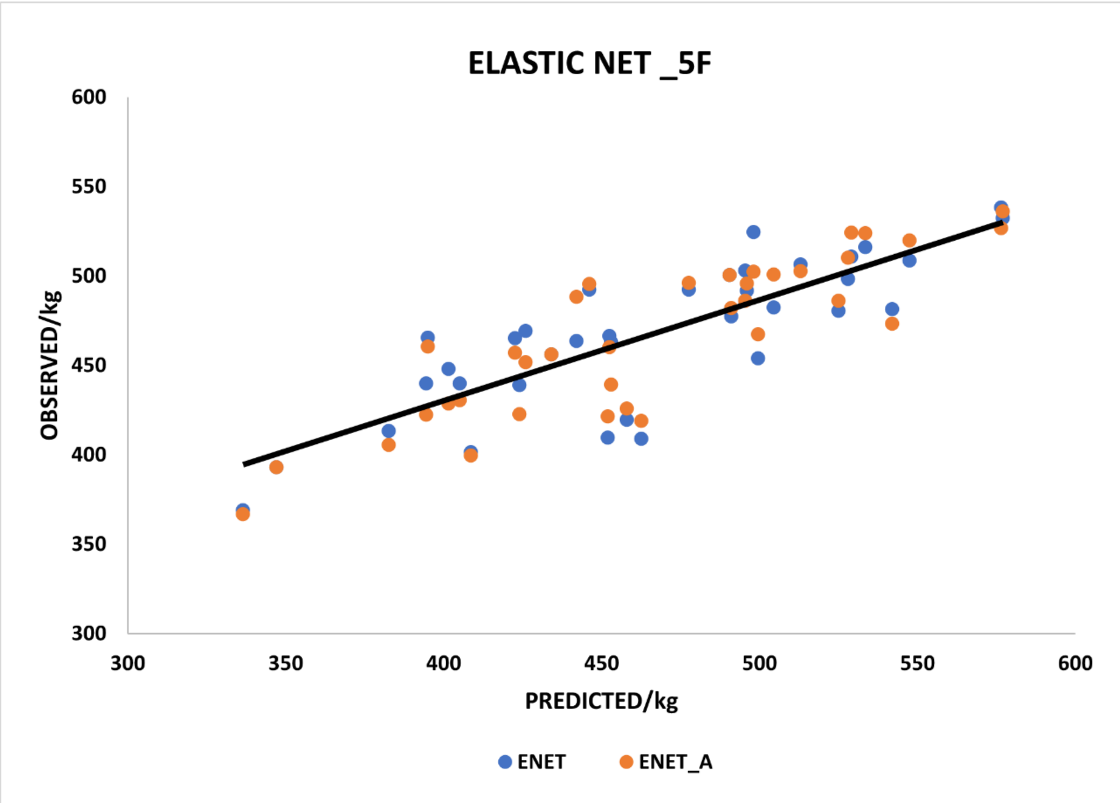


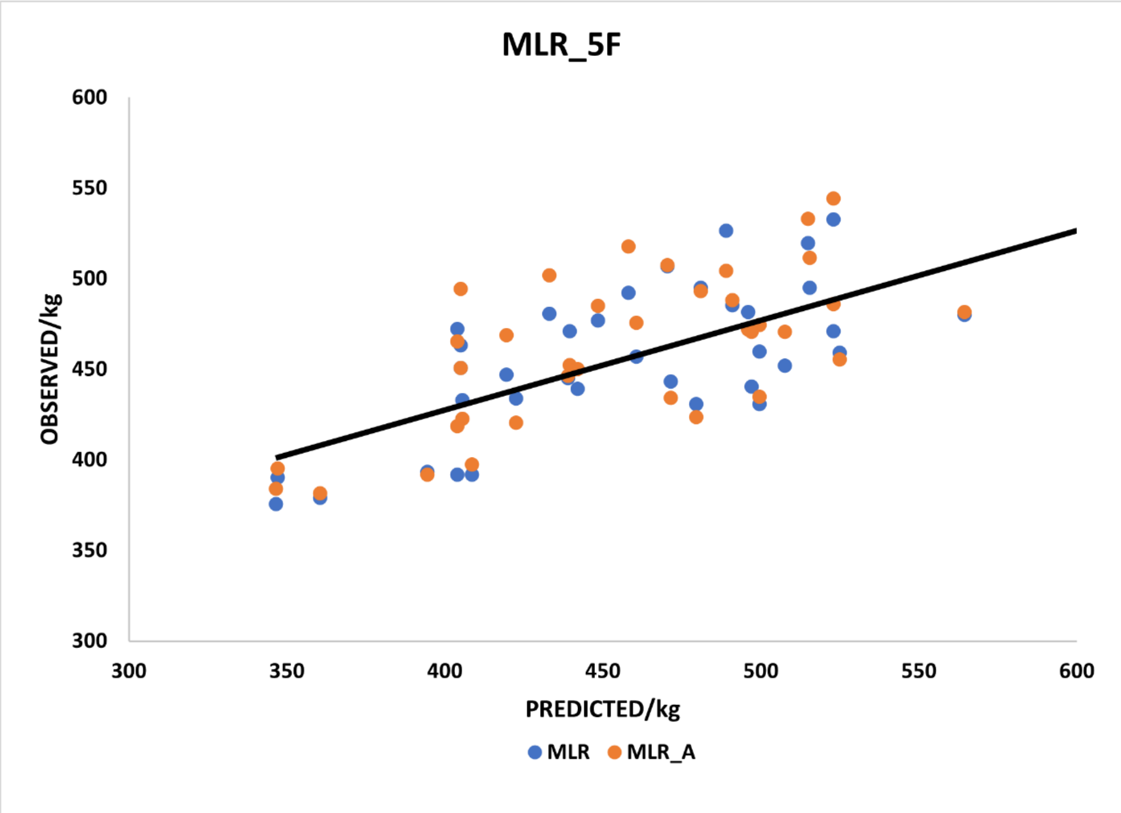
Fig 2. Observed vs. Predicted BW for each algorithm without and with AGE as independent variable and metrics results for 3F dataset. RIDGE: ridge using only shape descriptors; LASSO: LASSO: using only shape descriptors; ENET: ENET using only shape descriptors; MLR: MLR using only shape descriptors; ADAB: ADAB using only shape descriptors; RF: RF using only shape descriptors; RIDGE_A: ridge using AGE; LASSO_A: LASSO using AGE; ENET_A: ENET using AGE; MLR_A: MLR using AGE; ADAB_A: ADAB using AGE; RF_A: RF using AGE.

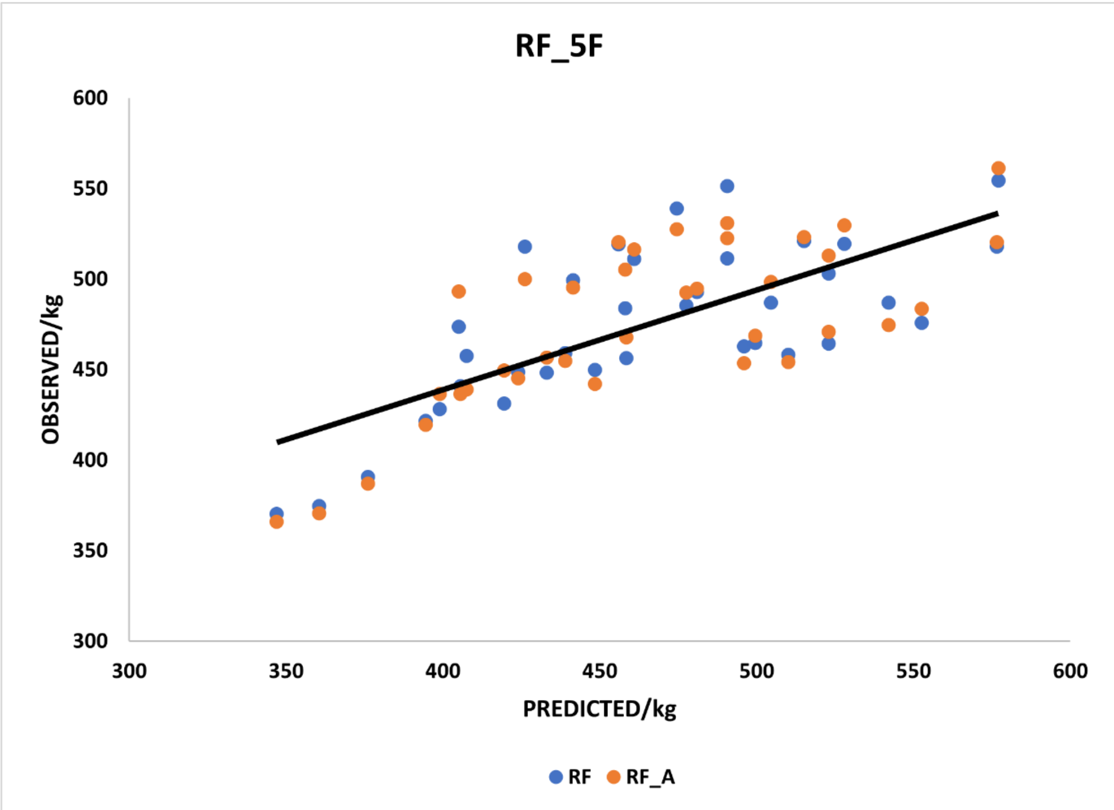
.




LASSO_5F
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
LASSO
LASSO_A


ELASTIC NET _5F
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
ENET
ENET_A


MLR_5F
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
MLR
MLR_A


RF_5F
600
550
500
450
400
350
300
OBSERVED/kg
300
350
400
450
500
550
600
PREDICTED/kg
RF
RF_A

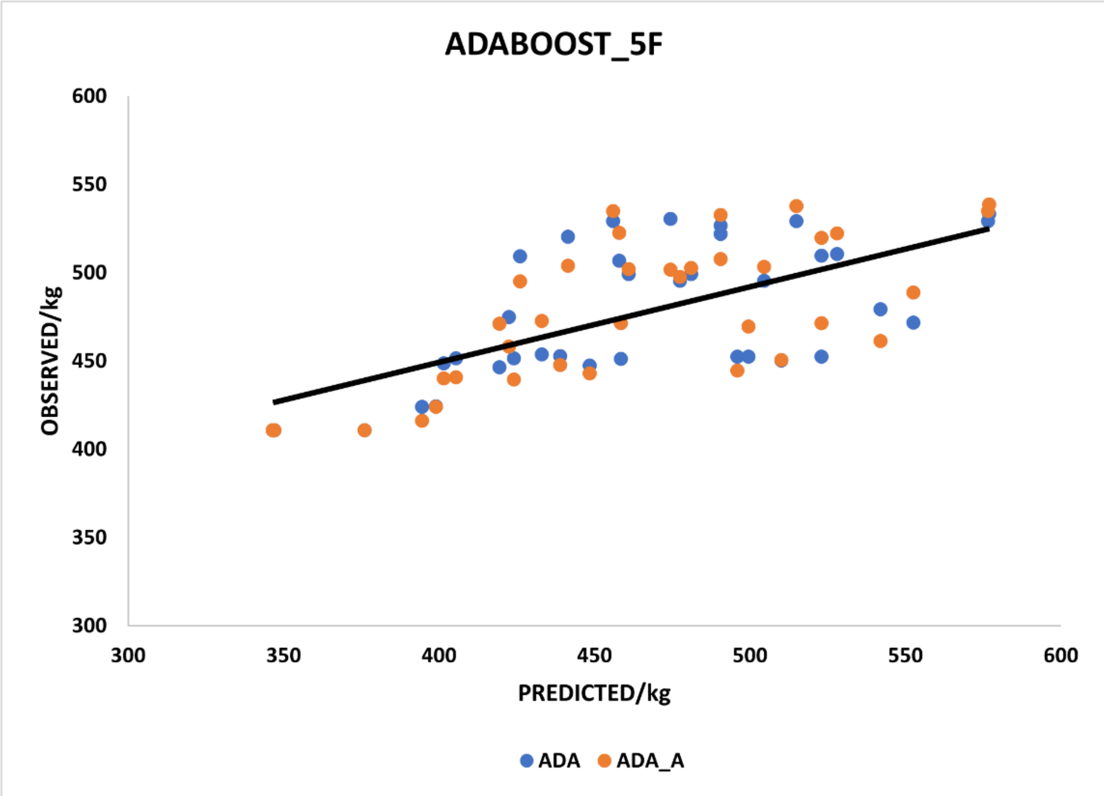
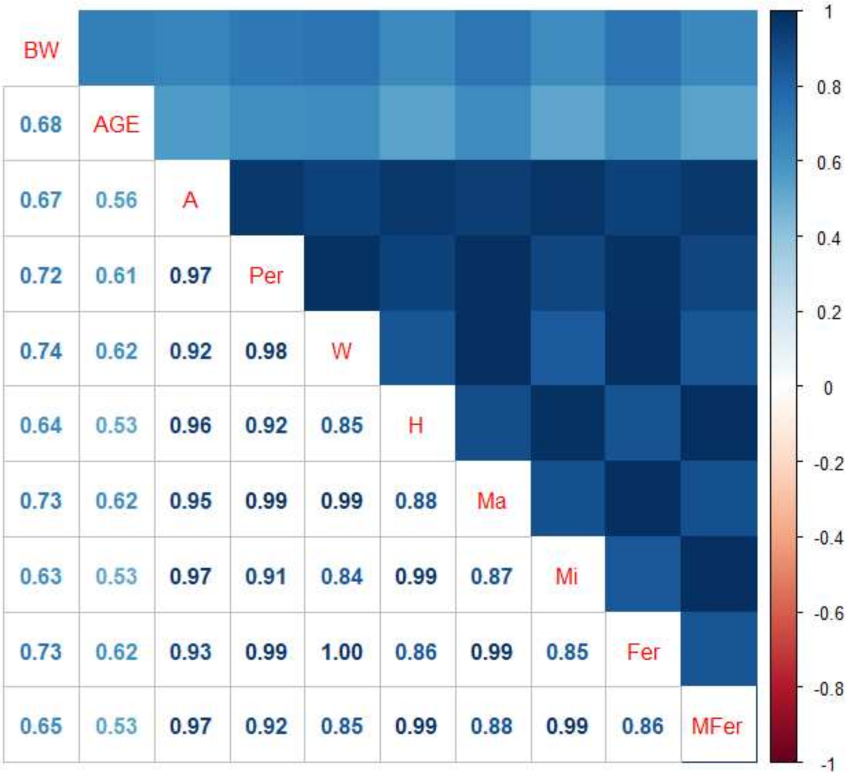
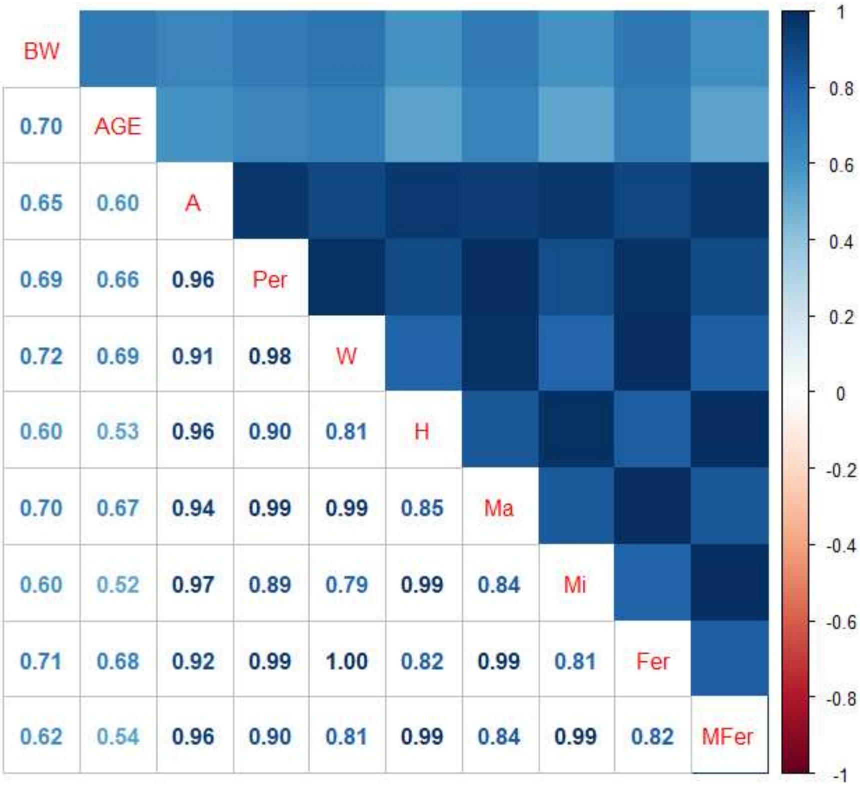


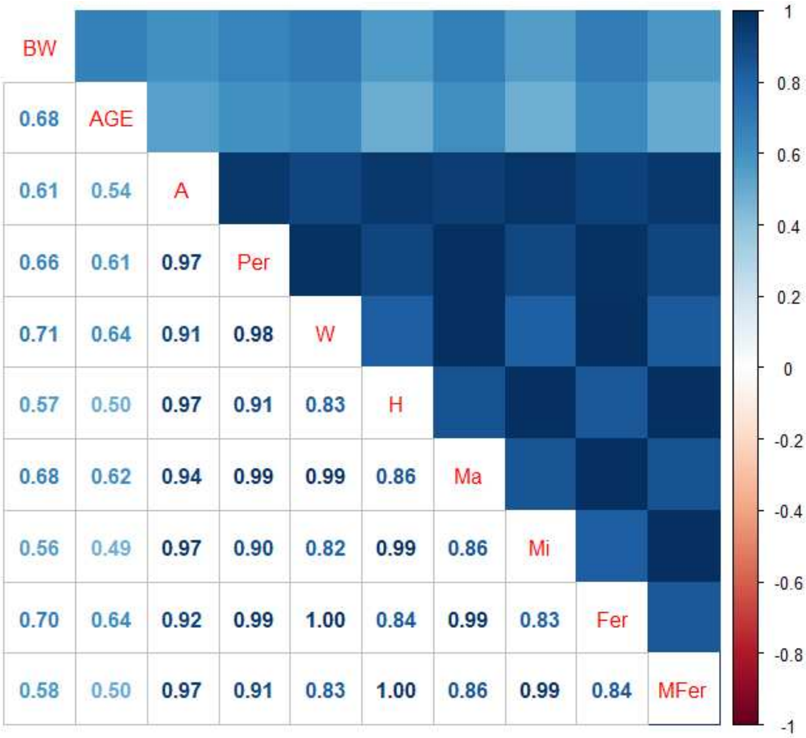
Fig 3. Observed vs. Predicted BW for each algorithm without and with AGE as independent variable and metrics results for 5F dataset. RIDGE: ridge using only shape descriptors; LASSO: LASSO: using only shape descriptors; ENET: ENET using only shape descriptors; MLR: MLR using only shape descriptors; ADAB: ADAB using only shape descriptors; RF: RF using only shape descriptors; RIDGE_A: ridge using AGE; LASSO_A: LASSO using AGE; ENET_A: ENET using AGE; MLR_A: MLR using AGE; ADAB_A: ADAB using AGE; RF_A: RF using AGE.



1F


BW
0.70 AGE
0.65 0.60 A
0.69 0.66 0.96 Per
0.72 0.69 0.91 0.98 W
0.60 0.53 0.96 0.90 0.81 H
0.70 0.67 0.94 0.99 0.99 0.85 Ma
0.60 0.52 0.97 0.89 0.79 0.99 0.84 Mi
0.71 0.68 0.92 0.99 1.00 0.82 0.99 0.81 Fer
0.62 0.54 0.96 0.90 0.81 0.99 0.84 0.99 0.82 MFer
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1

3F


BW
0.68 AGE
0.61 0.54 A
0.66 0.61 0.97 Per
0.71 0.64 0.91 0.98 W
0.57 0.50 0.97 0.91 0.83 H
0.68 0.62 0.94 0.99 0.99 0.86 Ma
0.56 0.49 0.97 0.90 0.82 0.99 0.86 Mi
0.70 0.64 0.92 0.99 1.00 0.84 0.99 0.83 Fer
0.58 0.50 0.97 0.91 0.83 1.00 0.86 0.99 0.84 MFer
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1

5F

Fig 4. Correlation between the dependent variable (BW) and the independent variables for body weight obtained through bidimensional images. 1F: 1 Frame Information, 3F: 3 Frames Information, 5F: 5 Frames Information.

SCRIPT

```
## Necessary libraries
library (imager)
library (dplyr)
## image input
Im = load.image('image.jpg')
Img2 = grayscale(im) ##grayscale
plot(imt) ##plotting

##Threshold segmentation
Imt= threshold(img2, "30%") %>% plot ## limiar segmentation

## Contour
Library (image.ContourDetector)
c= contours(imt) ## Edge detection
```