

Digital Phenotyping and Genomic Prediction Using Machine and Deep Learning in Animals and Plants

Ye Bi

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Animal and Poultry Sciences

Gota Morota, Chair

Mark D. Hanigan

Ismini Lourentzou

Nicholas Santantonio

September 18, 2024

Blacksburg, Virginia

Keywords: deep learning, digital phenotyping, genomic analysis, machine learning

Copyright 2024, Ye Bi

Digital Phenotyping and Genomic Prediction Using Machine and Deep Learning in Animals and Plants

Ye Bi

ABSTRACT

This dissertation investigates the utility of deep learning and machine learning approaches for livestock management and quantitative genetic modeling of rice grain size under climate change. Monitoring the live body weight of animals is crucial to support farm management decisions due to its direct relationship with animal growth, nutritional status, and health. However, conventional manual weighing methods are time consuming and can cause potential stress to animals. While there is a growing trend towards the use of three-dimensional cameras coupled with computer vision techniques to predict animal body weight, their validation with deep learning models as well as large-scale data collected in commercial environments is still limited. Therefore, the first two research chapters show how deep learning-based computer vision systems can enable accurate live body weight prediction for dairy cattle and pigs. These studies also address the challenges of managing large, complex phenotypic data and highlight the potential of deep learning models to automate data processing and improve prediction accuracy in an industry-scale commercial setting. The dissertation then shifts the focus to crop resilience, particularly in rice, where the asymmetric increase in average nighttime temperatures relative to the increase in average daytime temperatures due to climate change is reducing grain yield and quality in rice. Through the use of deep learning and machine learning models, the last two chapters explore how metabolic data can be used in quantitative genetic modeling in rice under environmental stress conditions such as high night temperatures. These studies showed that the integration of metabolites and

genomics provided an improvement in the prediction of rice grain size-related traits, and certain metabolites were identified as potential candidates for improving multi-trait genomic prediction. Further research showed that metabolic accumulation was low to moderately heritable, and genomic prediction accuracies were consistent with expected genomic heritability estimates. Genomic correlations between control and high night temperature conditions indicated genotype-by-environment interactions in metabolic accumulation and the effectiveness of genomic prediction models for metabolic accumulation varied across metabolites. Joint analysis of multiple metabolites improved the accuracy of genomic prediction by exploiting correlations between metabolite accumulation. Overall, this dissertation highlights the potential of integrating digital technologies and multi-omic data to advance data analytics in agriculture, with applications in livestock management and quantitative genetic modeling of rice.

Digital Phenotyping and Genomic Prediction Using Machine and Deep Learning in Animals and Plants

Ye Bi

GENERAL AUDIENCE ABSTRACT

This dissertation explores the application of deep learning and machine learning to computer vision-based livestock management and quantitative genetic modeling of rice grain size under climate change. The first half of the research chapters illustrate how computer vision systems can enable digital phenotyping of dairy cows and pigs, which is critical for informed management decisions and quantitative genetic analysis. These studies address the challenges of managing large-scale, complex phenotypic data and highlight the potential of deep learning models to automate data processing and improve prediction accuracy. Chapter 3 showed that a deep learning-based segmentation, Mask R-CNN, improved the prediction performance of cow body weight from longitudinal depth video data. Among the image features, volume followed by width correlated best with body weight. Chapter 4 showed that efficient deep learning-based supervised learning models are a promising approach for predicting pig body weight from industry-scale depth video data. Although the sparse design, which simulates budget and time constraints by using a subset of the data for training, resulted in some performance loss compared to the full design, the Vision Transformer models effectively mitigated this loss. The second half of the research chapters focuses on integrating metabolomic and genomic data to predict grain traits and metabolic content in rice under climate change. Through the use of machine learning models, these studies investigate how combining genomic and metabolic data can improve predictions, particularly under high night temperature stress in rice. Chapter 5 showed that the integration of metabolites and ge-

nomics improved the prediction of rice grain size-related traits, and certain metabolites were identified as potential candidates for improving multi-trait genomic prediction. Chapter 6 showed that metabolic accumulation was low to moderately heritable. Genomic correlations between control and high night temperature conditions indicated genotype-by-environment interactions in metabolic accumulation, and the effectiveness of genomic prediction models for metabolic accumulation varied across metabolites. Joint analysis of multiple metabolites improved the accuracy of genomic prediction by exploiting correlations between metabolite accumulation. Overall, the dissertation provides insight into how cutting-edge methods can be used to improve livestock management and multi-omic quantitative genetic modeling for breeding.

Dedicated to my parents and my husband.

Acknowledgments

Pursuing a Ph.D. has been a long journey for me. Over the past few years, I have explored many different research directions, but I have found that genuine interest is what really drives me to achieve my goals. I am deeply grateful to those who have supported me along the way. First, I am deeply grateful to my advisor, Dr. Gota Morota. I thank him for introducing me to the area of quantitative genetics and precision livestock farming, helping me realize my potential in research, and guiding me to discover my interests in animal sciences. Dr. Morota's creative insights and keen intuition have guided me through many challenging times in my research. He always approached my mistakes with kindness and patience, offering constructive feedback to help me improve. I still remember the countless hours he spent explaining complex concepts through numerous meetings, meticulously edited manuscripts, and more. His extensive support and guidance have been invaluable throughout my academic journey. He is not only an exceptional advisor, but also a trusted friend.

I sincerely appreciate the support of my committee members, Drs. Mark Hanigan, Ismini Lourentzou, and Nicholas Santantonio. Their invaluable insights and constructive feedback have been instrumental in the development of my research.

I am also thankful to Dr. Jianhua Xuan, who taught me the fundamentals of deep learning and computer vision and provided valuable suggestions for my research projects and guidance for my future career. I am also grateful to Drs. Xiaowei Wu, Song Li, and Anuj Karpatne, who introduced me to the fundamentals of statistics and data science in genomics, machine learning, and computer vision. The courses they taught were fundamental in shaping my theoretical understanding and analytical skills.

Special thanks to Drs. Harkamal Walia and Toshihiro Obata for providing the rice dataset

and their continued support of my research. I would also like to thank Dr. Haipeng Yu for providing knowledge on deep learning models and career development suggestions.

I would like to extend my gratitude to Drs. Yijian Huang and Di Liang for introducing me to the industry, which has been instrumental in expanding my professional network and opportunities.

Special thanks to my colleagues and friends at Virginia Tech for their unwavering support in teamwork, coursework, data collection, and animal management, as well as their invaluable friendship throughout my journey. Each of these individuals has played a significant role in my academic and personal development, and I am deeply grateful for their help and encouragement.

Above all, I would like to express my deepest gratitude to my parents for their unwavering and selfless support throughout my journey. Their belief in and devotion to me have been a constant source of strength. I am also deeply grateful to my husband, Dr. Jiabin Yu. His passion, love, and companionship have helped me get through my most challenging moments. His support has been a beacon of hope and resilience, without him, I would never have had the courage to pursue my Ph.D. dream. Although it may sound cliché, I truly consider myself lucky to have him in my life.

Contents

List of Figures	xv
List of Tables	xxvii
List of Abbreviations	xxix
1 Introduction	1
1.1 Background	1
1.2 Outline of dissertation	2
2 Literature Review	5
3 Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms	10
3.1 Abstract	10
3.2 Background	12
3.3 Materials and methods	15
3.3.1 Animal experiments	15
3.3.2 Video acquisition setup	15
3.3.3 Image segmentation	16

3.3.4	Biometric features	25
3.3.5	Body weight regression model	27
3.3.6	Model goodness of fit	30
3.3.7	Cross-validation strategies	30
3.4	Results	33
3.4.1	Pearson correlation analysis	33
3.4.2	Model goodness of fit	35
3.4.3	Cross-validation	36
3.5	Discussion	38
3.6	Conclusions	43
3.7	Code availability	44
4	Industry-scale prediction of video-derived pig body weight using efficient convolutional neural networks and vision transformers	45
4.1	Abstract	45
4.2	Background	47
4.3	Materials and Methods	51
4.3.1	Animals and data collection design	51
4.3.2	Video acquisition setup	52
4.3.3	Video preprocessing	53
4.3.4	Deep learning models	56

4.3.5	Cross-validation strategy	62
4.4	Results	64
4.4.1	Single time point cross validation	65
4.4.2	Time series forecasting cross validation	70
4.5	Discussion	74
4.5.1	Influence of video preprocessing for model training on large-scale data	75
4.5.2	Utility of efficient CNN and ViT	76
4.5.3	Impact of sparse design	79
4.5.4	Current limitations and future studies	79
4.6	Conclusions	81
5	Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice	82
5.1	Abstract	82
5.2	Introduction	84
5.3	Materials and Methods	86
5.3.1	Plant materials and growth conditions	86
5.3.2	Metabolic profiling	87
5.3.3	Statistical analyses	87
5.3.4	Subpopulation structure	95
5.4	Results	97

5.4.1	Correlation analysis	97
5.4.2	Evaluation of metabolic classification performance	99
5.4.3	Evaluation of metabolic prediction performance	100
5.4.4	Evaluation of genomic prediction performance	106
5.4.5	Estimates of metabolic and genomic heritability	108
5.4.6	Evaluation of multi-trait genomic prediction performance	108
5.4.7	The impact of subpopulation structure	112
5.5	Discussion	113
5.5.1	Utility of metabolites for classification	114
5.5.2	Utility of metabolites for prediction and heritability analysis compared to SNP markers	115
5.5.3	Utility of metabolites as auxiliary phenotypes in genomic prediction .	116
5.6	Conclusions	118
5.6.1	Data availability	118
6	Genomic prediction of metabolic content in rice grain in response to warmer night conditions	119
6.1	Abstract	119
6.2	Introduction	121
6.3	Materials and Methods	123
6.3.1	Plant materials and growth conditions	123

6.3.2	Metabolic profiling	123
6.3.3	Genomic heritability of metabolite accumulation	124
6.3.4	Single-trait genomic prediction of metabolites	125
6.3.5	Genomic correlation between the same metabolite accumulation in different treatments	126
6.3.6	Exploratory factor analysis	127
6.3.7	Simultaneous regression modeling of metabolites	128
6.3.8	Deep learning models	130
6.4	Results	132
6.4.1	Single-trait analysis of metabolites	132
6.4.2	Latent factors behind metabolites	137
6.4.3	Single-trait analysis using non-additive models	138
6.4.4	Joint analysis of multiple metabolites	140
6.5	Discussion	143
6.6	Conclusions	147
6.7	Data availability	148
7	Conclusion	149
	Appendices	153
	Appendix A Depth video data-enabled predictions of longitudinal dairy cow	

body weight using thresholding and Mask R-CNN algorithms	154
Appendix B Industry-scale prediction of video-derived pig body weight using efficient convolutional neural networks and vision transformers	157
Appendix C Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice	169
Appendix D Genomic prediction of metabolic content in rice grain in response to warmer night conditions	184
Bibliography	190

List of Figures

3.1	Image conversion process. A) The original depth image. B) Cropped depth image. C) The RGB color model was converted to the HSV color model. D) Hue was extracted and used for subsequent analyses.	18
3.2	Neck removal process. A) Binary image after thresholding. B) The initial segmentation that includes the neck. C) The ratio of body width was used to remove the neck. Blue line represents the column location that returns the maximum number of white pixels across all columns. Red line represents the column location that provides the ratio of 0.3 relative to the blue line. D) The final cow body area with the bounding box. The cyan box, yellow line, and blue line denote the bounding box, dorsal length, and abdomen width, respectively.	19
3.3	Mask R-CNN architecture used in the current study. conv1: initial convolutional block. conv_x: multiple residual convolutional blocks. C2-C5: feature maps generated from each residual block. P1-P5: feature maps generated after the feature pyramid network (FPN). RPN: regional proposal network. RoI: region of interest. FCN: fully connected neural network. Bbox reg: bounding box regression.	22
3.4	Overview of the biometric approach using thresholding.	25
3.5	Overview of the biometric approach using Mask R-CNN.	26

3.6	Three-dimensional visualization of the cow volume. The volume of the cow was obtained by integrating the pixels of the cow area over its depth.	28
3.7	Cross-validation design for body weight prediction. Time series forecasting (A). 90:10: the first 90% of time points were used to predict the remaining 10%. 80:20: the first 80% of time points were used to predict the remaining 20%. 70:30: the first 70% of time points were used to predict the remaining 30%. 60:40: the first 60% of time points were used to predict the remaining 40%. 50:50: the first 50% of time points were used to predict the remaining 50%. leave-three-cows-out (B). Of 12 cows, 9 and 3 cows were used as the training and testing sets, respectively.	32
3.8	Scale-based cow body weight records across time points. ID 1-10: lactating Holstein cows. ID 11-12: non-lactating Jersey cows.	33
3.9	Pearson correlation heat map between scaled-based body weight, length, width, centroid height, average height, and volume averaged across time points. A) single-thresholding, B) adaptive-thresholding, and C) Mask R-CNN.	35
3.10	Leave-three-cows-out cross-validation results. A) Prediction coefficient of determination. B) Mean absolute percentage error. ST: single-thresholding. AT: adaptive-thresholding. MRCNN: Mask R-CNN. OLS: Ordinary least squares. RR: Ridge regression. LASSO: Least absolute shrinkage and selection operator.	39

3.11	Range of body weight records observed in training and testing set partitioning for leave-three-cows-out cross-validation using Mask R-CNN and ordinary least squares. A) top 4 predictions based on the coefficient of determination. B) bottom 4 predictions based on coefficient of determination. C) top 4 predictions based on mean absolute percentage error. (D) bottom 4 predictions based on mean absolute percentage error.	43
4.1	Histogram of scale-based pig body weight records. T1 to T6 represent six time points analyzed in this study.	52
4.2	YOLOv8 detection examples. (A) No detection: the pig is not in the center of the image. (B) Pig detection rate of 0.89: acceptable image because the pig is in the center of the image. (C) Pig and block detection: both the pig and the block are in the image. (D) Pig and block detection: both the pig and the block are in the image, but the pig is not in the center of the image. (E) No detection: multiple pigs in the image. (F) Block detection rate of 0.63: Only block is detected. (G) Pig and block detection: the camera is positioned in the opposite direction. (H) Pig detection rate of 0.90: acceptable image because the pig is in the center of the image but the camera is positioned in the opposite direction.	55
4.3	Number of images for each pig across six farm visits. T1 to T6 represent six time points analyzed in this study. The dotted red line represents the number of images of 20.	56

4.4	Overview of the deep learning workflow for body weight prediction. Depth videos were processed into images and comma-separated values (CSV) files. YOLOv8 filtered out poorly positioned or obstructed images. The quality-controlled CSV files were converted to grayscale images, padded, and resized. Neural networks are then trained and evaluated using four metrics. MAPE: mean absolute percentage error; R^2 : coefficient of determination; MAE: mean absolute error; and RMSE: root mean squared error.	57
4.5	Time series forecasting cross-validation design. Subsets of pigs from earlier and most recent time points (shown in blue) were used to train models to predict the body weight of pigs at the most recent time points (shown in red). Five scenarios (A to E) were considered for predicting pigs from T2 to T6. T1 to T6 represent six farm visits.	64
4.6	Single time point cross-validation prediction results. Thirteen models were tested: three from the MobileNet series, three from the MobileNetV3 series, two from the MobileViT series, four from the MobileViT V2 series, and ResNet50 as the baseline model. The plots show the test mean absolute percentage error over six different time points (T1 to T6) for each model. . .	66
4.7	Single time point prediction results in the testing sets using the best model at each visit, along with the overall performance. The plots compare the predicted weight with the observed weight (ground truth). The red line represents the slope of 1. T1 to T6 represent six farm visits. RMSE: root mean squared error; R^2 : coefficient of determination; MAPE: mean absolute percentage error; and MAE: mean absolute error.	68

4.8	Single time point prediction results for body weight subgroups using the best model at each visit. The mean absolute percentage error (MAPE) on the test set for 10 kg weight intervals, along with the number of images in the training and testing sets. T1 to T6 represent six farm visits.	69
4.9	Gradient-weighted class activation mapping heatmaps for arbitrarily selected A) light, B) average, and C) heavy pigs from the T5 visit, generated with MobileViT-S.	70
4.10	Time series forecasting cross-validation prediction results. 0.25, 0.5, and 0.75 represent scenarios for the subset ratios 25%:75%, 50%:50%, and 75%:25%. Here, 75%:25% indicates that 75% of the data points from all previous time points were used as the training set to predict 25% of the current data points. The red squares represent full design cross-validation, and the bars represent sparse design cross-validation results.	73
5.1	Cross-validation (CV) design for binary classification of high night temperature stress conditions (A) and metabolomic and genomic prediction of grain size related phenotypes (B).	93
5.2	Cross-validation design for multi-trait prediction. Scenario 1 (A) and Scenario 2 (B).	96
5.3	Correlation between metabolites and grain size phenotypes in control (A) and high night time temperature stress (B) conditions.	98
5.4	Classification accuracy of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; RF: random forest; and XGBoost: extreme gradient boosting. . . .	100

5.5	Predictive correlations of grain length using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.	102
5.6	Predictive correlations of grain width using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.	103
5.7	Predictive correlations of grain perimeter using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.	104

5.8	Predictive correlations of grain length, grain width, and grain perimeter using metabolic, genomic, and multi-omic prediction models in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MBLUP: metabolic best linear unbiased prediction; GBLUP: genomic best linear unbiased prediction; and GMBLUP: genomic and metabolic best linear unbiased prediction.	107
5.9	Predictive correlations of Scenario 1 multi-trait (bivariate) genomic prediction for grain length, grain width, and grain perimeter when metabolites were used as a secondary phenotype under control and high night time temperature stress conditions. The horizontal lines indicate the predictive correlations of single-trait genomic prediction.	110
5.10	Predictive correlations of Scenario 2 multi-trait (bivariate) genomic prediction for grain length, grain width, and grain perimeter when metabolites were used as a secondary phenotype under control and high night time temperature stress conditions. The horizontal lines indicate the predictive correlations of single-trait genomic prediction.	111
6.1	Flowchart of converting single nucleotide polymorphisms to image data.	132

6.2	Genomic heritability estimates of metabolite accumulation in control and high night temperature stress conditions. A) Scatter plot. B) Density plot. Solid and dashed lines indicate mean and median, respectively. C) Agreement of heritability estimates between control and high night temperature stress conditions. Metabolite accumulation in green and red colors indicate that the heritability difference between control and high night temperature stress conditions was small (< 0.05) and large (> 0.1).	134
6.3	Genomic correlation estimates between the same metabolite accumulation measured under control and high night temperature stress conditions. A) Scatter plot. B) Bar chart. Solid and dashed lines indicate mean and median, respectively.	135
6.4	Genomic prediction accuracy of metabolite accumulation in control and high night temperature stress conditions. A) Box plot. The horizontal line indicates the mean value. B) Density plot. The solid and dashed lines indicate the mean and median, respectively. C) Agreement of genomic prediction accuracy between control and high night temperature stress conditions. Metabolite accumulation in green and red colors indicate that the genomic prediction difference between control and high night temperature stress conditions was small (< 0.05) and large (> 0.1).	136
6.5	Percentage difference of gain in prediction accuracy for single-trait reproducing kernel Hilbert spaces regression (RKHS), VGG16, ResNet50, EfficientNetB7, InceptionV3, MobileNetV2, and DenseNet201 relative to single-trait genomic best linear unbiased prediction.	139

6.6	Example of a set of single nucleotide polymorphisms transformed into image data for a randomly selected genotype. Images of 12 chromosomes were processed in the multi-channel convolutional neural networks.	140
6.7	Genomic correlation estimates between different metabolite accumulation in control and high night temperature stress conditions. The solid and dashed lines indicate mean and median, respectively.	141
6.8	Percentage difference of gain in prediction accuracy for multi-trait genomic best linear unbiased prediction (MegaLMM-G) and multi-trait reproducing kernel Hilbert spaces regression (MegaLMM-GK) relative to single-trait genomic best linear unbiased prediction (A). Density plots of percentage difference are shown for MegaLMM-G (B) and MegaLMM-GK (C).	142
A.1	Per day Pearson correlation heat map between scaled-based body weight and biometric features (length, width, centroid height, average height, and volume). A) single-thresholding, B) adaptive-thresholding, and C) Mask R-CNN.	155
A.2	Per day per AM/PM Pearson correlation heat map between scaled-based body weight and biometric features (length, width, centroid height, average height, and volume). A) AM data using single-thresholding, B) PM data using single-thresholding C) AM data using adaptive-thresholding, D) PM data using adaptive-thresholding, E) AM data using Mask R-CNN, and F) PM data using Mask R-CNN.	156
B.1	Single time point cross-validation training history using the best model at each visit. T1 to T6 represent six farm visits. Blue and orange curves represent the mean square error in the training and validation sets, respectively. . . .	167

B.2	Gradient-weighted class activation mapping heatmaps for the top 50 prediction results from the T5 visit, generated with MobileViT-S.	168
C.1	Pearson correlation heat map among metabolic profiles in control (A) and high night time temperature stress conditions (B).	174
C.2	Pearson correlation heat map among metabolic profiles expressed as the ratio of control and high night time temperature stress.	175
C.3	F1 scores of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; and RF: random forest; XGBoost: extreme gradient boosting.	176
C.4	Area under the curve of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; and RF: random forest; XGBoost: extreme gradient boosting. . . .	177
C.5	Classification accuracy of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost). . .	178
C.6	F1 scores of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost).	179
C.7	Area under the curve of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost). . .	180

C.8	Metabolic predictions across high night time temperature (HNT) stress conditions. Control: Train the model using phenotypes and metabolites in control and predict control phenotypes from metabolites in control conditions. Control-HNT: Train the model using phenotypes and metabolites in control and predict HNT phenotypes from metabolites in HNT conditions using the metabolic effect estimated in control conditions. HNT: Train the model using phenotypes and metabolites in HNT and predict HNT phenotypes from metabolites in HNT conditions. HNT-Control: Train the model using phenotypes and metabolites in HNT and predict phenotypes from metabolites in control conditions using the metabolic effect estimated in HNT condition. . .	181
C.9	Principal component analysis plot based on the genomic relationship matrix.	182
C.10	Principal component analysis plot based on the metabolic relationship matrix in control (A) and high night time temperature stress conditions (B).	183
D.1	Pearson correlation heat maps between metabolic profiles in control (A) and high night temperature stress (B).	186
D.2	Heat maps of factor loadings after removing cross-loading by setting a cutoff of $\lambda > 0.4 $. The rows of each panel correspond to the observed metabolite accumulation and the columns correspond to the five latent factors in control (A) and high night temperature stress conditions (B).	187
D.3	Genomic heritability estimates of metabolite contents in control and high night temperature stress conditions (A) and genomic prediction accuracy of metabolite contents in control and high night temperature stress conditions, separated by latent factors (B). F1-F5 and F6-F10 are latent factors in control and high night temperature stress conditions, respectively.	188

D.4 Genomic correlation estimates between different metabolites in control (A)	
and high night temperature stress (B).	189

List of Tables

3.1	Model goodness of fit.	36
3.2	Forecasting cross-validation results.	38
4.1	Hyperparameters of deep learning models.	61
4.2	Single time point prediction results in the testing set for 10 kg weight intervals.	71
5.1	Metabolic heritability (h_M^2) and genomic heritability (h_G^2) of grain size phenotypes in control and high night temperature stress conditions.	109
B.1	Single time point cross-validation results for deep learning models.	157
B.1	Single time point cross-validation results for deep learning models.	158
B.1	Single time point cross-validation results for deep learning models.	159
B.1	Single time point cross-validation results for deep learning models.	160
B.1	Single time point cross-validation results for deep learning models.	161
B.2	Time series forecasting cross-validation results for deep learning.	162
B.2	Time series forecasting cross-validation results for deep learning.	163
B.2	Time series forecasting cross-validation results for deep learning.	164
B.2	Time series forecasting cross-validation results for deep learning.	165
B.2	Time series forecasting cross-validation results for deep learning.	166

C.1	List of metabolites.	170
C.2	Conventional heritability estimates of grain size phenotypes in control and high night temperature (HNT) stress conditions.	171
C.3	Predictive correlations of grain-size phenotypes using metabolic, genomic, and multi-omic predictions in control and high night time temperature stress conditions.	172
C.4	Predictive correlations of leave-one-subpopulation-group-out cross-validation in control and high night temperature stress conditions.	173
D.1	List of metabolites analyzed in the current study.	185

List of Abbreviations

3D Three-dimensional

AUC Area under curve

BW Body weight

CNN Convolutional neural networks

CSV Comma-separated values

CV Computer vision

DL Deep learning

FPN Feature pyramid network

GBLUP Genomic best linear unbiased prediction

GK Gaussian kernel

GMBLUP Integrated genomic metabolic best linear unbiased prediction

Grad-CAM Gradient-weighted Class Activation Mapping

HDRA High-density rice array

HNT High night temperatures

HSV Hue, saturation, and value

LASSO Least absolute shrinkage and selection operator

LMM Linear mixed model

LR Logistic regression

MAE Mean absolute error

MAPE Mean absolute percentage error

MBLUP Metabolic best linear unbiased prediction

MegaLMM-G Multi-trait genomic best linear unbiased prediction

ML Machine Learning

MOLS Metabolic OLS

OLS Ordinary least squares

PLF Precision livestock farming

R² Coefficient of determination

RF Random forest

RKHS Reproducing kernel Hilbert spaces regression

RMSE Root mean squared error

RoI Region of interest

RoIAlign Region of interest align

RPN Region proposal network

RSS Residual sum of squares

SNP Single nucleotide polymorphism

SVM Support vector machines

T6P Trehalose-6-phosphate

ViT Vision Transformers

XGBoost Extreme gradient boosting

Chapter 1

Introduction

1.1 Background

Effective livestock breeding programs rely heavily on the systematic collection of comprehensive and accurate phenotypic data, which is crucial for genomic analysis of complex traits. Genetic improvement in livestock is directly influenced by the quality and quantity of phenotypic data collected [148, 164]. Precision livestock farming (PLF) is a transformative approach that uses digital technologies, such as computer vision systems, to facilitate the phenotyping and monitoring of large numbers of animals [120]. The integration of digital image analysis through these computer vision systems allows the extraction of valuable information from images, providing insights for a better understanding of critical traits, including morphological changes, growth dynamics, and activity status of animals [33]. This advanced technology supports a variety of image processing techniques that increase the precision and depth of phenotypic data, thereby improving the effectiveness of breeding programs [124]. Despite these advancements, digital phenotyping also presents several challenges, particularly in the management and processing of large data sets that require thorough quality control, cleaning, annotation, and model evaluation [169]. The big data problem of digital phenotyping also arises from its inherent high heterogeneity and time dependence. Unlike homogeneous genomic data, phenotypic measurements can vary widely and change rapidly over time [136].

Genomic prediction has been widely used to predict the responses of plants and animals [113]. It is a powerful quantitative genetic approach to predict the genetic value of unphenotyped lines for diverse arrays of traits in crops and animals [7]. Recent advances in biotechnology have introduced new omic layers that enrich our understanding of these traits. Multi-omic data integration, such as the combination of genomics and metabolomics, offers a novel approach to building quantitative genetic models, particularly for complex traits that are influenced by climate change. For example, in rice, the asymmetric increase in nighttime temperatures due to climate change has been shown to reduce grain yield and quality [68, 71, 180]. However, the integration of multi-omics data requires novel computational tools and algorithms to effectively extract meaningful insights from the many layers of biological information available.

To manage and interpret these complex datasets, including digital images and multi-omic datasets, the application of machine learning and deep learning models has become increasingly important. These models can automate the analysis of large amounts of data, identify patterns, and generate predictions [94, 158].

This dissertation investigates the utility of machine learning and deep learning methods for digital image and multi-omic data analysis. The overarching theme of this dissertation is to show how these novel approaches can be used to enhance phenotypic image analysis and quantitative genetic modeling for better animal management and the development of more climate-resilient crop varieties.

1.2 Outline of dissertation

Chapter 1 presents the background and motivation for the dissertation and also outlines the organization of this dissertation.

Chapter 2 provides a review of the relevant literature while identifying gaps in current research that this dissertation aims to address.

Chapter 3 focuses on the use of depth video data to predict dairy cow body weight, an important aspect of farm management. Traditional methods of body weight measurement, often using single time point data, do not take into account the dynamic nature of body weight changes. This chapter investigates the potential of using longitudinal video data collected by an Intel RealSense D435 camera and examines the effectiveness of different segmentation approaches, including thresholding and Mask R-CNN algorithms. By comparing these techniques, the study shows how deep learning-based segmentation can improve the accuracy of body weight predictions, providing dairy producers with a robust tool for herd monitoring, management, and phenotyping.

Chapter 4 shifts the focus to predicting pig body weight in an industry-scale commercial setting. Accurate weight measurement in pigs is critical for optimizing growth monitoring and health management, but traditional methods can be labor-intensive and stressful for the animals. This chapter evaluates the performance of efficient convolutional neural networks and vision transformers applied to video data collected in a commercial environment. Through a series of validations, including random repeated subsampling and time series forecasting, the study demonstrates how novel deep learning models can improve the prediction of pig body weight, providing a potentially practical solution for large-scale operations.

Chapter 5 presents the analysis of metabolic and genomic data in rice to address the challenges posed by high nighttime temperatures. Increases in nighttime temperatures due to climate change are known to adversely affect rice grain yield and quality. This chapter investigates how metabolic profiles and single nucleotide polymorphisms can be used to predict grain traits, such as length, width, and perimeter, under high night temperature conditions. The study highlights the effectiveness of integrating metabolic and genomic data to improve

prediction accuracy and provides insights into the use of metabolite accumulation measures in quantitative genetic analysis.

Chapter 6 continues the exploration of joint metabolic and genomic analysis, focusing on genomic prediction of metabolic content in rice grains under warmer nighttime conditions. This chapter investigates how well genomic markers can explain variation in metabolic profiles and examines the influence of genotype-by-environment interactions. The study uses novel statistical and deep learning methods to model the genome-metabolite relationship, providing valuable information on the genetic control of metabolic variation and its implications for crop improvement under changing climate conditions.

In summary, this dissertation integrates cutting-edge technologies and methodologies to address pressing issues in livestock management and crop improvement. By leveraging deep learning and machine learning-based computer vision and multi-omic data analysis, the research provides possible practical solutions for improving management practices and crop varieties in the face of sustainability and environmental challenges.

Chapter 2

Literature Review

This literature review focuses on the application of computer vision technologies for predicting animal body weight, highlighting advancements in 3D depth sensor cameras, segmentation techniques, and deep learning models while addressing the challenges of accuracy, practicality, and the need for open science in livestock management.

Regular monitoring of body weight (BW) is a key aspect of livestock management and plays a vital role in optimizing production efficiency and ensuring animal health and welfare. In dairy cattle, daily variations in BW provide important insights into several physiological and management factors, including milk production, feed efficiency, energy balance, growth, and overall health. For example, a negative energy balance, often reflected in a loss of BW, may indicate inadequate nutrition or underlying health and reproductive issues that require immediate attention. Therefore, individualized monitoring of BW throughout a cow's lifespan is essential for making informed management decisions that can improve animal productivity and well-being. However, traditional methods of BW monitoring, such as manual weighing or the use of automated walk-over weighing systems, are often labor-intensive, time-consuming, and stressful for the animals, making them impractical for routine use on large-scale farms [43].

Similarly, in swine production, variations in BW between pigs, particularly during the critical growing-finishing phase, present significant challenges to production efficiency. Pigs with slower growth rates may reach market weight later than their faster growing counterparts,

resulting in reduced income and increased production costs for producers. These economic impacts are largely related to meat quality grading and carcass pricing, which are directly influenced by the accuracy of BW measurements [31, 88]. Regular and continuous monitoring of BW in pigs not only helps to make timely marketing decisions, but also facilitates early detection of health issues, such as digestive disorders, that can hinder growth and development [28, 100]. Despite its importance, maintaining accurate and continuous records of individual pig weights in commercial production environments is often fraught with challenges, including the laborious nature of the process, time constraints, and the risk of increased animal stress and injury during manual weighing [43, 153].

The growing trend towards larger farm sizes and the increasing complexity of modern livestock operations have underscored the need for innovative, non-invasive, and cost-effective methods of BW monitoring. Precision livestock farming has emerged as a promising approach to address these challenges, leveraging advancements in information science and computer technology to monitor and manage animals in real-time [33]. Among the various technologies used in precision livestock farming (PLF), computer vision (CV) systems, particularly those using low-cost, non-contact, stress-free, three-dimensional (3D) depth sensor cameras, have shown significant potential for accurately assessing morphometric traits of livestock, including BW. These systems provide a non-invasive alternative to traditional weighing methods, enabling continuous monitoring of animals without the need for physical restraint or direct contact [124]. The integration of CV into PLF aligns with the broader goals of the field, which seeks to improve farm management through individualized monitoring and management of animals, ultimately improving productivity and animal welfare [120].

Early applications of CV often relied on the simultaneous use of multiple 2D cameras to capture images from different angles, requiring detailed calibration processes to ensure synchronization [89, 173]. While this approach produced reasonable results, it was complex and

time-consuming. The advent of 3D depth sensor cameras has revolutionized this process by providing additional depth information that captures the distances between the camera and objects, along with the standard two-dimensional data. This added depth dimension enables the generation of height- and volume-derived 3D images, which have been shown to significantly improve the accuracy of BW predictions. Research has shown that models using these 3D images have superior goodness-of-fit and cross-validation performance compared to those based solely on 2D data [51, 86, 90, 104, 123, 167, 191]. Thus, the use of 3D depth cameras has gained traction as an effective means of capturing real-time growth curves with minimal impact on animal welfare. Previous studies have explored different camera positioning strategies, including top-and side-views, to capture depth data suitable for BW estimation [32, 53, 59, 91, 132]. The primary focus has been on top-view images because they do not interfere with daily farm operations. However, most existing studies have focused on data collected at a single point in time, which limits the ability to evaluate the predictive performance of models over time and under varying conditions.

After acquiring depth images, the next step is to process these images using advanced image processing algorithms that analyze the extracted image features to accurately estimate body weight. Two dominant image processing methods have been proposed: semantic segmentation and supervised neural networks. Semantic segmentation methods aim to extract morphological measurements from segmented images, such as length, width, height, volume, chest depth, and curvature, which are then fitted into prediction models typically based on linear or non-linear regression or machine learning [138, 202]. The accuracy of these methods is highly dependent on the performance of the segmentation algorithms, with factors such as image obstructions, illumination variations, and the choice of segmentation techniques all playing a critical role [39, 108]. However, the potential of deep segmentation techniques, such as Mask R-CNN for instance segmentation, object detection, and classification in ani-

mal BW prediction, has yet to be fully explored [9, 57, 143, 157, 193, 195]. Also, supervised neural networks, including convolutional neural networks (CNN) and vision transformers (ViT), have become increasingly influential in PLF. Unlike traditional image processing methods that require extensive preprocessing and feature extraction, deep learning models can directly use raw image data, to automatically learn the most relevant features for BW prediction. Studies using CNN have shown that top-view depth images, can provide highly accurate BW predictions even without preprocessing [15, 76, 108]. ViT, which divides images into patches and incorporates positional embeddings with multiple self-attention blocks, has also demonstrated state-of-the-art performance in various computer vision tasks, including semantic segmentation, object detection, and image classification [30, 46, 98]. However, the application of ViT in BW prediction remains limited, primarily due to the challenges of acquiring diverse datasets from commercial farm environments.

Modern state-of-the-art neural networks typically involve more complex and deep designs, that require substantial computational resources, making them impractical for many on-farm applications [20, 58, 66, 165, 170, 171]. Therefore, the focus must be on applying more efficient architectures that can provide accurate BW predictions with fewer modalities and faster processing time, facilitating local data processing. For PLF applications, Afridi et al. [2] applied MobileNet to RGB, depth, fully segmented, and fully segmented depth images for cattle body weight prediction and [203] applied MobileNetV2 to pig depth images for body weight and size prediction. The application of MobileViT for facial and coat recognition in dairy cattle and sheep is demonstrated in the literature [92, 181]. However, the application of MobileViT for predicting body weight in livestock remains scarce. In addition, few studies have compared the feature extraction performance of efficient CNN and ViT.

Several challenges remain that hinder the broader adoption and refinement of the above technologies. A significant barrier is the lack of open source code in animal science, particularly

in cross-validation-based BW prediction research, which limits transparency, reproducibility, and collaboration within the scientific community [121]. Open science, which advocates making research products and processes available to all, is critical to overcoming these barriers [125]. Promoting the adoption of open source practices in animal science, along with further research to refine deep learning models and explore their longitudinal application, is essential for the continued advancement of PLF and its impact on livestock management and phenotyping.

Chapter 3

Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms

3.1 Abstract

Monitoring cow body weight is crucial to support farm management decisions due to its direct relationship with the growth, nutritional status, and health of dairy cows. Cow body weight is a repeated trait, however, the majority of previous body weight prediction research only used data collected at a single point in time. Furthermore, the utility of deep learning-based segmentation for body weight prediction using videos remains unanswered. Therefore, the objectives of this study were to predict cow body weight from repeatedly measured video data, to compare the performance of the thresholding and Mask R-CNN deep learning approaches, to evaluate the predictive ability of body weight regression models, and to promote open science in the animal science community by releasing the source code for video-based body weight prediction. An Intel RealSense D435 camera was installed on Virginia Tech dairy complex to collect top-view videos of 10 lactating Holstein cows twice a day for 28

days after milking sessions and 2 non-lactating, pregnant Jersey cows once a week upon calving. At the same time, the ground truth body weight records were collected using a walk-over weighing system. A total of 40,405 depth images and depth map files were obtained. Three approaches were investigated to segment the cow's body from the background, including single thresholding, adaptive thresholding, and Mask R-CNN. Four image-derived biometric features, such as dorsal length, abdominal width, height, and volume, were estimated from the segmented images and fitted to predict body weight using ordinary least squares, ridge regression, least absolute shrinkage and selection operator, and linear mixed models. Two cross-validation designs, forecasting and leave-three-cows-out, were used to evaluate prediction performance. The Pearson correlation between image-derived biometric features and scale-based body weight ranged from 0.74 to 0.95. On average, the Mask-RCNN approach to extract body features combined with a linear mixed model to relate features to body weight resulted in the best prediction coefficient of determination and mean absolute percentage error of 0.98 and 2.03%, respectively, in the forecasting cross-validation. The Mask-RCNN approach was also the best in the leave-three-cows-out cross-validation, followed by adaptive and single thresholding. The prediction coefficients of determination and mean absolute percentage error of the Mask-RCNN coupled with the linear mixed model were 0.90 and 4.70%, respectively. Our results suggest that deep learning-based segmentation improves the prediction performance of cow body weight from longitudinal depth video data.

3.2 Background

The body weight of dairy cattle varies on a daily basis due to variations in gut fill, growth, gravid uterus development, and changes in fat accumulation in adipose stores. Regular monitoring of body weight is important because estimates of body weight can provide insight into management factors that may affect milk production, feed efficiency, energy balance, growth, and health status. For example, a negative energy balance can potentially indicate an inadequate diet or health and reproductive problems. For this reason, individual monitoring of body weight throughout a cow's lifespan is essential to make better management decisions. However, body weight monitoring is not performed as often as it should be because it is labor- and time-consuming when measured manually or cost-prohibitive if using an automated walk-over weighing system. In addition, the trend toward larger average farm sizes provides more opportunities for increasing the efficiency of identification of cow health problems. The use of a low-cost, non-contact, stress-free 3 dimensional (3D) depth sensor camera coupled with computer vision analysis has been proposed as an alternative method to address the challenges associated with assessing cow morphometric traits, including predicting body weight [124]. Computer vision is considered one of the sub-fields in a broader effort of precision livestock farming, which aims to integrate advances in engineering and technology, and monitor animals at individual levels to positively intervene in farm management [120]. Although earlier computer vision applications for dairy cows involve the simultaneous use of multiple 2D cameras, a detailed calibration process is required to synchronize them [89, 173]. A 3D depth sensor camera recently used in computer vision provides distances between the camera and objects in addition to standard two-dimensional information, including body contours. Height- and volume-derived 3D images have been shown to improve the performance of body weight predictions as assessed by model goodness of fit [51, 86, 90, 104, 191] and cross-validation [89, 123, 167].

Body weight is a repeated trait. However, previous body weight prediction research has used data collected at a single point in time without repeated measurements from the same cow. This type of experimental design does not allow evaluation of the predictive performance of models for observations that vary over time. Previous studies in dairy cattle that collected repeated body weight measurements used limited sample sizes, limited time points, manually selected the best image from each video, or did not employ cross-validation to evaluate predictive ability [90, 104, 191]. Therefore, an alternative approach is needed that can better explore longitudinal records and improve body weight prediction.

Deep learning has been recognized as mainstream in image and video data analysis. Yet, the usefulness of deep learning for dairy cow body weight prediction has not been assessed. Image segmentation is often a critical first step in computer vision to separate the cow from background information. The quality of cow biometric features extracted from segmented images heavily influences the predictability of traits of interest. Among various deep learning models, Mask R-CNN is used for instance segmentation, object detection, and object classification [57]. Although Mask R-CNN has been witnessing growth in cow detection [9, 143, 157, 193, 195], research evaluating the utility of Mask R-CNN for body weight prediction is in its early stages and has not been evaluated in great detail. Additionally, applying deep learning-based computer vision to animal video data often involves extensive data analysis using custom computer codes. Open science promotes research products and processes that are accessible to all levels of the community [125]. One way to practice open science is to make the source code used in research publications available to the public. However, it has been pointed out that the adoption of open source code in animal science is lagging behind other disciplines [121]. In particular, none of the computer vision-based body weight prediction work in the published literature to date has released its source code. The lack of available open source code makes animal computer vision data analysis processes

less transparent and less reliable. Breaking down these barriers requires making open source code used in analysis accessible to the animal science community without any restrictions.

Therefore, the objectives of this study were 1) to predict cow body weight from repeatedly measured video data, 2) to compare the performance of automated thresholding and Mask R-CNN approaches, 3) to evaluate the predictive ability of body weight regression models, and 4) to release open source code for depth video-based body weight prediction to the animal science community.

3.3 Materials and methods

3.3.1 Animal experiments

Animal handling and media recording were approved and conducted in accordance with the Virginia Tech Institutional Animal Care and Use Committee. A total of 12 multiparous animals were used. Ten lactating Holstein cows (286 days in milk) with average initial body weight (SD) of 793 (148) kg and two pregnant non-lactating Jersey cows with initial body weight of 552 (33) kg at 12- and 21-days ante-parturition from Kentland Farm Dairy Complex (Virginia Tech, Blacksburg, VA) were used in the current study. Lactating Holsteins and pregnant Jerseys were chosen to ensure that our computer vision pipeline is robust across breeds and physiological stages. Cows were housed in a free-stall barn, milked twice daily, fed in *ad libitum* regime with a common diet once daily, and had free access to water. Data were collected in 2020. Lactating animal body weight were recorded daily for 28 days after cows left the milking parlor following the 12:00 am and 12:00 pm milking sessions. Non-lactating animal body weight was recorded once a week. Both animals were handled and restrained individually over the walk-through scale, where the scale-based body weight was concomitantly and automatically recorded.

3.3.2 Video acquisition setup

An Intel RealSense D435 depth sensor camera (Intel, Santa Clara, CA, USA) connected to a laptop via a USB 3.1 cable was used to obtain depth video data from cows in this study. The sensor camera provides 87° horizontal and 58° vertical fields of view and uses two stereo image sensors to determine depth. A single-sensor camera was placed in a temperature-controlled box to maintain camera temperatures within the normal operating range. The

box was mounted 2.95 meters above a one-way exit lane equipped with a walk-over weighing system between the milking parlor and pen housing, positioned straight down to capture top views of the cows. Equipment installation and location did not interfere with daily farm activities. Each cow was manually restrained to collect depth data in a short video format of 10-12 s at 30 frames per second and a resolution of 848×480 . Although the cow's body motion was stable throughout the videos, the head and tail were relatively active. The total size of all the video data was 50.1 GB. During the video recording, the cow's body weight was automatically recorded by a walk-through weighing system (Afimilk, Kibbutz Afikim, Israel). Lactating animals were identified via pedometers located at the animals' ankle. Due to pedometer absence in non-lactating animals body weight, identification, and time recordings using the same scale were executed after milking time and manually recorded.

3.3.3 Image segmentation

After the completion of on-farm video acquisition, off-farm data process was initiated. The rs-convert program was used to convert .bag video files into PNG depth images and comma-separated value (CSV) depth map files [70]. The size of the video data increased to 103.2 GB after converting to PNG and CSV. Different colors in each depth image are proportional to distances from the camera to the object. The corresponding CSV file of the same size contains the distance in meters for each pixel. Three image segmentation techniques, single-thresholding, adaptive-thresholding, and Mask R-CNN, were used to segment the cow from its background. The background was considered extraneous information in the images that does not represent the cows' body (i.e., constraint fences).

Single-thresholding

The OpenCV library (version 4.7.0) was used in Python (version 3.10.9) for both thresholding methods. First, all depth images were cropped while maintaining the same ratio to remove the unnecessary surrounding area (i.e., background) while keeping the body part of the cow intact. The cropped images were converted from RGB (red, green, and blue) to HSV (hue, saturation, and value) color models (Figure 3.1). The hue channel was extracted to segment the cow's body [77]. Hue values below the threshold were converted to black and those above were converted to white. Single-thresholding is often used in the literature for image-based body weight prediction [104, 202]. In single-thresholding, the mean of all hue values was applied to all images as a fixed threshold. After thresholding, the OpenCV findcontour function was used to identify and isolate the largest area connecting all contiguous points of equal value, commonly known as the contour. The contour was then drawn on a black mask to remove depth-blank areas (i.e., missing data chunks) in the segmented cow binary image. The body width ratio was used to remove the neck in each image (Figure 3.2) because previous studies have shown that removing the neck improves the prediction of body mass [51, 104]. Specifically, the number of white pixels in each column was calculated and the maximum number of white pixels across all columns was kept. We then calculated the ratio of the total number of white pixels in each column to the maximum value, starting from the center to the head side of the image. If the ratio was less than 0.3 (according to a preliminary analysis), we added several columns of black pixels to remove neck and also ensured that the image size remained the same. Finally, the contour was identified and the bounding box was drawn on the binary image. The bounding box and depth information associated with the identified contour were retained for body weight regression.

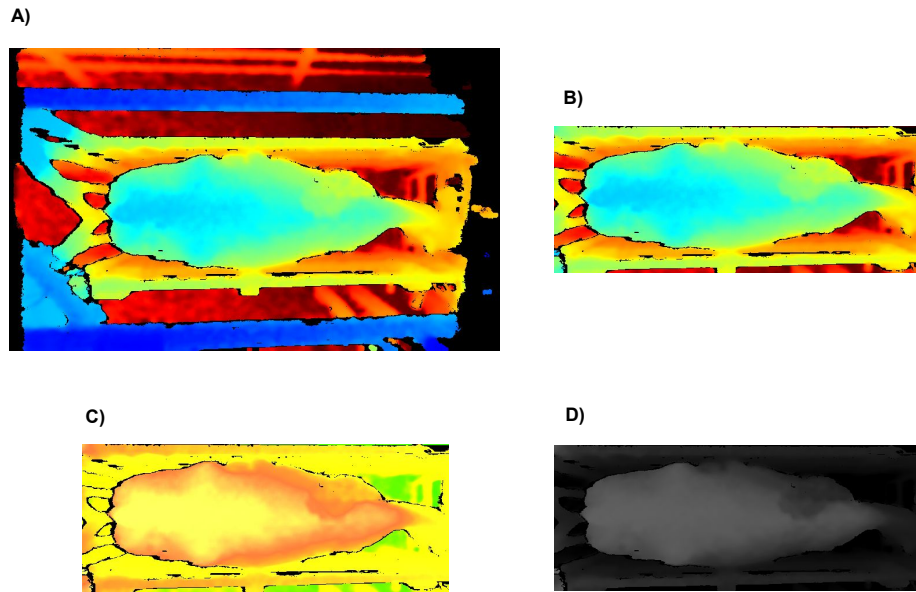


Figure 3.1: Image conversion process. A) The original depth image. B) Cropped depth image. C) The RGB color model was converted to the HSV color model. D) Hue was extracted and used for subsequent analyses.

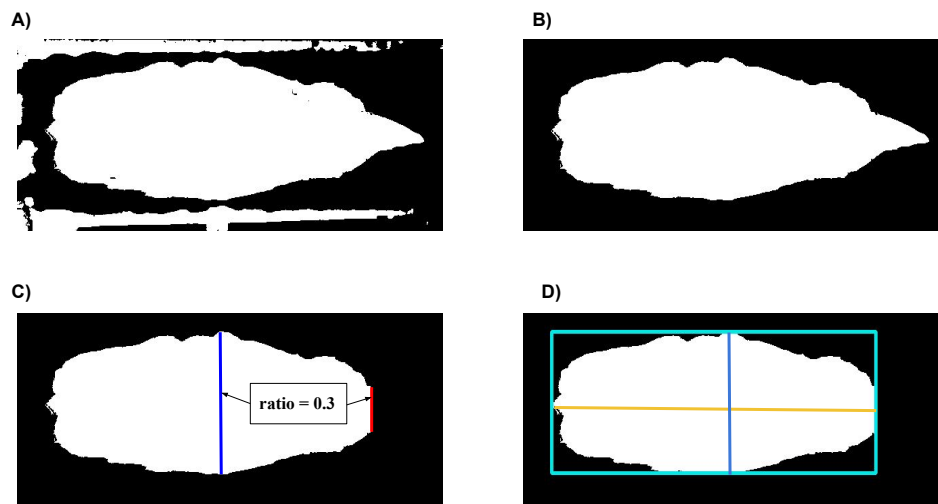


Figure 3.2: Neck removal process. A) Binary image after thresholding. B) The initial segmentation that includes the neck. C) The ratio of body width was used to remove the neck. Blue line represents the column location that returns the maximum number of white pixels across all columns. Red line represents the column location that provides the ratio of 0.3 relative to the blue line. D) The final cow body area with the bounding box. The cyan box, yellow line, and blue line denote the bounding box, dorsal length, and abdomen width, respectively.

Adaptive-thresholding

A potential drawback of the single-thresholding approach is that the best threshold can vary dramatically from image to image due to differences in cow position and lighting conditions. For example, an inadequate threshold will result in incomplete background removal, often resulting in the cow's body being connected to the fence rail, while over-removal of the background will reduce the area of the cow's body part. Lighting conditions at noon and midnight also affect the pixel intensity of an image. To address this challenge, single-thresholding was extended by exploring the optimal hue threshold from minimum to maximum in the adaptive-thresholding approach rather than keeping a fixed threshold. We started with the lowest hue threshold for each image to create a binary image. We drew the bounding box using the OpenCV `boundingRect` function around the largest contour to determine an appropriate threshold. If the coordinates of four corners were less than or equal to five pixels from the edge, the hue threshold was increased by one unit until the four corners were greater than five pixels from the edge of image, because in this case the entire cow body contour is obtained without being connected to the fence rails. The OpenCV `findcontour` function was then used to find the contour with the largest area, and the contour was drawn on a black mask to remove empty areas in the segmented cow binary image. Finally, we used the OpenCV `minAreaRect` function to draw the minimum bounding rectangle of the cow's body contour. The same image processing step was used to remove the neck in each image, identify the contour, and draw the final bounding box (Figure 3.2).

Mask R-CNN

Mask R-CNN is developed on top of Faster R-CNN [146] by adding a new branch that generates a binary mask on each region of interest (RoI). In addition, Mask R-CNN also

incorporates RoIAlign, which enhances feature alignment and aids in accurate mask generation. The overall architecture of Mask R-CNN is shown in Figure 3.3. It includes the backbone network to extract features from the input image, followed by a region proposal network (RPN) that generates object proposals, and a head network that produces the final outputs of the model, including object class labels, bounding box regression, and binary mask predictions for object instances in the input image.

Backbone: All depth images were provided to the backbone of Mask R-CNN after being resized to $1024 \times 1024 \times 3$ with zero padding. The backbone of Mask R-CNN was implemented using ResNet50, a 50-layer residual convolutional neural network model, to perform feature extraction over an entire image [58]. Its architecture includes five stages that extract features from low level (edges and corners) to high level (rich semantics). The first stage includes a convolutional layer and a max-pooling layer. The second through fifth stages contain three, four, six, and three iterations of three-layer convolutional blocks with different designs, respectively, followed by an average pooling layer. In total, there are 48 convolutional layers, one max pooling layer, and one average pooling layer. Stages 2 to 4 in ResNet50 include skip connections, which allow information to be passed directly from one stage to another without going through the intermediate layers, to avoid a vanishing gradient problem, improve prediction, and facilitate convergence. The bottom-up pathway in ResNet50, coupled with the rectified linear unit activation function from the last residual block of each stage, produces a pyramid of feature maps. They were further processed via the top-down pathway to develop the feature pyramid network (FPN) with the aim of improving segmentation performance [97].

Region proposal network: The feature maps generated from the backbone and FPN are fed into a region proposal network (RPN), which is a fully convolutional network, to gen-

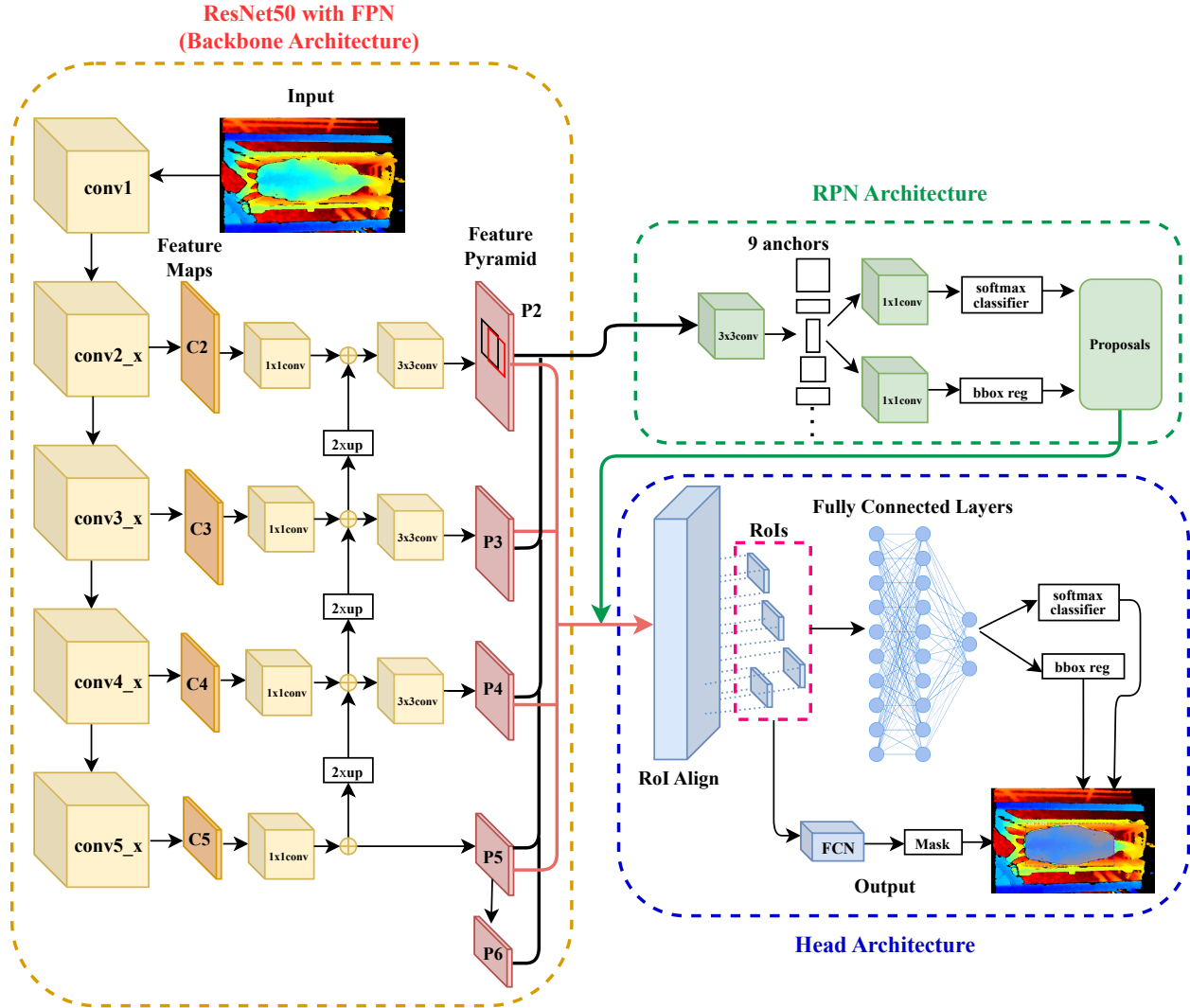


Figure 3.3: Mask R-CNN architecture used in the current study. conv1: initial convolutional block. conv_x: multiple residual convolutional blocks. C2-C5: feature maps generated from each residual block. P1-P5: feature maps generated after the feature pyramid network (FPN). RPN: regional proposal network. RoI: region of interest. FCN: fully connected neural network. Bbox reg: bounding box regression.

erate the region proposals [146]. It uses sliding windows over the feature maps to detect whether the target object (i.e., cow) is present within the region proposals to define ROIs and determine their coordinates. In particular, anchors (rectangles) with nine different sizes, three scales, and three aspect ratios were used. In total, about 21,000 anchors were used per image, since the feature map size after feature extraction was 40×60 . The softmax function was used for each anchor to classify the cow as foreground or background. Here, foreground means that the cow is located in that particular anchor. For anchors with the most possibilities to contain the cow, the bounding box size and location were corrected to get a better fit. Non-maximum suppression was applied to remove bounding boxes that overlapped with other bounding boxes with higher foreground scores. The final set of ROIs are regions in the feature map that contain the cow.

Region of interest align: Region of interest align (RoIAlign) was used to properly project the ROIs onto the extracted features to produce better pixel segmentation results [57]. It adjusts the size of the bounding box to a fixed size of 7×7 and splits the proposal into 7×7 sub-regions within each bin. The bilinear interpolated value of four top points from the image grid was used to keep a single value per bin.

Mask head: After the RoIAlign step, the fixed size ROIs were fed into a fully connected layer for object classification and bounding box regression. The ROIs were also fed to a small fully convolutional network for cattle body segmentation via mask prediction. Specifically, at each pixel location in the ROI, the fully convolutional neural network produced a mask prediction indicating the probability of that pixel belonging to the object class. The mask predictions are then post-processed to obtain the final cow segmentation by thresholding the mask probabilities and assigning each connected component to a unique object instance (i.e.,

cow). This process is equivalent to generating a mask for each RoI, allowing for pixel-level image segmentation.

Model training: A multi-task loss function was used to determine whether the training process of bounding box regression, classification, and mask prediction in a Mask R-CNN model was complete. It was calculated as $L = L_c + L_b + L_m$, where L_c is the classification loss, L_b is the bounding box regression loss, and L_m is the mask loss. The training process of Mask R-CNN was considered complete when the loss converged. The Mask R-CNN model was pre-trained on the COCO dataset, which is widely used for object detection, image segmentation, and image annotation [96]. It contains more than 300,000 annotated images. In addition to the COCO dataset, we randomly selected 300 depth images per cow per time point per day to include more specific situations for cow identification. The use of a pre-trained model is known as transfer learning, where we transferred the parameters from the previous training as initial values. We used Labelme to manually label the ground truth for the cow contour from the background [179]. These labeled cow images were added to the training dataset to fit the Mask R-CNN model. Segmentation performance was evaluated using a confidence score, which ranges from 0 to 1, with higher values indicating better segmentation accuracy [1]. The Mask R-CNN was implemented using Keras [1]. We used a batch size of 1, a step number of 300, an epoch of 42, learning rate 0.001. Since the original version was based on Python 3.4, TensorFlow 1.3, and Keras 2.0.8, we manually upgraded to Python 3.9, TensorFlow 2.9, and Keras 2.9, and updated the Mask R-CNN source code.

Prediction The estimated mask R-CNN parameters were used to segment cow contours in the remaining depth images. To reduce computational time, we chose to predict a subset of depth images for contour segmentation. Specifically, we excluded the first two frames in each video and retained one frame every 15 frames, resulting in predictions of 2 frames per

second. The total number of newly predicted frames was $(40405 - 300)/15 = 2674$.

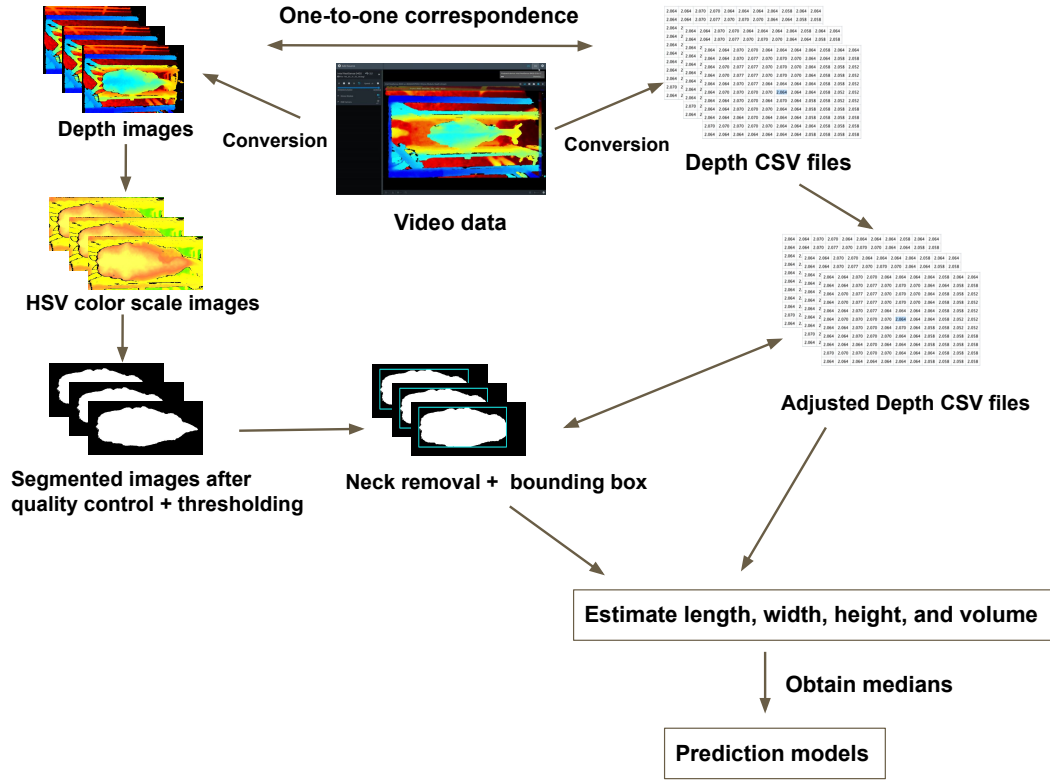


Figure 3.4: Overview of the biometric approach using thresholding.

3.3.4 Biometric features

Instead of selecting the single best image from each video, all images were processed to estimate biometric features. The biometric features of the cows were estimated from the single-thresholding, adaptive-thresholding, and Mask R-CNN segmented contours. Overviews of the biometric approaches using thresholding and Mask R-CNN are shown in Figures 3.4 and 3.5, respectively. The following four biometric features were extracted from each image.

Dorsal length and abdominal width: The dorsal length and abdominal width of the cow in each image were estimated by drawing the bounding box with minimum area using the

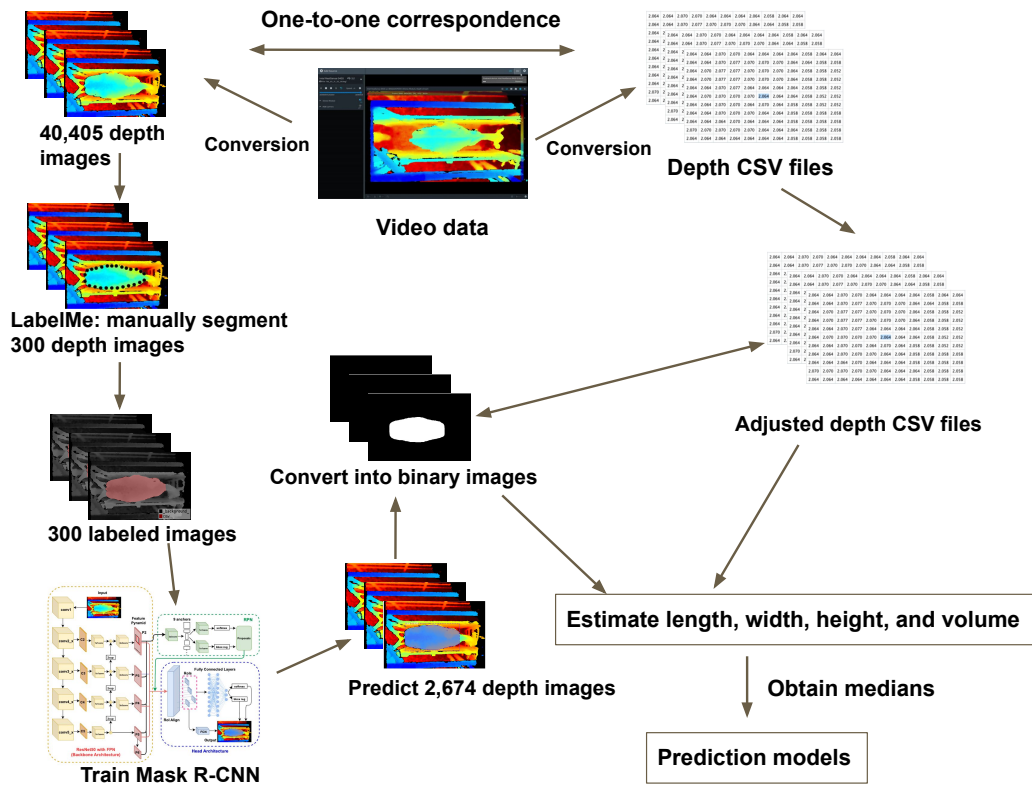


Figure 3.5: Overview of the biometric approach using Mask R-CNN.

OpenCV `minAreaRect` function.

Height: Two methods were evaluated to measure cow height: centroid-based and average-based, to measure cow height. In the first method, the centroid of the cow’s contour was determined using the OpenCV `moments` function. The height of the cow was defined as 2.95 m (height of the camera) minus the distance from the camera to the centroid point using the corresponding CSV depth map file. In the second method, we replaced all zeros (missing data chunks) within contour pixels in the CSV file with the average distance, and the average of all distances between the camera and elements within the entire contour was obtained from the corresponding CSV file. The average cow height was defined as 2.95 m minus the average distance from the camera to the cow contour.

Volume: Because a 3D depth image captures the curvature of the back, the volume of the cow was obtained by summing the height pixels within the cow contour [123]. In mathematics, this is analogous to obtaining the volume by integrating the pixels of a cow area over its depth. An example of the estimated cow volume is shown in Figure 3.6. Once all images were processed and biometric features were estimated, the medians of dorsal length, abdominal width, height, and volume were calculated from all images to derive the single final estimates for each video. The median, rather than the mean, was chosen to make the final biometric estimates robust to outliers.

3.3.5 Body weight regression model

After obtaining all four biometric parameters (length, width, height, and volume), we used them as predictors to build regression models to predict dairy cow body weights. We used ordinary least squares (OLS), ridge regression (RR), and least absolute shrinkage and selection operator (LASSO) to estimate model parameters.

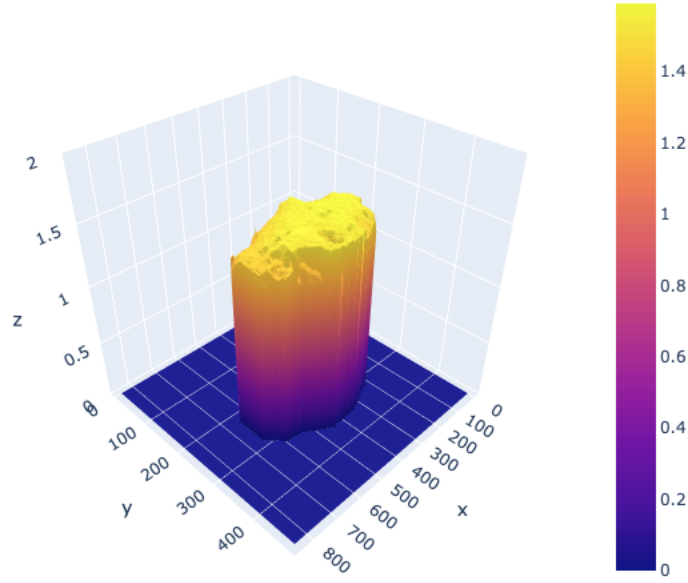


Figure 3.6: Three-dimensional visualization of the cow volume. The volume of the cow was obtained by integrating the pixels of the cow area over its depth.

Ordinary least squares: OLS was constructed using width, length, height, and volume as predictors.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{w} + \mathbf{l} + \mathbf{h} + \mathbf{v} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is a vector of body weight (kilograms), μ is the intercept, $\mathbf{w}$ is a vector of width (pixels), $\mathbf{l}$ is a vector of length (pixels), $\mathbf{h}$ is a vector of height (meters), $\mathbf{v}$ is a vector of volume (cubic meters), ϵ is a vector of residuals. The residual sum of squares (RSS) to be minimized for ordinary least square model is

$$\text{RSS}_{\text{OLS}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\mathbf{X}$ is a matrix including the intercept, length, width, height, and volume; and $\boldsymbol{\beta}$ is a vector of regression coefficients for biometric predictors.

Penalized regression: Two types of penalized regression, RR [60] and LASSO [174], add a penalty function to the OLS objective function. Specifically, an L2 penalty was used for RR to induce shrinkage, while LASSO adds an L1 penalty to have shrinkage and sparsity (variable selection) properties. The RSS for RR and LASSO are

$$\begin{aligned} \text{RSS}_{\text{RR}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \lambda\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} \\ \text{RSS}_{\text{LASSO}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \lambda|\hat{\boldsymbol{\beta}}| \end{aligned}$$

where λ is the regularization parameter, which was fine-tuned using five-fold cross-validation. We implemented OLS, RR, and LASSO using the caret package of R [83].

Linear mixed model: A linear mixed model (LMM) including random intercepts and random slopes was applied to better capture cow variability or cow longitudinal trajectories of body weight records. The OLS was extended by considering individual cow differences as a random effect (random intercepts) and further assuming that the slopes between body weight and time points vary between cows (random slopes). The model used was the following.

$$\begin{aligned} y_{ij} &= \alpha_{0j} + \alpha_{1j}t_{ij} + \epsilon_{ij} \\ \alpha_{0j} &= \gamma_{00} + \sum_{\theta=1}^4 x_{j\theta}\beta_{\theta} + u_{0j} \\ \alpha_{1j} &= \gamma_{10} + u_{1j}, \end{aligned}$$

where y_{ij} is the i th observation from the j th animal, t_{ij} is the animal centered time points, γ_{00} is the grand mean, $\sum_{\theta=1}^4 x_{j\theta}\beta_{\theta}$ is the 4 biometric predictors described earlier, u_{0j} is the random intercept of the j th animal, γ_{10} is the fixed effect of time points, and u_{1j} is the random slope of the j th animal associated with time points. The following variance-covariance matrix of

the random effects were assumed

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{10} \\ \tau_{10} & \tau_1^2 \end{pmatrix} \right],$$

where τ_0^2 is the variance in intercepts between animals, τ_1^2 is the variance in slopes between animals, and τ_{10} is the covariance between intercepts and slopes. LMM was implemented using the lme4 package of R [8].

3.3.6 Model goodness of fit

Goodness of fit of OLS, RR, LASSO, and LLM was evaluated using the full data. This measure determines how well a model fits the current data (inference). In other words, goodness of fit assesses how well the model can predict body weight records that have already been used to estimate model parameters. Coefficient of determination (R^2) and mean absolute percentage error (MAPE) were used as goodness of fit evaluation metrics.

3.3.7 Cross-validation strategies

Cross-validation evaluates how well a model can predict new body weight records that were not used to estimate model parameters. Two cross-validation design strategies were used to test model prediction performance: time series forecasting and leave-three-cows-out (Figure 3.7). In time series forecasting, data from earlier time points were trained to predict the body weight of later time points. All data were divided into training and testing sets according to time points. We varied the ratio of training and testing sets to have five scenarios: 90%:10%, 80%:20%, 70%:30%, 60%:40%, and 50%:50%. Here, 50%:50% indicates that the first half of the time points were used to predict the remaining half. In leave-three-cows-out, nine

and three cows were used as the training and testing sets, respectively. This corresponds to partitioning the total data into 75% and 25%. All possible combinations of training and testing set partitioning were considered ($\binom{12}{3} = 220$). Prediction R^2 and MAPE were used to evaluate the model.

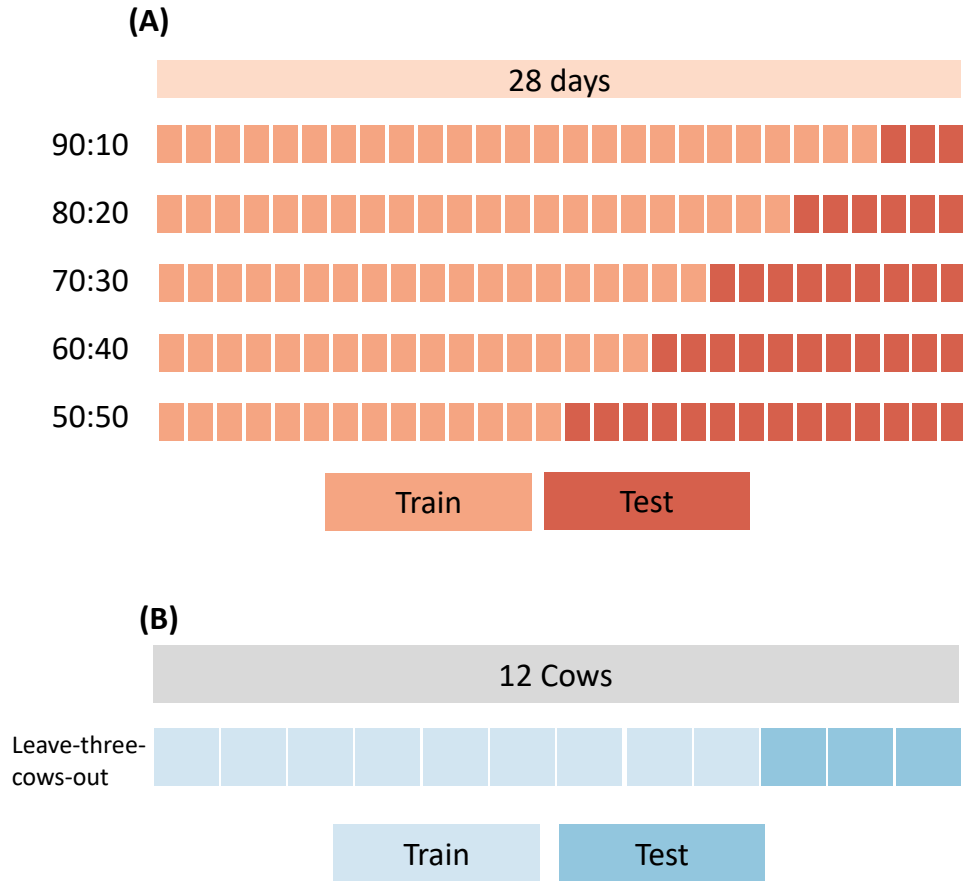


Figure 3.7: Cross-validation design for body weight prediction. Time series forecasting (A). 90:10: the first 90% of time points were used to predict the remaining 10%. 80:20: the first 80% of time points were used to predict the remaining 20%. 70:30: the first 70% of time points were used to predict the remaining 30%. 60:40: the first 60% of time points were used to predict the remaining 40%. 50:50: the first 50% of time points were used to predict the remaining 50%. leave-three-cows-out (B). Of 12 cows, 9 and 3 cows were used as the training and testing sets, respectively.

3.4 Results

Scale-based body weight measurements automatically collected via the walking-through weighting system across time are shown in Figure 3.8. The record taken on Day 5 am was removed because of a systematic error due to the scale. The bottom two cows were non-lactating Jerseys, which had a lower body weight than the Holsteins.

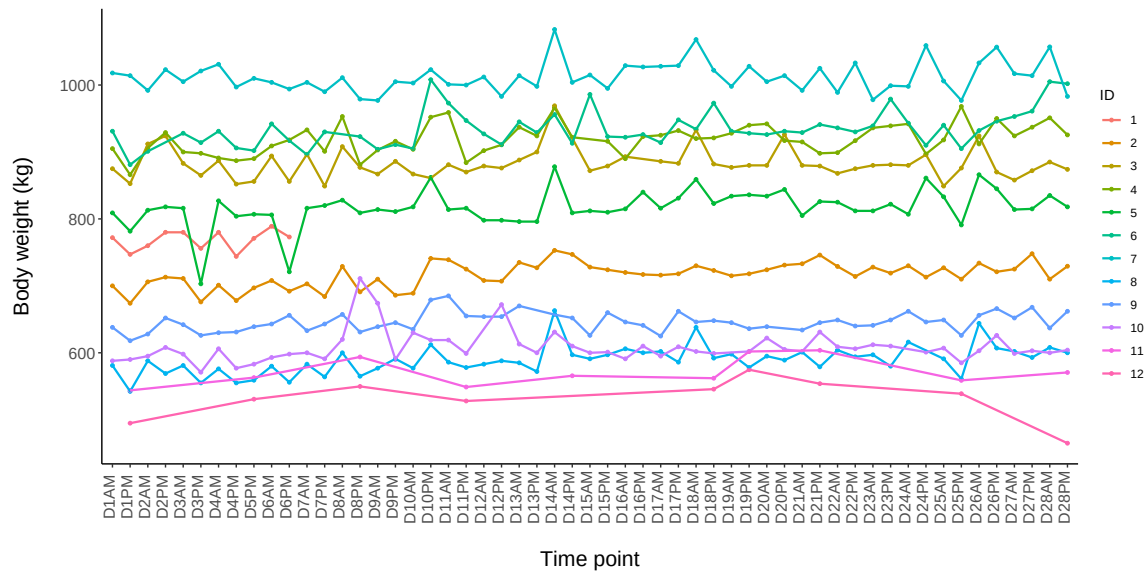


Figure 3.8: Scale-based cow body weight records across time points. ID 1-10: lactating Holstein cows. ID 11-12: non-lactating Jersey cows.

3.4.1 Pearson correlation analysis

The confidence score of the trained Mask-RCNN model using 300 images was 0.99, indicating that the model training process was successful. Pearson correlation coefficients between scale-based body weight and five biometric features obtained from the image analysis averaged across time points ranged from 0.74-0.93, 0.80-0.93, and 0.80-0.95 for single-thresholding, adaptive-thresholding, and Mask R-CNN, respectively (Figure 3.9). Volume was the most

highly correlated with body weight (≥ 0.93) for all single-thresholding and Mask R-CNN and the second most correlated for adaptive thresholding. Width was the most correlated biometric feature for adaptive-thresholding and the second most correlated for Mask R-CNN. Length was the least correlated for single-thresholding. Average heights were consistently more correlated than centroid heights. Overall, Mask R-CNN, which provides pixel-level segmentation, resulted in the highest correlation with biometric features and body weight. The patterns of Pearson correlations between scale-based body weight and image-based biometric features were stable over time (Figure A.1). Volume and width showed high correlations in all the methods. Again length from single-thresholding was not strongly correlated with body weight. We did not observe major differences in correlations when comparing midnight and noon body weights (Figure A.2).

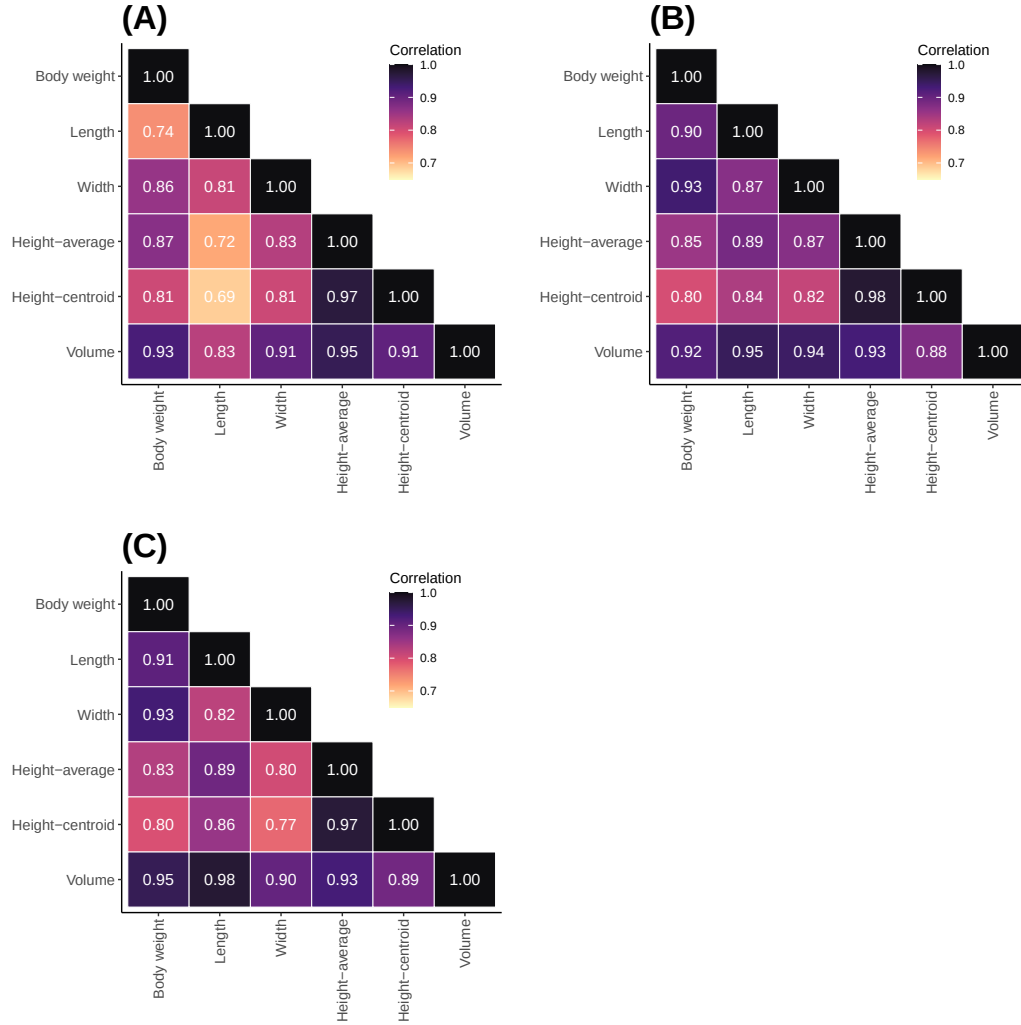


Figure 3.9: Pearson correlation heat map between scaled-based body weight, length, width, centroid height, average height, and volume averaged across time points. A) single-thresholding, B) adaptive-thresholding, and C) Mask R-CNN.

3.4.2 Model goodness of fit

Based on the correlation analysis, average height rather than centroid height was selected to build regression models for goodness of fit analysis and cross-validation. The extent to which the inferred image-based biometric features explained the observed body weight was

evaluated by goodness of fit using the full data. Regardless of the regression models used, the Mask R-CNN segmentation yielded the best goodness of fit according to R^2 and MAPE (Table 3.1). Adaptive-thresholding consistently outperformed single-thresholding when OLS, RR, and LASSO were used as regression models. In contrast, the three segmentation methods performed equally well when LMM was fitted ($R^2 = 0.98$ and $\text{MAPE} = 1.80\%$), which was the best of all. Regarding the regression models, LMM was always the best, followed by OLS, LASSO, and RR.

Table 3.1: Model goodness of fit.

Method	R^2 ¹				MAPE (%) ²			
	OLS ³	RR ⁴	LASSO ⁵	LMM ⁶	OLS	RR	LASSO	LMM
Single-thresholding	0.88	0.85	0.87	0.98	5.93	6.42	6.04	1.80
Adaptive-thresholding	0.91	0.89	0.91	0.98	4.72	5.40	4.84	1.80
Mask R-CNN	0.95	0.94	0.95	0.98	3.57	4.36	3.71	1.80

¹ Coefficients of determination

² Mean absolute percentage error

³ Ordinary least squares

⁴ Ridge regression

⁵ Least absolute shrinkage and selection operator

⁶ Linear mixed model

3.4.3 Cross-validation

Forecasting cross-validation results are shown in Table 3.2. Three regression models, OLS, RR, and LASSO, included only the image-based biometric features as predictors, while LMM included individual cow variability in the model. When the biometric features alone were used to predict body weight, the larger the training data, R^2 and MAPE were slightly larger and smaller, respectively, suggesting that having more past records is useful for predicting future body weight. However, this trend was not observed when single-thresholding biometric features were fitted with RR or LASSO. Mask R-CNN-based segmentation consistently

produced the best prediction results. The next best segmentation method was adaptive-thresholding followed by single-thresholding. When R^2 was used to evaluate the prediction performance, OLS and LASSO performed equally well. The prediction performance of RR was slightly lower than OLS and LASSO. The highest R^2 of 0.96 was achieved by the combination of Mask R-CNN and OLS. According to MAPE, OLS performed best, followed by LASSO and RR. Similar to R^2 , the lowest MAPE of 3.41% was obtained by the combination of Mask R-CNN and OLS. When individual cow variability was added to the prediction model using LMM, the prediction performance increased. Similar to the models including biometric features alone (OLS, RR, and LASSO), lower MAPE was observed when historical records were included in the training data. However, there was no obvious difference between the three segmentation methods used. Overall, LMM achieved the highest prediction R^2 of 0.98 and the lowest MAPE of 2.03.

The results of the leave-three-cows-out cross-validation are shown in Figure ???. Within each regression model, segmentation-derived parameters from Mask R-CNN consistently gave the best prediction according to R^2 and MAPE. Adaptive thresholding was the second best, followed by single-thresholding. Although there was not much difference between OLS, RR, and LASSO when evaluated by R^2 , OLS produced the best predictions based on MAPE. The combination of Mask R-CNN and OLS produced the highest average R^2 of 0.87 and the lowest average MAPE of 4.53.

Table 3.2: Forecasting cross-validation results.

Method	Scenario ¹	R ² ²				MAPE (%) ³			
		OLS ⁴	RR ⁵	LASSO ⁶	LMM ⁷	OLS	RR	LASSO	LMM
Single-thresholding	90:10	0.89	0.86	0.88	0.98	5.94	6.91	6.29	2.10
	80:20	0.89	0.86	0.88	0.98	5.82	6.51	5.99	2.17
	70:30	0.87	0.86	0.87	0.98	6.22	6.57	6.29	2.06
	60:40	0.88	0.86	0.87	0.98	6.07	6.49	6.13	2.14
	50:50	0.88	0.87	0.88	0.98	6.02	6.40	6.03	3.28
Adaptive-thresholding	90:10	0.93	0.92	0.93	0.98	4.74	5.30	4.75	2.12
	80:20	0.91	0.89	0.91	0.98	5.16	5.72	5.22	2.22
	70:30	0.91	0.89	0.91	0.98	5.05	5.69	5.14	2.14
	60:40	0.91	0.89	0.90	0.98	5.00	5.72	5.16	2.39
	50:50	0.91	0.89	0.90	0.98	4.96	5.68	5.10	3.44
Mask R-CNN	90:10	0.96	0.94	0.95	0.98	3.41	4.30	3.53	2.03
	80:20	0.96	0.94	0.95	0.98	3.45	4.40	3.66	2.17
	70:30	0.96	0.93	0.95	0.98	3.51	4.58	3.87	2.18
	60:40	0.96	0.93	0.95	0.98	3.54	4.58	3.84	2.33
	50:50	0.96	0.93	0.95	0.98	3.60	4.59	3.90	3.59

¹ 90:10: 90% training and 10 % testing; 80:20: 80% training and 20 % testing; 70:30: 70% training and 30 % testing; 60:40: 60% training and 40 % testing; and 50:50: 50% training and 50 % testing

² Coefficients of determination

³ Mean absolute percentage error

⁴ Ordinary least squares

⁵ Ridge regression

⁶ Least absolute shrinkage and selection operator

⁷ Linear mixed model

3.5 Discussion

In this study, we evaluated three segmentation methods for their preciseness in segmenting body images of dairy cows from background. This segmentation process is the important first step for the application of computer vision in precision livestock farming, because the quality of the biometric features (length, width, height, and volume) extracted after the image segmentation step greatly influences the subsequent data analysis. We fed the biometric features obtained from the three segmentation methods into four different regression models

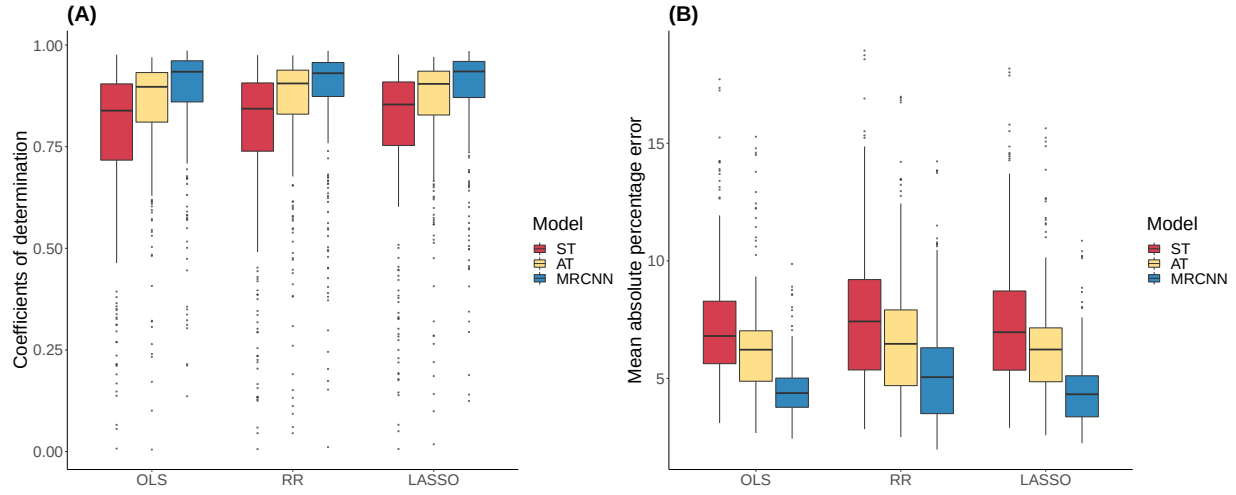


Figure 3.10: Leave-three-cows-out cross-validation results. A) Prediction coefficient of determination. B) Mean absolute percentage error. ST: single-thresholding. AT: adaptive-thresholding. MRCNN: Mask R-CNN. OLS: Ordinary least squares. RR: Ridge regression. LASSO: Least absolute shrinkage and selection operator.

to evaluate the predictability of longitudinal body weight. The main contribution of the current work was to evaluate the utility of Mask R-CNN instance segmentation and contour extraction for cow body weight prediction in great detail and to compare it with two types of thresholding approaches.

Of the four biometric measurements, volume was found to correlate most closely with body weight. Although volume is not always used to predict body weight in dairy cows, the importance of volume is consistent with the recent literature [89, 90, 191]. For instance, Le Cozler et al. [89] developed a 3D reconstruction system using cloud points to perform body weight prediction on 64 lactating Holstein cows. They found that body weight and volume were most strongly correlated. Theoretically, volume simultaneously accounts for all other biometric features (i.e., length, width, and height). Width was the most correlated and second most correlated with body weight in adaptive-thresholding and Mask R-CNN, respectively. The importance of width is consistent with Song et al. [167] reporting among rump length, hip height, and hip width, hip width had the largest correlation with body

weight. Height estimates obtained from averages were more correlated with body weight than height estimates obtained from centroids of body regions. This is because, unlike pigs, the height of a cow is often not flat with a curved surface. For this reason, average-based height estimates were used in the regression analyses in our study.

The quality of biometric features was evaluated from two aspects: goodness of fit and predictability. In the current study, the combination of Mask R-CNN and LMM provided the best model fitness of R^2 of 0.95 and MAPE of 3.57% when cow variability was not considered. When cow variability was included in the prediction model, R^2 and MAPE improved to 0.98 and 1.80%, respectively. These values obtained from without or with the inclusion of cow variability were similar or better than previous studies that reported R^2 of 0.80 using six different types of body length and width measurements from eight cows [86], MAPE of 6.1% using the volume of 185 cows [51], R^2 of 0.96 using the rump width, thorax width, and dorsal area of 28 Holstein cows and 27 Holstein heifers [104], R^2 of 0.92 using the volume of 16 cows [191], R^2 of 0.98 using the volume of 16 cows coupled with LMM including cow variability [191], R^2 of 0.94 using the volume, partial surface area, hip width, chest depth, diagonal length, and heart girth, and body condition score of 16 cows [191], R^2 of 0.98 using the volume, partial surface area, hip width, chest depth, diagonal length, and heart girth and body condition score of 16 cows coupled with cow variability [191], and R^2 of 0.98 using the volume and hip width of 5 heifers [90]. Although the degree of goodness of fit varied among the three segmentation methods, the choice of segmentation method was no longer relevant when LMM was used as the regression model. This suggests that the inclusion of cow variability can compensate for the disadvantages of thresholding compared to Mask R-CNN.

The first type of prediction scenario was evaluated using forecasting cross-validation. Overall, better prediction results were obtained when a larger number of past records were included

in the training sets, suggesting the importance of training data size in forecasting. The best R^2 of 0.96 and MAPE of 3.41% was achieved by the combination of Mask R-CNN and OLS when cow variability was not included. When cow variability was added to a prediction model, R^2 and MAPE improved to 0.98 and 2.03%, respectively. Similar to the goodness of fit results, adding cow variability increased predictions and obscured the differences between the segmentation methods. Because body weight is a repeatedly measured trait, the random effect of the animals contains information about the body mass of each animal. Thus, this information is something we can use when predicting the same cow. We found that the performance of OLS was better than RR or LASSO. This is probably because shrinkage or variable selection was not strongly needed since the total number of predictors used was only four. Our results show that it is possible to predict future body weight of cows from their past biometric features. One potential application is when ground truth body weight records can only be sparsely collected manually on different days for each cow due to the lack of a walk-over weighing system. In such a case, forecasting can be used to predict body weight on specific days from image data alone.

Predictability of new cows not used in the training data was evaluated using leave-three-cows-out cross-validation in the second type of a prediction scenario. Since new cows are predicted in the testing sets, the use of LMM was not feasible in this scenario unless there is genetic information (pedigree or genomics) that allows the cows in the training and testing sets to be connected [4]. The combination of Mask R-CNN and OLS achieved the highest average R^2 of 0.86 and the lowest average MAPE of 4.59%. Overall, our models were able to deliver acceptable predictive performance, but we observed some negative outliers where the predictions were not as successful (Figure ??). These outliers (low R^2 and high MAPE) were further investigated by taking a closer look at the specific partitioning of the training and testing sets that delivered the top 4 predictions and the bottom 4 predictions (outliers)

based on R^2 and MAPE (Figure 3.11). We found that the bottom 4 predictions always occurred when extrapolation was performed. In such a case, all three cows in the testing sets were in either the upper tail or lower tail of the body weight records. For example, the testing set contained two Jersey cows and one small Holstein cow. In contrast, the top 4 predictions were observed when interpolation was performed where all three cows in the testing sets were within the body weight range of the training set. This contrast occurred because of the relatively small sample size used in this study. It is likely that predictions will be more stable as more data are added to leave-several-cow-out cross-validation. To our knowledge, no previous prediction studies employing cross-validation similar to leave-three-cows-out cross-validation with repeated body weight records were found. Although this prevents a direct comparison of the predictive performance obtained in this study and previous work, our results were within the range or slightly better than previous studies that used larger numbers of cows. For example, R^2 of 0.95 and MAPE of 5.60% were reported with a hold out method using length, width, area, volume, withers height, hip height, and the additional non-biometric variable age of 107 cows [123]. Additionally, a MAPE of 7.4% was obtained in leave-cow-out cross-validation including the hip width of 30 cows [167] and R^2 of 0.93 and a MAPE of 2.72% was obtained in repeated random sub-sampling validation replicated 100 times including volume, area, backside width, hip width, heart girth, and wither height of 64 cows [89].

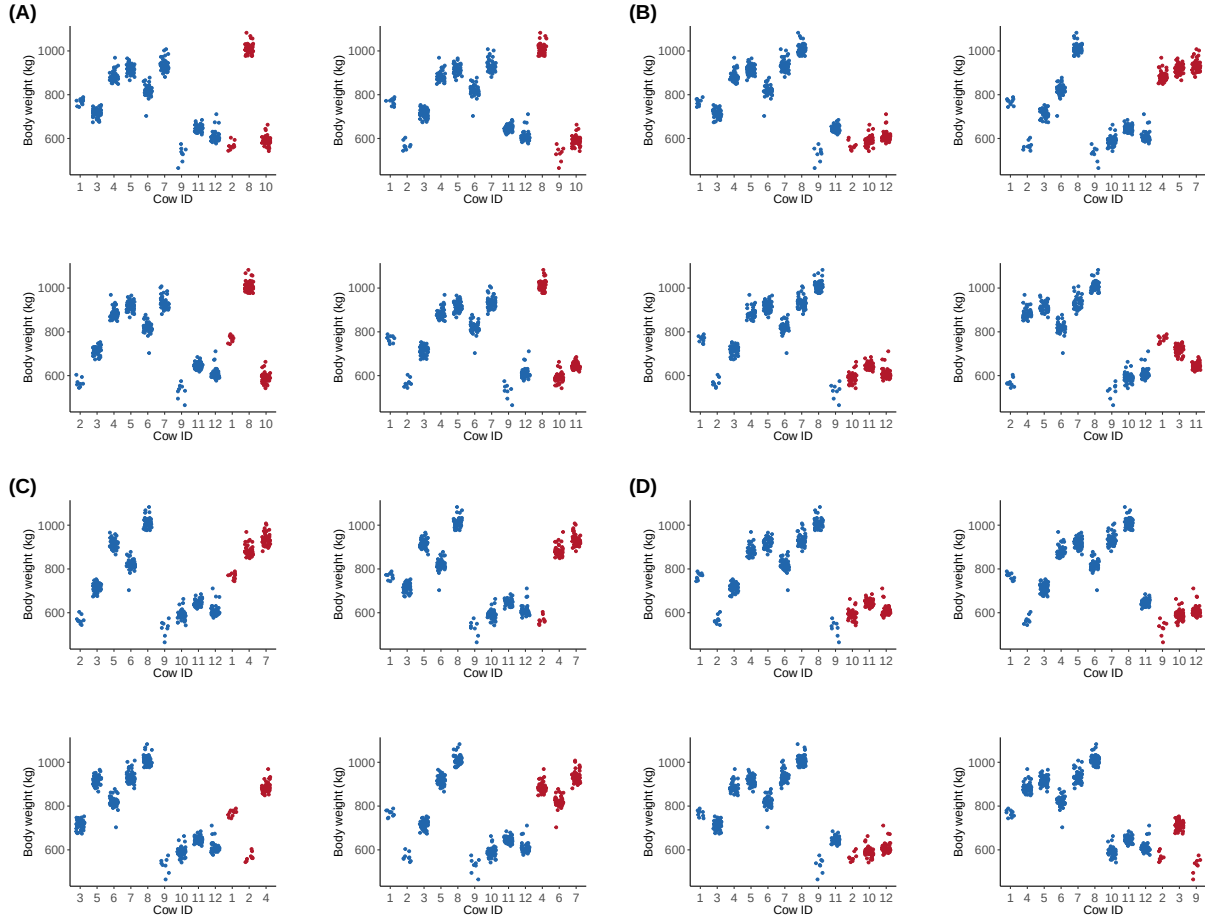


Figure 3.11: Range of body weight records observed in training and testing set partitioning for leave-three-cows-out cross-validation using Mask R-CNN and ordinary least squares. A) top 4 predictions based on the coefficient of determination. B) bottom 4 predictions based on coefficient of determination. C) top 4 predictions based on mean absolute percentage error. (D) bottom 4 predictions based on mean absolute percentage error.

3.6 Conclusions

The current study evaluated model goodness of fit and cross-validation predictability of longitudinal body weight records using automatically extracted image-based biometric features. We found that volume followed by width, correlated best with body weight. The combina-

tion of Mask R-CNN and OLS or Mask R-CNN and LMM provided the best predictive performance when the model did not include cow variability or when the model included cow variability, respectively. We believe that the availability of the source code used in the current analyses will promote open science in the animal science community.

3.7 Code availability

The Python and R code used to perform the video processing and regression analyses are available on GitHub (https://github.com/yebigithub/BW_dairy).

Chapter 4

Industry-scale prediction of video-derived pig body weight using efficient convolutional neural networks and vision transformers

4.1 Abstract

Accurate pig body weight measurement is critical for pig growth, health, and marketing, but conventional manual weighing methods are time-consuming and may cause potential stress to the animals. Although there is a growing trend towards the use of three-dimensional cameras coupled with computer vision approaches for pig body weight prediction, such as supervised neural networks, their validation with large-scale data collected at multiple time points in commercial environments is still limited. Therefore, the main objective of this study was to predict pig body weight from repeatedly measured video data obtained in a commercial environment using efficient convolutional neural networks and efficient vision transformers. An Intel RealSense D435 camera was installed on a commercial farm to collect top-view videos of over 600 pigs every two weeks at six different time points over three months. Manually measured body weight records were also collected using a digital weighing system. We used

an automated video conversion pipeline and fine-tuned YOLOv8 to preprocess the raw depth videos. We then acquired a total of 204,477 depth map files. We applied transfer learning by initializing the weights of two families of lightweight deep neural networks, including MobileNet and MobileViT, with pre-trained weights from ImageNet. ResNet50 was used as the baseline model. We then fine-tuned these models on the pig depth images. The last layer of each model was adjusted to add a dropout layer and a linear regression layer with L2 regularization, allowing direct prediction of body weight for each image without the need for additional image feature extraction steps. We used single time point random repeated subsampling cross-validation (CV1) and time series forecasting cross-validation (CV2) considering limited budget scenarios to evaluate the prediction performance. In CV1, the best prediction mean absolute percentage error (MAPE) for each time point was 4.71%, 3.80%, 3.08%, 5.60%, 3.42%, and 3.77%, respectively. In CV2, the best MAPE was 4.27%, 3.40%, 5.12%, 3.10%, and 3.24%, respectively. On average, the MobileViT-S model produced the best prediction MAPE of 3.08% and 3.10% in CV1 and CV2, respectively. ViT models always showed better prediction performance than CNN models. Our results suggest that efficient deep learning-based supervised learning models are a promising approach for predicting pig body weight from industry-scale depth video data.

4.2 Background

Variability in body weight (BW) in swine production poses significant challenges to farm efficiency, particularly in growing-finishing facilities. Pigs with slower growth rates that reach market weight later than faster counterparts reduce producer income and efficiency, with economic consequences largely tied to quality grading and carcass pricing. This is particularly true for the lightest pigs in a batch, which face significant devaluation [31, 88]. Accurate and frequent live weight measurements play an important role in pig farm management when making decisions about marketing pigs. In addition, regular and continuous weight monitoring allows early detection of digestive and health problems in pigs [28, 100]. In commercial production environments, continuous measurement of accurate records of individual pig weights often presents challenges, including labor-intensive processes, time constraints, animal stress and anxiety, and the potential risk of injury to pigs [43, 153].

Due to the low cost and high reliability of depth measurements, RGB-D cameras are increasingly used for various applications in precision livestock farming (PLF), such as BW prediction [21, 22, 32]. This non-invasive approach allows effective assessment of live BW with minimal negative impact on the animals [124]. In addition, to build an efficient computer vision system for BW prediction, it is crucial to identify the optimal camera positioning, as capturing images of pigs from different angles can introduce variations due to differences in their shapes within the images, such as object distortion. Previous studies have successfully used 3D depth cameras positioned from the top view to obtain depth images of pigs to perform BW prediction [32, 45, 53, 59, 91, 132, 139, 202].

After capturing the depth images, the next step is to process these images using advanced image processing algorithms to accurately predict BW. Previous studies have shown positive correlations between BW records and manually measured body dimensions, such as length,

heart girth, and width [44, 131]. Semantic segmentation builds on this concept to estimate these morphological measures from depth images by first segmenting the animal body from the background and then extracting biometric features, such as length, width, height, volume, chest depth, and curvature. These metrics are then used in prediction models, which are mostly linear or non-linear regressions or machine learning models [6, 79, 138, 202]. However, this type of two-step approach has several drawbacks. Some major drawbacks are that its performance is strongly influenced by the choice of a segmentation model, obstacles covering parts of the animal’s body, and variations in illumination that affect the performance of the depth camera [10, 39, 48, 108]. Another drawback is inefficiency, as this approach involves both feature extraction and regression modeling, making it less efficient for real-time applications.

To date, the growing influence of deep learning techniques in PLF has been noteworthy [17, 184]. In particular, one area of research is to investigate the use of raw image data as direct input to predictive models for BW prediction using convolutional neural networks (CNN) [15, 76, 108]. These studies have shown that top-view depth image data without feature extraction is sufficient for accurate BW prediction. In particular, CNN offers an advantage over semantic segmentation approaches in that it automatically extracts the most relevant features, eliminating the need for predetermined feature preprocessing and mitigating the challenges posed by imperfect data acquisition environments. In addition to CNN, Vision Transformer (ViT) is another well-received neural network for image recognition tasks in computer vision [30]. By dividing the image into patches with positional embeddings as input and incorporating multiple multi-head self-attention blocks, this approach has achieved performance at least comparable to CNN-based approaches. For example, networks using ViT as a backbone have demonstrated state-of-the-art performance in tasks such as semantic segmentation, object detection, and image classification, underscoring the superiority of ViT

in computer vision [46, 98].

However, modern state-of-the-art networks typically involve more complex and deep designs that require substantial computational resources, making them impractical for many on-farm applications [20, 58, 66, 165, 170, 171]. Therefore, the focus must be on deploying more efficient architectures that can deliver accurate BW predictions with fewer modalities, fewer resources, and faster processing time.

The use of the MobileNet and MobileViT series provides a new opportunity to design efficient neural networks tailored to pig BW prediction. MobileNet uses depthwise separable convolutions, which significantly reduces model size and computational requirements while maintaining reasonable accuracy [63]. MobileNetV2 further optimizes this approach with inverted residuals and linear bottlenecks, improving both performance and efficiency [160]. MobileNetV3 combines automated search techniques with a novel architecture design, resulting in superior performance with minimal resource requirements [62]. It offers two variants, MobileNetV3-Large and MobileNetV3-Small, designed for high and low resource environments, respectively. Compared to MobileNetV2, MobileNetV3-Large was 3.2% more accurate in ImageNet classification while reducing latency by 15%, and MobileNetV3-Small was 4.6% more accurate while reducing latency by 5%. For PLF applications, [2] applied MobileNet to RGB, depth, fully segmented, and fully segmented depth images for cattle BW prediction and [203] applied MobileNetV2 to pig depth images for BW and size prediction.

MobileViT integrates lightweight self-attention mechanisms that view transformers as convolutions, allowing it to take advantage of both convolutions (e.g., inductive biases) and transformers (e.g., long-range dependencies) to build a lightweight network for mobile devices [110]. Similarly, MobileViTV2 further refines attention mechanisms into separable self-attention with linear complexity to optimize mobile vision applications. Taken together, these advances highlight the potential of using efficient neural networks for on-farm appli-

cations. The application of MobileViT for facial and coat recognition in dairy cattle and sheep is presented in the literature [92, 181]. However, the application of MobileViT for BW prediction remains scarce. In addition, few studies have compared the feature extraction performance of efficient CNN and ViT. Therefore, the main objective of this study was to predict pig BW from repeatedly measured large video data obtained in a commercial environment using efficient CNN and ViT.

4.3 Materials and Methods

4.3.1 Animals and data collection design

The data used in this study were collected from pigs raised in a nucleus farm operated by Smithfield Premium Genetics (Rose Hill, NC, USA). Therefore, animal use approval was not required. A total of 623 animals were used in the current study. Data were collected on a commercial farm over three months during six farm visits (July to September 2023), referred to as T1 to T6. The number of pigs at these visits was 623, 612, 600, 591, 353, and 196, respectively. The decrease in the number of pigs was mainly due to sales for marketing. A small number of pigs were removed from the later collections due to being moved to the sick pens. The pigs were fed daily in a *ad libitum* regime with a common diet, and had free access to water. Scale-based BW measurements were recorded using a single-animal walk-in digital scale, with pigs individually handled and restrained over the stand-on scale. Histogram plots of scale-based BW records for all pigs across six farm visits are shown in Figure 4.1. The weights ranged from 25 to 154 kg. These were used as ground truth for the BW prediction models.

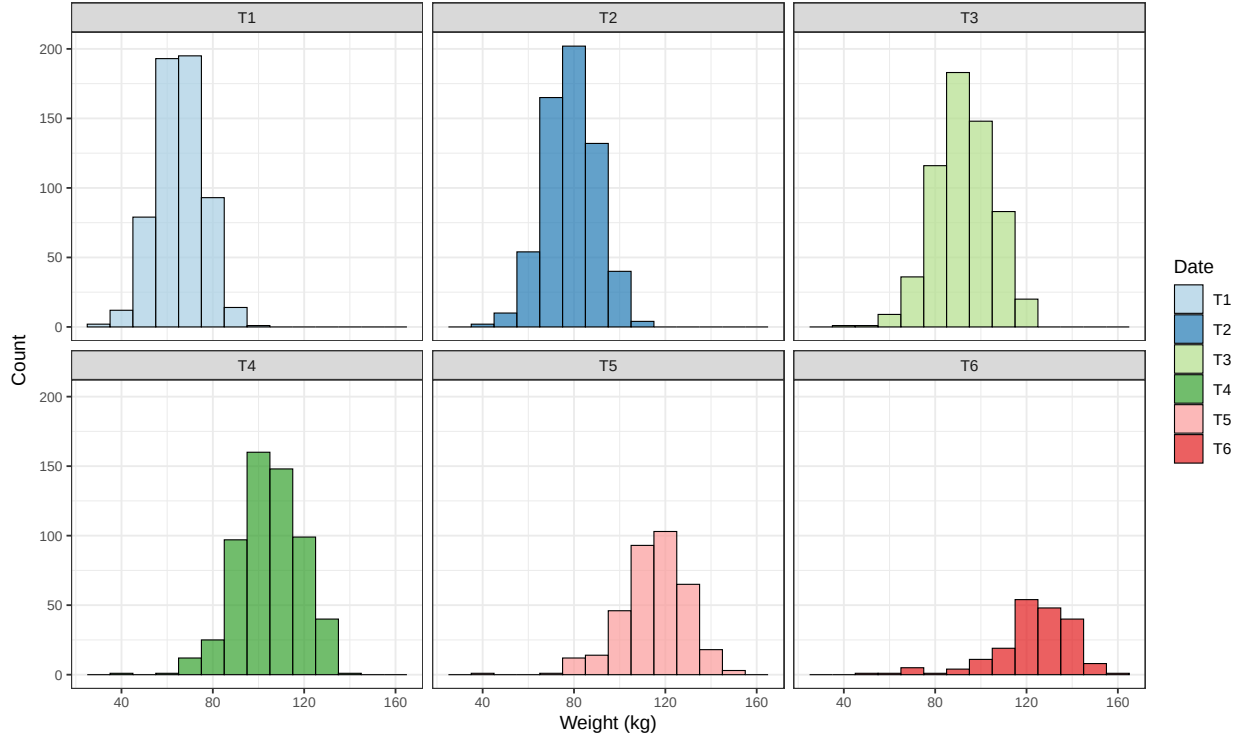


Figure 4.1: Histogram of scale-based pig body weight records. T1 to T6 represent six time points analyzed in this study.

4.3.2 Video acquisition setup

A video acquisition setup closely followed that of Bi et al. [10] for dairy cattle. An Intel RealSense D435 (Intel, Santa Clara, CA, USA) connected to a laptop via a USB 3.1 cable was used to capture depth video data from pigs in this study. With a horizontal field of view of 87° and a vertical field of view of 58° , the sensor camera uses two stereo image sensors to determine depth. Positioned approximately 1.40 meters above a one-way exit lane equipped with a stand-on weighing system, a single-sensor camera was mounted in a camera holder and angled downward to capture top views of the pigs. Each pig was manually restrained to collect depth data, resulting in short video recordings of 10-12s at a frame rate of 30 frames per second and a resolution of 1280×720 . While the pigs' body movements remained stable

throughout the videos, their heads and tails showed relatively active movement. The total size of all video data was 354 GB. Simultaneously with the video recording, the pigs' BW was measured using a weight scale, and their identification was determined by ear tag detection. To ensure the robustness of our analysis, we removed outliers from the ground truth BW records by filtering out the lower and upper third percentiles of the weight distribution at each time point.

4.3.3 Video preprocessing

The depth data collected in .bag video format were converted to PNG depth images and comma-separated value (CSV) depth map files using the *rs-convert* program[70], increasing the video data to 620 GB. Different colors within each depth image correspond to different distances from the camera to the object, providing a visual representation of spatial depth. Each image was accompanied by a CSV file of the same size, providing distance measurements in meters for each pixel. The total number of depth images was 52,513, 91,446, 95,776, 81,922, 55,305, and 71,720 for T1 to T6, respectively.

Since a large amount of videos were collected in a commercial environmental setting, the videos were preprocessed to maintain overall quality for subsequent image analysis. Two main challenges were 1) the videos contained not only pig bodies but also other unrelated objects, such as human bodies and 2) the positions of the pigs varied considerably from image to image. In addition, some videos contained two pigs entering at the same time. Thus, we fine-tuned YOLOv8s [74] to perform preprocessing to address these issues. Specifically, we annotated 100 depth images by labeling "pig" and "block" categories in the Roboflow software [95]. Here "pig" means the main body of the pig and "block" means the worker blocking the camera view. Since we found in the preliminary analysis that YOLOv8s was

good enough to detect these two class objects compared to other YOLOv8 versions, we fine-tuned the YOLOv8s model on these labeled images. After training, we fit the model to detect objects in all depth images. Three criteria were used simultaneously to remove images from the data: block detection rate greater than 0.1, pig detection rate less than 0.5, and no detection for either pig or block. Some examples are shown in Figure 4.2. After preprocessing, the size of the depth images was reduced to 20,356, 45,814, 44,525, 54,580, 23,968, and 15,234 for T1 to T6, respectively. Due to the varying length and quality of each video, the number of depth images for each animal varied, as summarized in Figure 4.3. To avoid data imbalance caused by pigs with too many images, we uniformly selected 10 or 20 images for each animal to train the deep learning models, depending on the cross-validation (CV) designs employed (see below). For pigs with less than or equal to 10 or 20 images, all images were included. For pigs with more than 10 or 20 images, we used a sampling strategy by uniformly selecting images at regular intervals so that the number of images per pig was exactly 10 or 20.

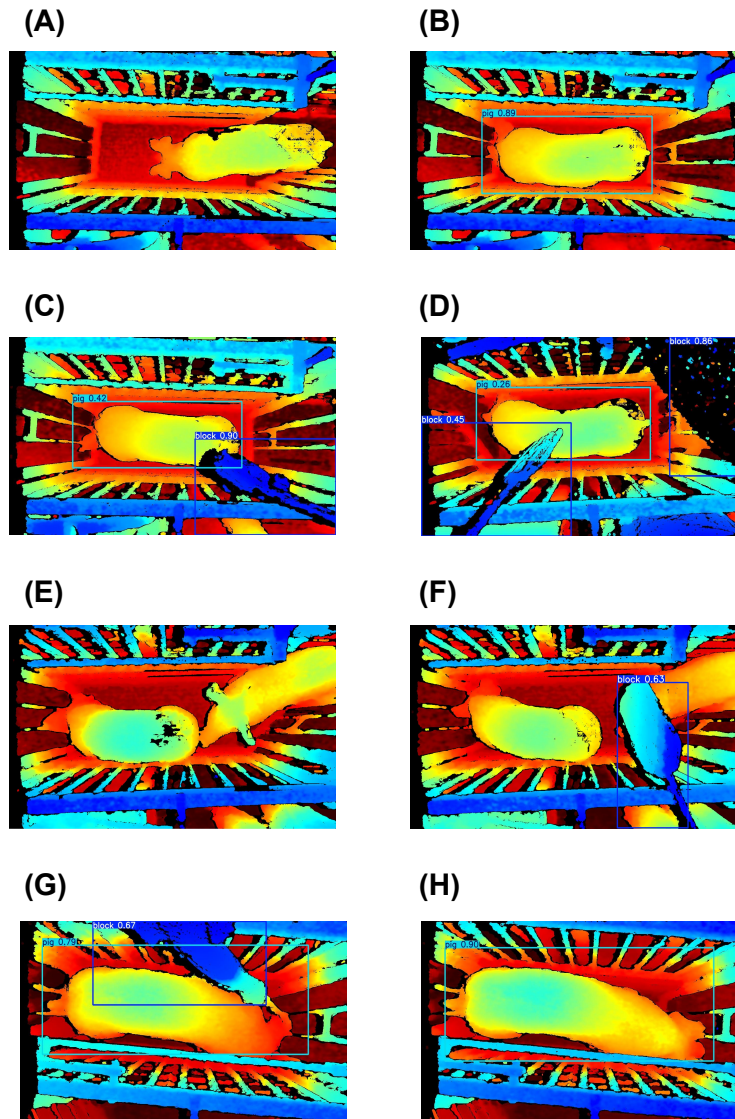


Figure 4.2: YOLOv8 detection examples. (A) No detection: the pig is not in the center of the image. (B) Pig detection rate of 0.89: acceptable image because the pig is in the center of the image. (C) Pig and block detection: both the pig and the block are in the image. (D) Pig and block detection: both the pig and the block are in the image, but the pig is not in the center of the image. (E) No detection: multiple pigs in the image. (F) Block detection rate of 0.63: Only block is detected. (G) Pig and block detection: the camera is positioned in the opposite direction. (H) Pig detection rate of 0.90: acceptable image because the pig is in the center of the image but the camera is positioned in the opposite direction.

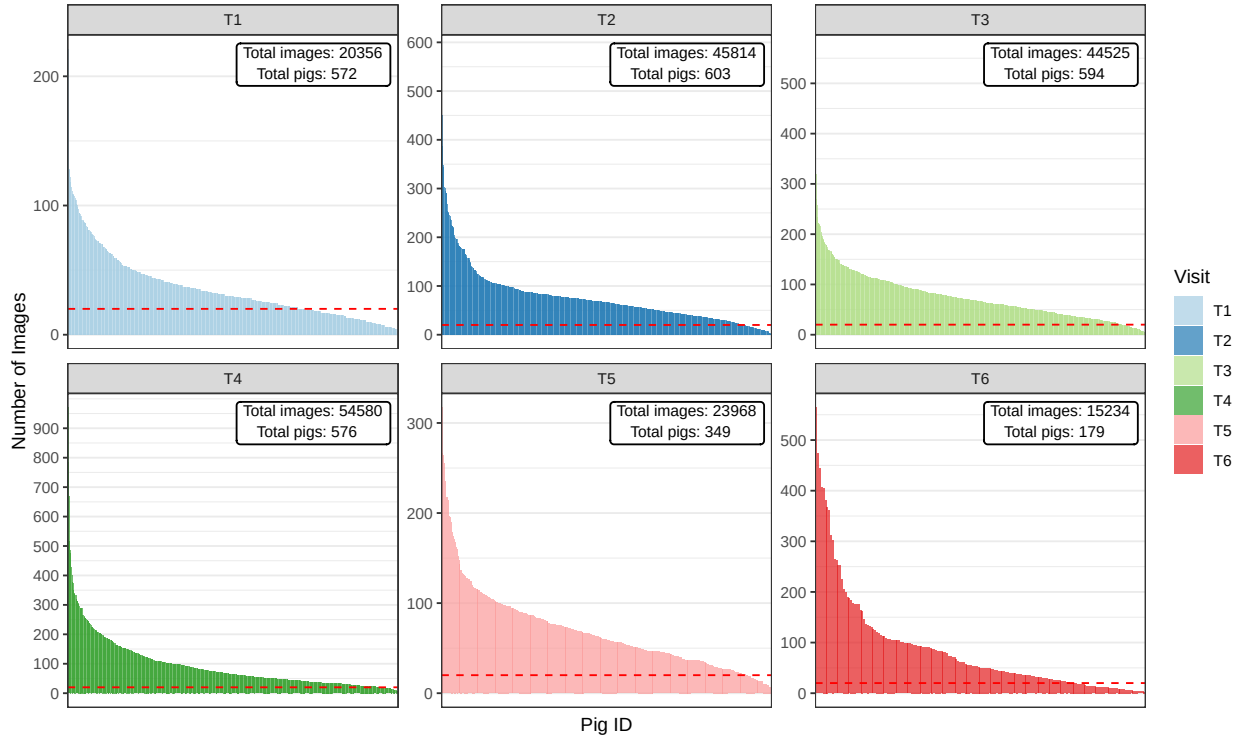


Figure 4.3: Number of images for each pig across six farm visits. T1 to T6 represent six time points analyzed in this study. The dotted red line represents the number of images of 20.

4.3.4 Deep learning models

We used deep learning models to predict BW from depth images. The main idea was to feed the entire depth images to neural network models to directly predict pig BW. This approach bypasses the need to manually adjust internal parameters, such as cropping and thresholding through image segmentation, as in a biometric counterpart. Additionally, deep learning models integrate feature extraction and regression into a single-step process. The overall workflow of the proposed method is shown in Figure 4.4. Depth CSV files were first converted into depth grayscale images, then unsupervised features were extracted from these images and used as input to neural networks for BW prediction.

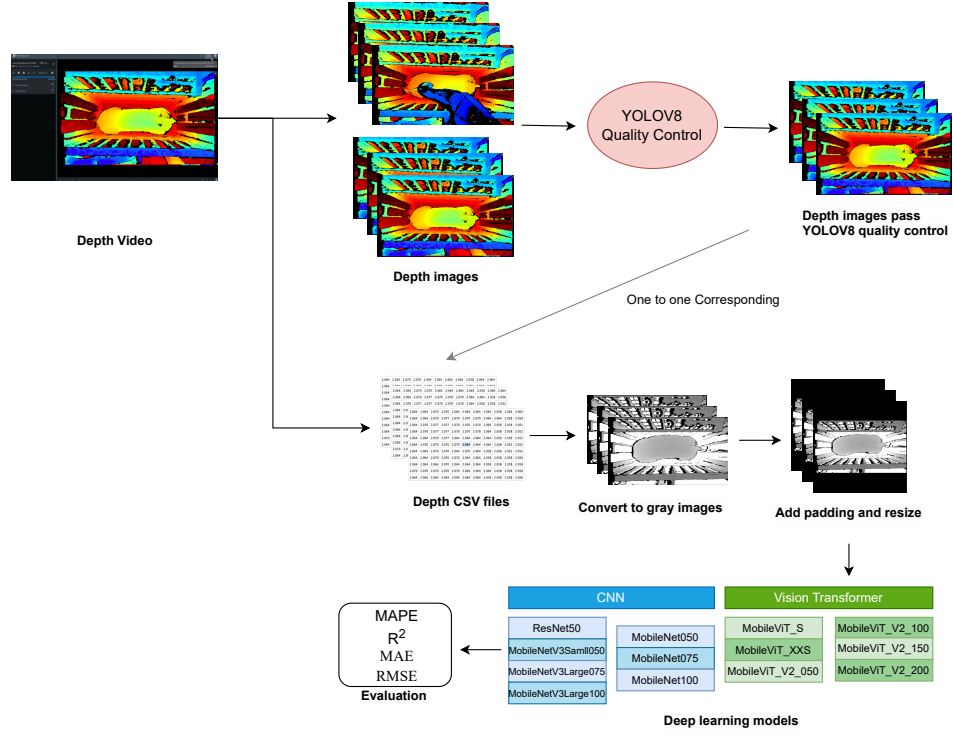


Figure 4.4: Overview of the deep learning workflow for body weight prediction. Depth videos were processed into images and comma-separated values (CSV) files. YOLOv8 filtered out poorly positioned or obstructed images. The quality-controlled CSV files were converted to grayscale images, padded, and resized. Neural networks are then trained and evaluated using four metrics. MAPE: mean absolute percentage error; R^2 : coefficient of determination; MAE: mean absolute error; and RMSE: root mean squared error.

Depth file preprocessing

Since histogram equalization processing was applied to the depth images produced by the *rs-convert* program, the color scheme in the images became inconsistent depending on whether the object was present or not. However, we preferred that depth images convey only stable distance information. For this reason, we converted depth CSV files containing distance information in meters to depth grayscale images. This conversion was performed using the *pyplot* function in the matplotlib library (version 3.7.1) in Python (version 3.8.16). The

resulting images retained a size of 1270×720 pixels. We then resized the images to a smaller size and added padding at the top and bottom to convert them to a square dimension.

Supervised neural networks

Given the remarkable performance of many CNNs in the ImageNet Large Scale Visual Recognition Challenge [24], we evaluated the feature abstraction capabilities of three deep learning model families initialized with transfer learning. These models included the MobileNet and MobileViT series [110, 111], and ResNet50 as the baseline model, all pre-trained with ImageNet weights. Since these models were originally designed for image classification tasks, the following specific modifications were made: 1) the classifier layer was removed from each model; 2) their base model was retained as the backbone; 3) a dropout layer with a 50% dropout rate was added; and 4) a single node dense layer with a linear activation function and L2 regularization was added to each model to generate BW outputs. The main architecture of each model is summarized below.

MobileNet uses depthwise and pointwise separable convolutions as an efficient replacement for traditional full convolutional layers. As a lightweight model, MobileNet prioritizes a balance between model compression and accuracy [63]. For two hyperparameters, width multiplier and resolution multiplier, we set them to 1.0, meaning that we applied the same number of filters used in each layer [63] and 3×3 depthwise separable convolutions. MobileNetV2 uses the linear bottleneck and inverted residual structure of MobileNet, and it consists of three key components: a 1×1 expansion convolution, depthwise convolutions, and a 1×1 projection layer [160].

MobileNetV3 builds on MobileNetV2 by adding modified swish nonlinearities as the activation function and adding squeeze and excitation in the residual layers [62]. We evaluated two architectures, MobileNetV3-Small and MobileNetV3-Large, where the number of filters in each layer was proportionally reduced by 0.5 and 0.75. The specific models tested were MobileNetV3Small050, MobileNetV3Small075, MobileNetV3Small100, MobileNetV3Large075, and MobileNetV3Large100. Here, "100" denotes the original model, while "050" or "075" denotes a 0.5 or 0.75 reduction in the number of filters.

MobileViT is a hybrid approach combining convolutions and transformers, therefore it combines the strengths of CNNs and ViTs, which are spatial inductive biases and less sensitivity to data augmentation and input-adaptive weighting and global processing, respectively [110]. In brief, this approach combines the MobileViT block with the MobileNetV2 block. The MobileViT block starts by applying a standard convolutional layer followed by a pointwise convolutional layer to the input tensor, capturing local spatial information and projecting the tensor into a higher dimensional space. The resulting tensor is then unfolded into non-overlapping flattened patches, to which transformers are applied to model long-range dependencies. The processed patches are re-folded into their original spatial dimensions, and this tensor is projected back into a lower dimensional space using another pointwise convolution. The resulting low dimensional tensor is concatenated with the original input tensor, and an additional convolutional layer fuses these features. We evaluated all three architectures of Mehta and Rastegari [110], including MobiViT-XXS, MobiViT-XS, and MobiViT-S.

MobileViTV2 The architecture of MobileViTV2, which is based on MobileViTV1 includes two major changes. It replaces multi-headed self-attention with separable self-attention and removes the fusion block and skip connection because they only marginally improve perfor-

mance. The expansion factor in MobileNetV2 blocks and feed-forward layers is two, and Swish is used as the non-linear activation function [111]. Unlike MobileViTV1, which creates three specific architectures for mobile devices, MobileViTV2 scales the network width uniformly using a width multiplier α ranging from 0.5 to 2.0 to create models of different complexity. In this study, we tested MobileViTV2-050, MobileViTV2-100, and MobileViTV2-150. Here, "100" denotes the original model, while "050" or "150" denotes α equal to 0.5 or 1.5.

ResNet50 was used as the baseline model. It uses residual blocks and applies an identity shortcut connection to skip one or more layers to mitigate the degradation problem, allowing for easier optimization and increased accuracy with depth [58]. We also compared ResNet50 with deeper ResNet models, such as ResNet101, ResNet152, and found that ResNet50 was more suitable for the current data.

Activation heatmaps visualization

We employed gradient-weighted class activation mapping (Grad-CAM) [162]) to visualize the activation heatmaps of our deep learning models. Grad-CAM is a technique that provides visual explanations for decisions made by neural networks. By computing the gradients of the target concept with respect to the final convolutional layer, Grad-CAM produces a coarse localization map that highlights the important regions in the input pig image that influence the model's predictions. The generated heatmaps allowed us to compare model performance by visually inspecting which parts of the input pig images were considered important by the models. This comparison facilitated an understanding of the models' decision-making processes and helped identify potential biases or areas for improvement. The visualizations were particularly useful in validating the reliability and interpretability of our models.

Table 4.1: Hyperparameters of deep learning models.

Parameter	Value
Image Size	100, 150, 200
Epoch	200, 500, 1000
Batch size	100, 200, 1000
Loss function	mean squared error
Learning rate	0.001 to 0.000001
Optimizer	Adam, AdamW

Implementation

All models were implemented using Tensorflow-gpu version 2.14.0 and Keras version 2.14.0 in Python version 3.9.0. All code was run on the Virginia Tech Advanced Research Computing cluster with two 80 Gb NVIDIA DGX A100 graphics cards. To compare the performance of each model under the same conditions, we first tested the configuration of the possible hyperparameters used in each model, as shown in Table 4.1. The image size of 150×150 , a batch size of 100, and 300 epochs were found to be appropriate for model convergence. The mean squared error was used as the loss function. In the training process, we used the *ReduceLROnPlateau* callback to dynamically adjust the learning rate, reducing it by a factor of 0.1 if the validation loss did not improve for a quarter of the total epochs, with a minimum learning rate threshold set at 0.000001 and the *ModelCheckpoint* callback to save the best model weights based on the validation loss with a patience value of 75 epochs. The loss of each model on the validation set was computed as the evaluation standard to select and retain the best parameters in the training process for each model.

4.3.5 Cross-validation strategy

Two CV designs were used to evaluate model prediction performance: single time point (single farm visit, CV1) and time series forecasting CV (CV2).

For CV1, random repeated subsampling cross-validation was used by dividing 80% of the pigs for training and 20% for testing to evaluate model prediction performance at each time point. CV2 is a sparse design that simulates scenarios where budget or time constraints prevent us from weighing all pigs throughout the growth process (Figure 4.5). Specifically, the subsets of pigs from earlier and most recent time points (blue colors) were used to train models to predict BW of some pigs at the most recent time points (red colors). We considered a total of five scenarios (A to E) for predicting pigs at T2 through T6. For example, in Scenario A, pigs marked with "X" were excluded and pigs marked with blue at T1 and T2 were used as the training set to predict pigs marked with red at T2. Similarly, in Scenario B, pigs from T1 to T3 were used for training and T3 was used for testing. We varied the ratio of the number of excluded pigs: 75%:25%, 50%:50%, and 25%:75%. Here, 75%:25% indicates that 75% of the data points from all previous time points were used as the training set to predict 25% of the current data points being tested. The scenarios without excluding pigs from previous time points were also tested, and their performance served as the baseline. To avoid the problem of gradient explosion, we reduced the number of images to 10 images per individual for CV2. Other hyperparameters are the same as CV1, as shown in Table 4.1.

Root mean squared error (RMSE), coefficient of determination (R^2), mean absolute percentage error (MAPE), and mean absolute error (MAE) were used to evaluate model performance in the testing set.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.2)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4.3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.4)$$

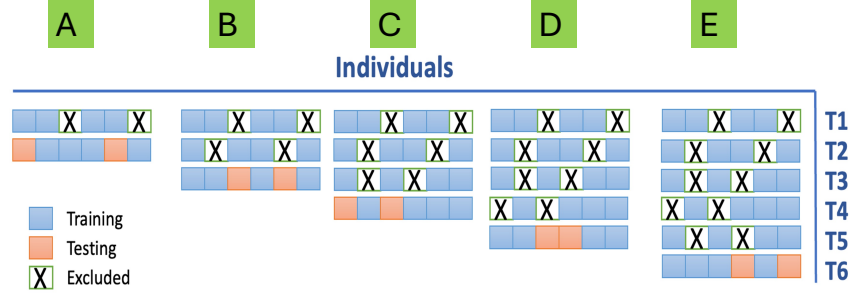


Figure 4.5: Time series forecasting cross-validation design. Subsets of pigs from earlier and most recent time points (shown in blue) were used to train models to predict the body weight of pigs at the most recent time points (shown in red). Five scenarios (A to E) were considered for predicting pigs from T2 to T6. T1 to T6 represent six farm visits.

4.4 Results

A total of 623 individual pigs were analyzed in this study. Data were collected on a commercial farm during six consecutive weighing and video recording visits over a three-month period. A total of 2,751 videos were used. To avoid an unbalanced dataset skewed towards certain pigs, we limited the selection to a maximum of 20 images per video. This uniform selection process resulted in a total of 47,529 depth images analyzed. After applying quality control to select images, we used the pre-trained neural networks, including ResNet50, MobileNet series, MobileNetV3 series, MobileViT series, and MobileViTV2 series, each augmented with an additional regression output branch for weight estimation. We fine-tuned these networks for their new task by using the selected images as input and the corresponding weights as target values.

4.4.1 Single time point cross validation

We first applied neural network models to each single time point dataset. We removed pigs at each time point with scale-based BW less than the bottom 3% of all weights and selected a maximum of 20 images per pig (Figure 4.3). After filtering and selection, 9,872, 11,286, 11,248, 11,007, 6,570, and 3,205 images were retained for T1 to T6, respectively (Table B.1). The CV results for MAPE are shown in Figure 4.6. The best MAPE for the six time points were 4.71%, 3.80%, 3.08%, 5.95%, 3.35%, and 3.77%, respectively and the best achieved MAE were 3.03kg, 2.91kg, 2.95kg, 6.14kg, 3.71kg, and 4.29kg, respectively. Among the six farm visits, T3 and T4 always had the best and worst prediction performance, respectively. The differences in performance between visits may be due to differences in image acquisition settings between farm visits. For example, the videos from T6 captured less of the BW scale and more of the background, and the videos from T4 had the depth camera set in the opposite direction compared to the other visits, with part of the barn blocking the animal's body, resulting in the overall performance of the models during these visits being inferior to that of T2-T3. This highlights the importance of establishing an optimal video acquisition system and fixed camera position for consistent model performance.

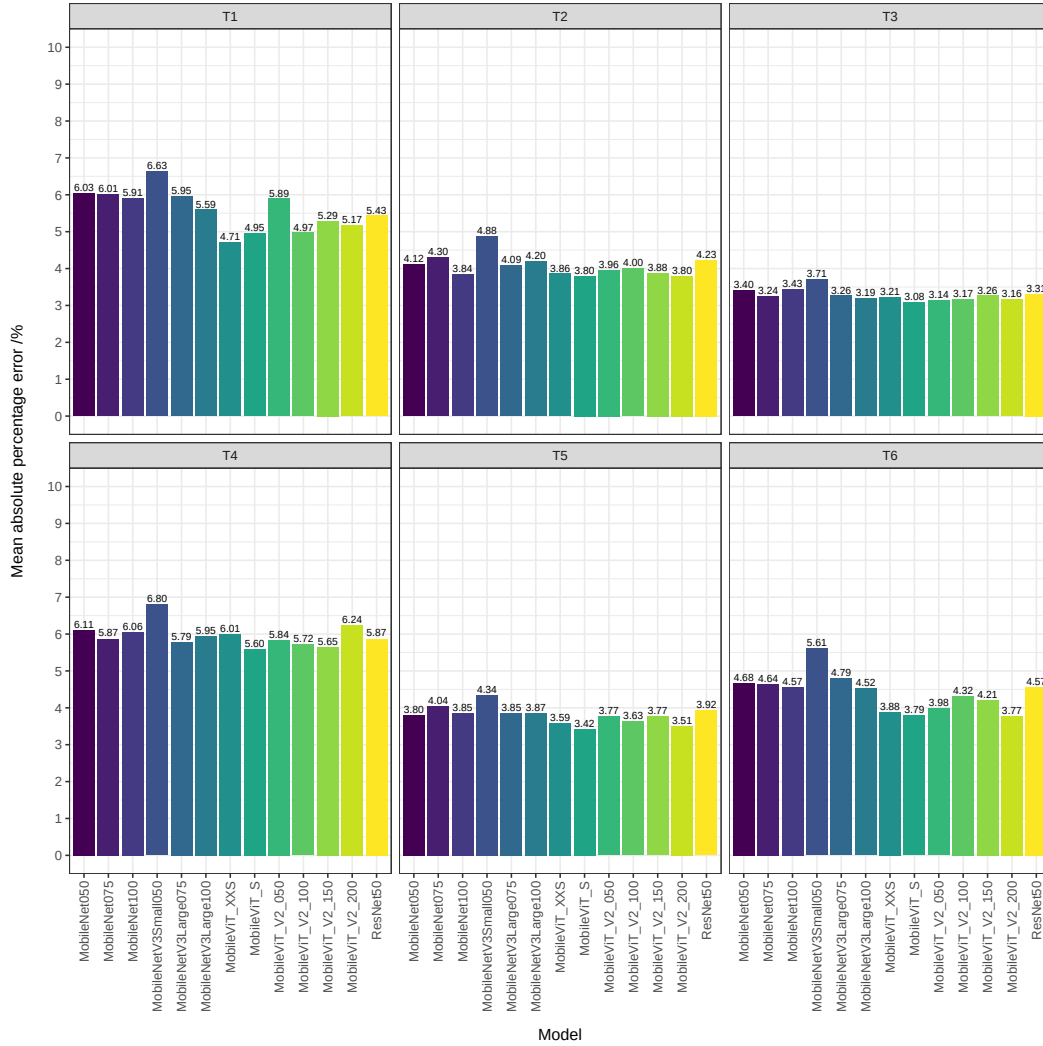


Figure 4.6: Single time point cross-validation prediction results. Thirteen models were tested: three from the MobileNet series, three from the MobileNetV3 series, two from the MobileViT series, four from the MobileViT V2 series, and ResNet50 as the baseline model. The plots show the test mean absolute percentage error over six different time points (T1 to T6) for each model.

When comparing different deep learning models on the same farm visit, despite having significantly more trainable parameters in its architecture, the performance of ResNet50 consistently lagged behind the other models across all six visits. This is evident in MAPE, where ResNet50 did not achieve the lowest error rates in any of the visits.

In contrast, the MobileViT models consistently outperformed the CNN-based MobileNet models. Specifically, the MobileViTV2-100 model showed the best performance on most visits and achieved the lowest MAPE values, indicating superior predictions for BW. The MobileViT models, whether V1 or V2 variants, consistently showed better performance than the MobileNet and MobileNetV3 variants, highlighting the effectiveness of the ViT approach over traditional convolutional methods for a prediction task. The analysis of each visit clearly shows a remarkable pattern. In T1, MobileViT-XXS achieved the lowest MAPE of 4.71%, better than the highest MAPE of 6.63% achieved by MobileNetV3-Small-050. In T2, MobileViT-S led with a MAPE of 3.80%, compared to the highest MAPE of 4.88% for MobileNetV3-Small-50. In T3, MobileViTV2-100 achieved a MAPE of 3.08%, while MobileNetV3-Small-050 was the worst performer with a MAPE of 3.71%. In T4, MobileViTV2-150 achieved a MAPE of 5.95%, while MobileNetV3-Small-050 achieved the highest MAPE of 7.43%. In T5, MobileViTV2-150 had the lowest MAPE of 3.35%, while MobileNetV3-Small-050 had the highest MAPE of 4.33%. In T6, MobileViTV2-200 had a MAPE of 3.77%, while MobileNetV3-Small-050 had the highest MAPE of 5.61%.

The averaged training history of CV1 for 300 epochs is shown in Figure B.1, which shows that the degree of overfitting is rather small. The training sets achieved the lowest MSE of 0.88 kg^2 , 0.62 kg^2 , 0.76 kg^2 , 2.12 kg^2 , 2.08 kg^2 , 2.33 kg^2 , and 2.53 kg^2 over six time points, indicating a good linear fit between the estimated and ground truth BW.

For a more detailed evaluation of the model performance on the test data, we selected the best model at each visit (Figure 4.6) and compared the predicted and observed weight in the test set without filtering or removing any individuals. The results are shown in Figure 4.7. We then divided the data into 10 kg intervals. The results are shown in Table 4.2 and Figure 4.8. Overall, the middle weight intervals showed better prediction results compared to the tails of the weight interval distribution. This means that increasing the number of

images in the training set subgroups generally improved the prediction performance of the test set subgroups. In contrast, when the training set subgroups contained the small number of images, the model's performance on the test set subgroup was unstable.

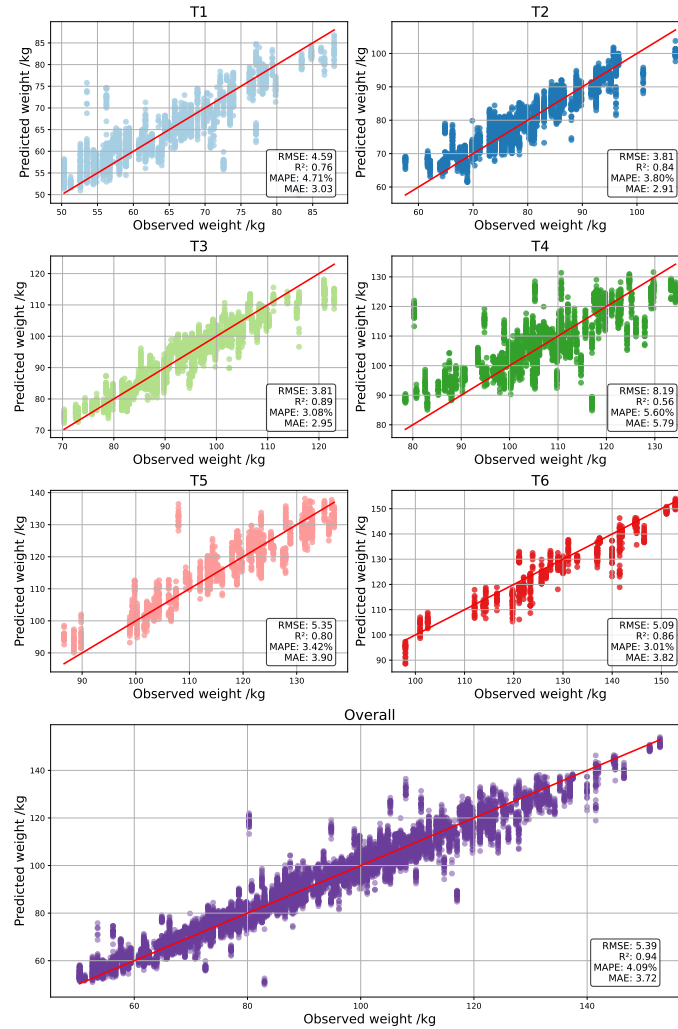


Figure 4.7: Single time point prediction results in the testing sets using the best model at each visit, along with the overall performance. The plots compare the predicted weight with the observed weight (ground truth). The red line represents the slope of 1. T1 to T6 represent six farm visits. RMSE: root mean squared error; R^2 : coefficient of determination; MAPE: mean absolute percentage error; and MAE: mean absolute error.

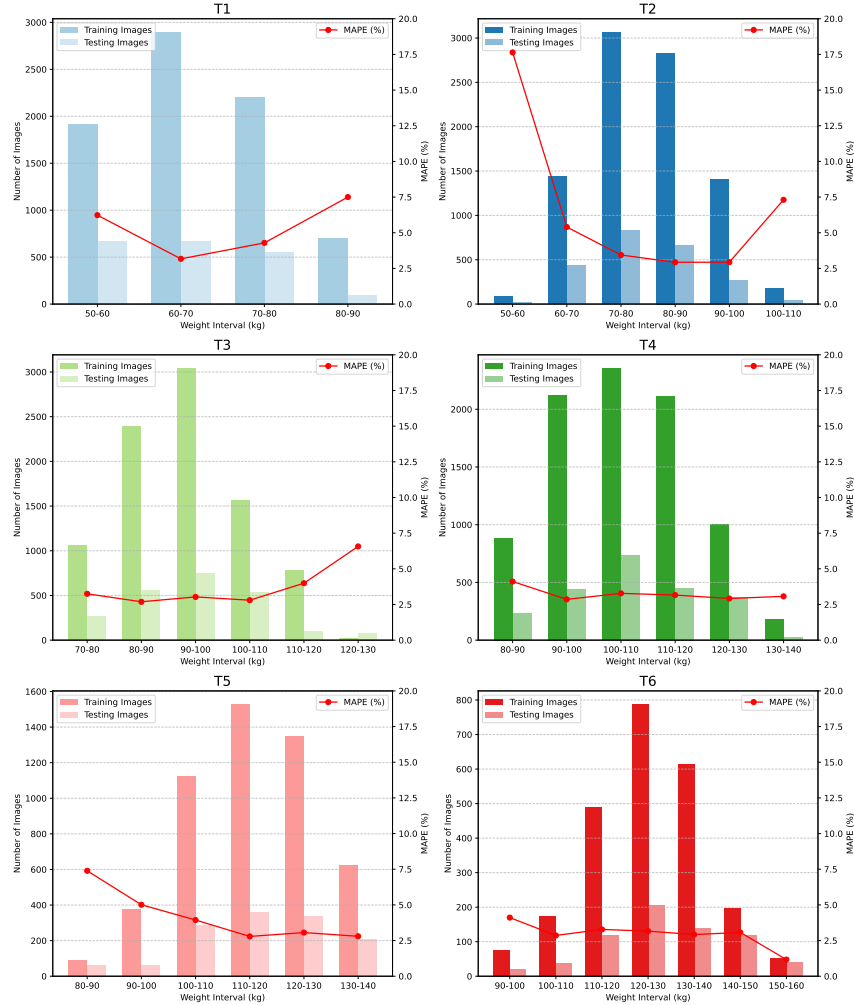


Figure 4.8: Single time point prediction results for body weight subgroups using the best model at each visit. The mean absolute percentage error (MAPE) on the test set for 10 kg weight intervals, along with the number of images in the training and testing sets. T1 to T6 represent six farm visits.

We used Grad-CAM to visualize the activation heatmaps for one of the best performers (MobileViT-S) to compare model performance. Figure 4.9 presents activation maps for arbitrarily selected light, average, and heavy pigs from T5 with the best predictions. The Grad-CAM patterns were consistent across different weight classes, indicating that the model's

attention remains similar regardless of the pig’s weight. In particular, the central part of the pig’s body was frequently highlighted, suggesting that this region is critical to the model’s predictions. Furthermore, Figure 4.9 illustrates that for pigs with lighter BW (e.g., less than 101 kg), the highlighted areas of the central body are relatively smaller and weaker compared to those of heavier pigs. This observation is consistent with the model’s performance metrics shown in Figure 4.7 and Figure 4.8, suggesting that the reduced prediction accuracy for pigs with lighter BW may be linked to the model’s reduced attention to the central body part in these cases. Figure B.2 shows Grad-CAM heatmaps for the top 50 predictions from T5.

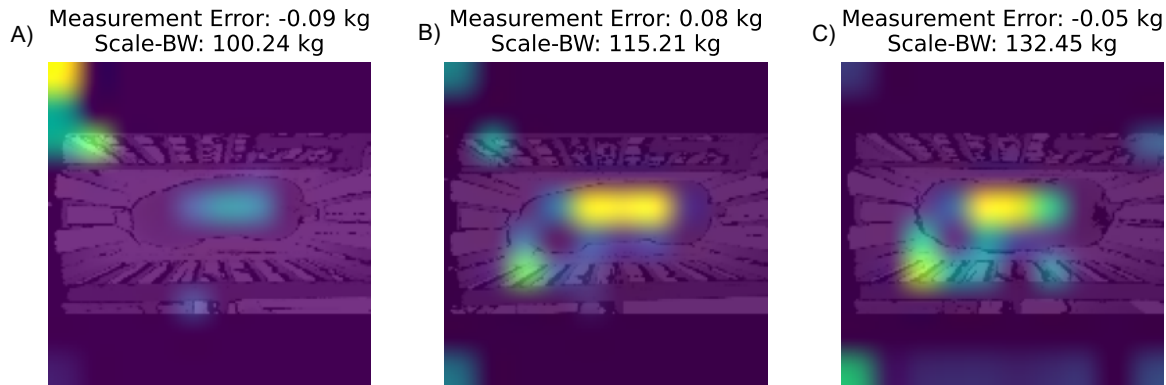


Figure 4.9: Gradient-weighted class activation mapping heatmaps for arbitrarily selected A) light, B) average, and C) heavy pigs from the T5 visit, generated with MobileViT-S.

4.4.2 Time series forecasting cross validation

Of the 13 models we used for single time point CV, MobileViT-S and MobileViTV2-200 consistently performed the best over most time points. Therefore, we selected these two models, along with ResNet50 as the baseline model, for CV2. The purpose of this CV design was to simulate scenarios where budget and time constraints prevent manual weighing of all pigs during their growth period and to assess whether the model could still maintain reasonable performance.

Table 4.2: Single time point prediction results in the testing set for 10 kg weight intervals.

Visit ¹	Weight interval (kg)	Images in training	Images in testing	RMSE (kg) ²	MAPE ³	MAE (kg) ⁴
T1	50-60	1913	668	4.76	6.23%	3.41
	60-70	2894	668	2.58	3.17%	2.06
	70-80	2206	550	4.74	4.29%	3.20
	80-90	703	92	10.31	7.50%	6.36
T2	50-60	88	20	10.21	17.64%	10.16
	60-70	1436	435	4.67	5.40%	3.53
	70-80	3064	827	3.26	3.44%	2.59
	80-90	2831	658	3.06	2.93%	2.48
	90-100	1402	271	3.73	2.94%	2.77
	100-110	176	38	7.83	7.31%	7.59
T3	70-80	1056	264	2.97	3.25%	2.46
	80-90	2393	560	2.99	2.69%	2.31
	90-100	3040	745	3.60	3.03%	2.88
	100-110	1562	531	3.76	2.80%	2.93
	110-120	786	93	5.64	3.99%	4.55
	120-130	20	80	8.36	6.57%	8.00
T4	80-90	793	240	3.90	3.34%	2.94
	90-100	2084	479	3.49	2.88%	2.76
	100-110	2606	486	5.53	3.35%	3.51
	110-120	2027	538	3.12	2.15%	2.46
	120-130	959	415	5.71	3.13%	3.91
	130-140	180	20	1.44	0.91%	1.20
T5	80-90	87	60	7.07	7.39%	6.50
	90-100	377	60	6.57	5.01%	4.99
	100-110	1125	285	7.10	3.94%	4.18
	110-120	1527	362	3.97	2.79%	3.24
	120-130	1348	336	4.77	3.06%	3.82
	130-140	623	207	4.47	2.80%	3.73
T6	90-100	75	20	4.83	4.11%	4.02
	100-110	174	38	3.34	2.86%	2.91
	110-120	489	120	4.87	3.28%	3.85
	120-130	787	206	5.05	3.16%	3.91
	130-140	614	140	5.06	2.92%	3.97
	140-150	196	120	6.44	3.07%	4.40
	150-160	52	40	1.96	1.17%	1.77

¹ T1 to T6 denote six farm visits² root mean squared error³ mean absolute percentage error⁴ mean absolute error

The results of CV2 by training the images of pigs from earlier and most recent time points to predict the BW of pigs at most recent time points are shown in Figure 4.10 and Table B.2. The best MAPE for scenarios A to E were 4.27%, 3.30%, 5.12%, 3.10%, and 3.28%, respectively. The best MAE values were 3.13kg, 3.10kg, 5.25kg, 3.53kg, and 4.01kg, respectively (Table B.2). Similar to CV1, scenarios C and E showed the worst prediction performance. When comparing the sparse and full designs, the full design showed better model performance than the corresponding sparse design in most cases.

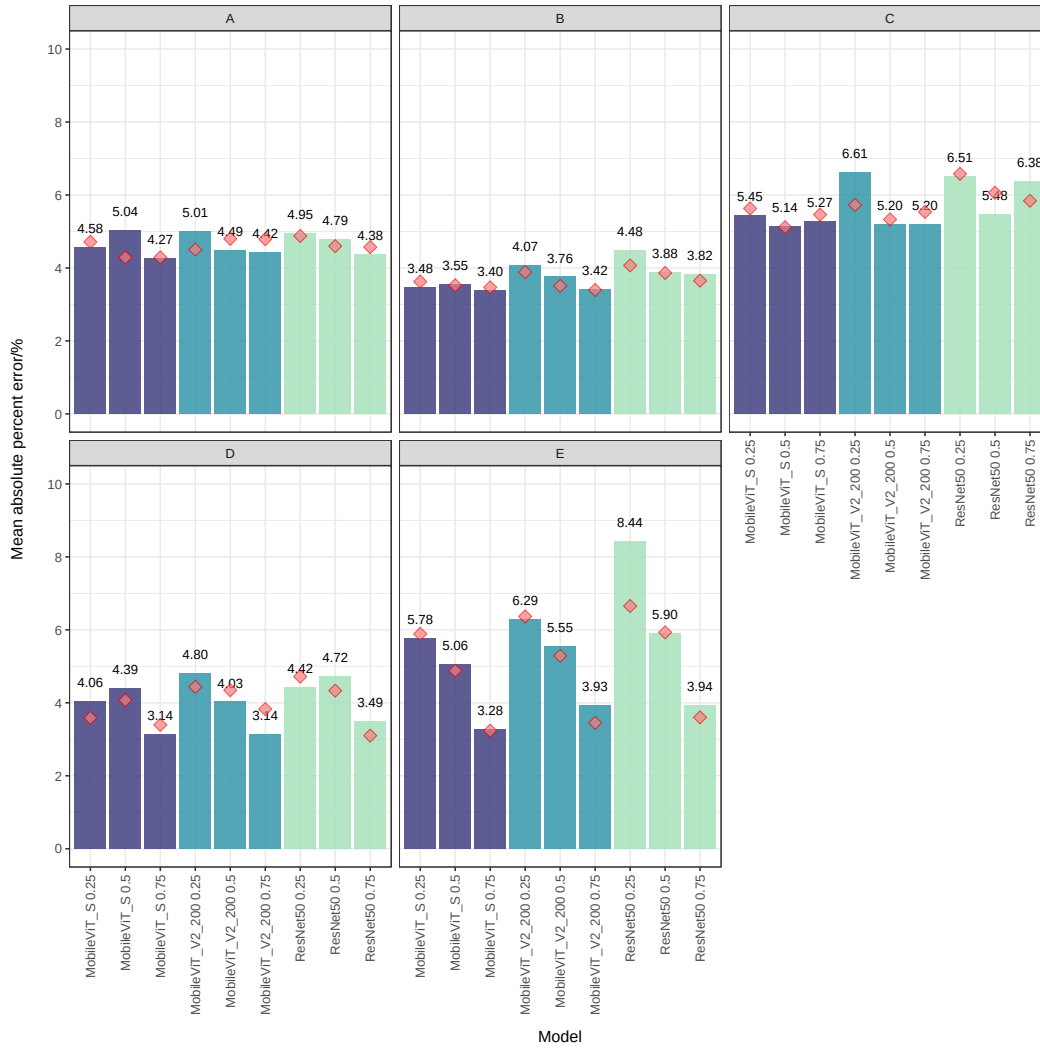


Figure 4.10: Time series forecasting cross-validation prediction results. 0.25, 0.5, and 0.75 represent scenarios for the subset ratios 25%:75%, 50%:50%, and 75%:25%. Here, 75%:25% indicates that 75% of the data points from all previous time points were used as the training set to predict 25% of the current data points. The red squares represent full design cross-validation, and the bars represent sparse design cross-validation results.

When comparing models, MobileViT-S consistently showed superior performance across all visits and training set ratios, with lower RMSE, MAPE, and MAE compared to ResNet50, while the MobileViTV2-200 model showed similar performance to MobileViT-S. ResNet50 had the highest MAPE over all visits, indicating that it is the least effective model among

those evaluated, especially at lower training ratios and in scenarios C and E.

When analyzing the performance at different training set ratios, it was apparent that the models generally performed better at a higher training ratio of 0.75, particularly in scenarios A, D, and E. For example, in scenario E, all three models had a lower MAPE at a training ratio of 0.75 compared to ratios of 0.5 and 0.25, suggesting that a larger training set improves model predictions. However, in scenario C, both MobileViT and MobileViT-V2 performed better with a training ratio of 0.5. A higher training ratio did not improve performance, suggesting that it may be more beneficial to use fewer images from previous visits in different video capture systems when training prediction models to avoid the potential misguidance of deep learning models. In contrast, the lowest training ratio of 0.25 tended to produce the worst performance, especially for the ResNet50 model, resulting in much higher MAPE values, suggesting that too small a training set can decrease prediction performance, especially for the CNN model. In summary, higher training ratios were associated with better model performance for most models and time points. MobileViT-S and MobileViTV2-200 showed robust performance across all time points and training ratios, with MobileViT-S slightly outperforming in most scenarios, while ResNet50 showed significant performance degradation at lower training ratios.

4.5 Discussion

The study aimed to demonstrate the use of deep learning models, specifically efficient CNNs and ViTs, for predicting pig BW from repeatedly measured large-scale video data obtained in a commercial environment. This research focused on using top-view depth image data as direct input to predictive deep learning models for pig BW prediction. These models provide an advantage over traditional biometric approaches by automatically extracting the

most relevant features from the images, thereby eliminating the need for predetermined feature preprocessing and overcoming the challenges posed by imperfect data acquisition environments. To the best of our knowledge, this study is the first to test the state-of-the-art efficient ViT models for predicting pig BW in relatively noisy large-scale datasets. In CV1, the average best model achieved across the test sets at six time points resulted in a MAPE of 3.08%, MAE of 2.95kg, RMSE of 3.81kg, and R^2 of 0.89 (Figure 4.7). Unlike most previous studies, which typically used fewer pigs with non-repeated BW measurements from non-commercial data, our study used a dataset of over 600 pigs with repeated measurements at six different time points over three months. For example, [22] included 234 pigs in their study, [87] included 70 pigs to build point clouds, and [122] studied 733 pigs from three locations. The dataset in the study by [108], which included about 400 pigs measured at four time points, is comparable in size to the current study.

4.5.1 Influence of video preprocessing for model training on large-scale data

The major differences between experimental farm data and commercial data are data quality and data size. For this reason, the first step was to remove noisy images before applying deep learning models for feature extraction. Traditionally, useful images were manually selected [122, 132]. To automate BW prediction and streamline the entire process, we applied a fine-tuned YOLOv8 model, which achieves real-time object recognition with lower error rates compared to manual selection. Recent studies have used YOLO series detectors for animal identification and motion detection using images [72, 187, 207]. As shown in Figure 4.2, the customized fine-tuned YOLOv8 can detect objects in depth images. The fine-tuned YOLOv8 is continuously adjusted based on validation results, ensuring robustness and adaptability to varying industry-scale depth image conditions.

To reduce the effects of varying camera positions and viewing angles between different farm visits, we converted depth images from CSV files containing distance information from the object to the camera in meters into depth grayscale images. Previous studies typically used the depth images directly, but the presence of objects or the default histogram equalization setting could cause inconsistencies in the depth information. Another major difference from the previous study is that we added zero padding to the top and bottom of each image to avoid distortion during model fitting [203]. This ensures that the images maintain their aspect ratio and that the depth information remains consistent and reliable for analysis.

To avoid misleading results, we ensured that the training and testing sets were separated so that images of the same pigs were only present in one, but not both. This separation helps to maintain the independence of the datasets. Additionally, to avoid data imbalance caused by pigs with a disproportionate number of similar images, we uniformly selected the same number of images per animal. This uniform selection ensures that each animal is equally represented, thereby increasing the robustness and fairness of our model training and evaluation processes. An alternative method that can be used to exclude similar frames in a video is the structure similarity index measure, which is used to remove similar images in tasks such as pig facial recognition, video cleaning, and image compression [52, 108, 185].

4.5.2 Utility of efficient CNN and ViT

Recent studies have shown that depth image data without feature extraction is sufficient for accurate BW prediction based on CNN models [2, 15, 203]. The earliest end-to-end model for this purpose was developed in 2019 [15]. They used image features extracted from the Faster R-CNN model and built a regression layer to predict BW. Subsequently, other CNN models, including DenseNet201, ResNet152, Xception, Inception, MobileNetV2, and EfficientNet-B0,

were applied to depth or RGB images [108, 203]. Compared to the similarly sized dataset as ours, Meckbach et al. [108] used EfficientNet-B0 to directly predict BW from depth images, achieving a MAPE of 3.9% and RMSE of 3.8 kg. Also, Nguyen et al. [122] used handheld 3D cameras to capture depth videos from different angles and built point clouds to estimate pig BW. Their model performed best on unseen data with MAE of 9.25kg, RMSE of 12.3kg, MAPE of 7.54%. The advent of ViTs marks a significant shift in computer vision. However, ViTs have seen limited adoption in animal BW prediction due to the small size of available datasets and the lack of pre-trained weights tailored to depth images in livestock. Currently, only one study has used the modified Swin transformer with cross-attention design to predict BW from depth and RGB images together for pigs living in groups of approximately 12-13 pigs [59]. They achieved an MAE of 3.24kg, MAPE of 4.08%, RMSE of 5.99kg, and R^2 of 0.74. However, it requires manual selection of the target pig in a pen, which prevents full automation and practical implementation.

Most of the existing models for pig BW prediction are computationally intensive, with numerous parameters and high resource requirements. Our study focused on evaluating efficient neural networks that can save computational resources while maintaining high BW prediction performance, with the goal of finding the most suitable models for practical deployment. We investigated the MobileNet and MobileViT model series. MobileNet, especially its later versions such as MobileNetV2 and MobileNetV3, uses depthwise separable convolutions and other techniques to reduce computational load and increase efficiency. MobileViT combines the strengths of CNNs and ViTs, leveraging convolutions for local feature extraction and transformer blocks for long-range dependencies. MobileViTV2 further refines these mechanisms with separable self-attention. Using two CV scenarios, we demonstrated that the ViT-based MobileViT models consistently outperformed the CNN-based MobileNet models. Specifically, we evaluated several predictive performance metrics, including MAPE, MAE,

RMSE, and R^2 . For example, at T3 (Figure 4.6 and Table B.1), MobileViT-S achieved the lowest MAPE of 3.08% and the lowest MAE of 2.95kg, while MobileNetV3Large100 had a MAPE of 3.19% and MAE of 3.05 kg. For other visits, the differences were more pronounced.

There are ongoing efforts in computer vision to improve the efficiency of transformers, such as Swin Transformer, DeLight, and Efficient Transformer [93, 98, 109, 183]. These approaches often reduce the number of tokens in the transformer block using methods, such as down-sampling and pyramidal structures. Transformer-based models can also benefit from mixed-precision training, efficient optimizers, and knowledge distillation to enhance performance [25, 114]. Therefore, there is still considerable potential to discover more suitable ViT models for future work in PLF.

In addition to selecting a suitable predictive model, model performance also depends on the ground truth BW distribution. We evaluated model performance in subgroup analyses with weight intervals of 10 kg, which can reveal how models perform across different categories and highlight any biases or inconsistencies. Overall, Figure 4.8 shows that model performance is related to the true group BW distribution, with increasing the number of images in the training set subgroups generally improving prediction performance on the testing set subgroups. Having the same number of images for each pig weight class could improve overall model performance and ensure consistent accuracy across weight classes. For example, intentionally collecting a larger number of videos for pigs with extreme weights (i.e., very light or heavy pigs) in the training dataset may be an option.

Pretraining on large datasets such as ImageNet has become a standard practice to enhance model performance [56, 201]. ImageNet provides a large volume of labeled data that allows for learning rich feature representations [24]. We observed that without transfer learning from ImageNet, MobileViT struggled to converge on our dataset. In our study, we used pre-trained weights from ImageNet and generated model weights specifically tailored to depth

image feature extraction. These generated model weights can serve as a valuable resource for future research on computer vision applications in PLF, allowing other researchers to take advantage of advanced architectures without the need for extensive computational resources.

4.5.3 Impact of sparse design

Sparse design was used to simulate situations where budget or time constraints prevent us from manually weighing all pigs throughout the growth process. By strategically selecting a subset of data for training, this approach aimed to balance the constraints while maintaining predictive performance. In our analysis, we found that higher training ratios generally corresponded to improved model performance across most models and time points. Although a sparse design resulted in some loss of performance compared to the full design, the use of ViT models mitigated this degradation. Specifically, MobileViT-S, MobileViTV2-200, and ResNet50 experienced average percentage decreases of 1.25%, 1.09%, and 4.40%, respectively, compared to their full design counterparts. The ViT architectures, in contrast to the traditional ResNet model, allowed for more effective utilization of limited data, thereby minimizing the performance gap between sparse and full design implementations.

4.5.4 Current limitations and future studies

The current method effectively predicts BW of individual pigs, but cannot handle multiple pigs simultaneously due to the lack of automated pig identification. To address this, we used RFID information while collecting ground truth scale-based weights. This approach can be integrated with other precision livestock equipment, such as automatic feeders and water troughs, by connecting them to the computer vision system. For future studies, an experimental setup that allows pigs to remain in their pens, such as placing a camera above

the drinking trough or feeder, would be beneficial to keep them in their familiar environment. Further efforts in pig identification can be made using computer vision methods, such as those based on the recognition of key body landmarks or visual digital ear tags. Another limitation of the current method is its ability to predict only the BW of pigs, without considering tissue components or other body dimensions. To address this gap, researchers could explore the integration of more advanced technologies, such as thermal imaging or ultrasound measurements, which could provide additional insight into tissue composition and body structure. Additionally, the use of novel computer vision techniques, such as 3D reconstruction from point clouds generated by depth cameras could provide a more comprehensive analysis of an animal's body dimensions and overall physical condition. By applying these precision technologies, it would be possible to enhance the accuracy and scope of BW predictions, leading to more detailed and useful assessments. Because more complex models typically require greater computational resources and more diverse datasets, it is important to identify efficient architectures that deliver accurate BW predictions with fewer modalities and faster processing time.

4.6 Conclusions

This study focused on predicting BW of pigs using only standardized depth videos in a commercial environment and compared the effectiveness of efficient neural networks. Through training and testing on an extensive industry-scale dataset of more than 600 pigs over six farm visits, our study demonstrated the potential of efficient neural networks. Specifically, we obtained a MAPE of approximately 3.08% and an MAE of 2.95kg on the testing data using ViT-based models CV1. In CV2, although the sparse design, which simulates budget and time constraints by using a subset of the data for training, resulted in some performance loss compared to the full design, the ViT models effectively mitigated this degradation. We believe that this study will inspire further research on the use of deep learning in PLF applied to pigs.

Chapter 5

Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice

5.1 Abstract

The asymmetric increase in average nighttime temperatures relative to increase in average daytime temperatures due to climate change is decreasing grain yield and quality in rice. Therefore, a better genome-level understanding of the impact of higher night temperature stress on the weight of individual grains is essential for future development of more resilient rice. We investigated the utility of metabolites obtained from grains to classify high night temperature conditions of genotypes, and metabolites and single nucleotide polymorphisms to predict grain length, width, and perimeter phenotypes using a rice diversity panel. We found that the metabolic profiles of rice genotypes alone could be used to classify control and high night temperature conditions with high accuracy using random forest or extreme gradient boosting. Best linear unbiased prediction and BayesC showed greater metabolic prediction performance than machine learning models for grain-size phenotypes. Metabolic prediction was most effective for grain width, resulting in the highest prediction performance. Genomic prediction performed better than metabolic prediction. Integrating metabolites and

genomics simultaneously in a prediction model slightly improved prediction performance. We did not observe a difference in prediction between the control and high night temperature conditions. Several metabolites were identified as auxiliary phenotypes that could be used to enhance the multi-trait genomic prediction of grain-size phenotypes. Our results showed that, in addition to single nucleotide polymorphisms, metabolites collected from grains offer rich information to perform predictive analyses, including classification modeling of high night temperature responses and regression modeling of grain size-related phenotypes in rice.

5.2 Introduction

Sustainable increase in food production is paramount to meet the demands of the growing population. However, rising temperatures threaten the productivity of major food crops including rice (*O. Sativa*) [133, 190, 204]. Rice is the staple food in many countries, however, its productivity is threatened by an increase in the average minimum (nighttime) temperatures. There has been a greater rise in the rate of nighttime temperatures than that of daytime temperatures [29, 177, 192]. Recent studies have indicated that high night temperatures (HNT) negatively impact photosynthesis and respiration, and hence, rice grain yield [68, 71, 134, 180, 188]. Importantly, HNT not only impacts grain yield-related traits but also grain width [27] and grain quality [168, 178] in rice. Given the increasing trend of global warming, understanding the variety of omic responses and their associations during grain development in rice is essential for improving its resilience to HNT.

Genomic prediction has been widely used to predict responses of plants and animals [113]. It is a powerful quantitative genetic approach to predict the genetic value of unphenotyped lines for diverse arrays of traits in rice [7]. In addition to DNA polymorphisms, metabolites have emerged as omics data sources that can be used to investigate biological responses. Plant metabolites play a multitude of critical roles in growth and development, and abiotic and biotic stress responses. The metabolites of plants are associated with nutrition, fragrance, and agronomic performance [126]. Differential metabolic abundance has been reported in rice grains between the control and HNT treatment [26], suggesting that the differences in biochemical or physiological signals between the two conditions are reflected in the metabolic profiles. Hence, it is worthwhile to investigate whether the metabolic profiles of genotypes alone can be used to classify control and HNT conditions (two growing conditions). When single nucleotide polymorphism (SNP) data are used as predictors, a classification accuracy of 0.5 is expected because genomics is irrelevant to the presence or absence of HNT stress.

Prediction of phenotypes using metabolites, known as metabolic prediction, has been carried out in maize [49, 149, 161, 189] and rice [196], obtaining an encouraging result for its predictive ability. Metabolic prediction captures the molecular composition of a plant, such as changes in biochemical or physiological signals that influence phenotypes, which may not be directly explained by genomic prediction [149]. Thus, metabolic data can be used to evaluate plant growth- or stress-related phenotypes in response to HNT.

Despite its potential, metabolic responses to HNT stress and the use of metabolites as covariates for complex trait prediction have not been fully explored relative to genetic analysis in rice yet. Overall, we hypothesized that the inclusion of all available metabolites would be useful for metabolic classification and prediction in HNT studies. Therefore, the objectives of this study were threefold: 1) evaluate the classification ability of metabolites to distinguish HNT conditions, 2) compare the predictive ability of metabolic prediction, genomic prediction, and their multi-omic integration for grain-size phenotypes, and 3) investigate whether the use of metabolites as auxiliary phenotypes improves the predictive performance of multi-trait genomic prediction of grain-size phenotypes under control and HNT conditions.

5.3 Materials and Methods

5.3.1 Plant materials and growth conditions

Rice diversity panel 1 lines [205] were phenotyped for grain length (major axis), grain width (minor axis), and grain perimeter in this study. These lines represent the following sub-population groups: tropical japonica (25.11%), temperate japonica (22.37%), indica (18.72%), aus (17.35%), admixed japonica (9.13%), aromatic (3.20%), admixed indica (2.74%), and admixed (1.38%) [107]. Six seedlings per accession were transplanted into 4-inch pots containing natural soil. The pots were arranged in a randomized complete block design. The HNT experiment was performed as previously described [27]. Briefly, all the plants were grown under controlled conditions until flowering. When approximately 50% of the primary panicle completed fertilization, half of the plants from each accession were transferred to HNT conditions until maturity. All the plants were harvested at physiological maturity. Dehulled mature grains from primary panicles were scanned using an Epson Expression 12000 XL scanner (Epson America Inc., Los Alamitos, CA, USA) at a resolution of 600 dpi. Morphometric measurements, including grain length, width, and perimeter, were obtained from mature grains using the MATLAB software [208]. The conventional heritability was calculated for grain-size phenotypes by treating replication and genotypes as random effects. Morphometric phenotypes were adjusted for downstream genetic analyses by deriving the best linear unbiased estimators for each accession in each condition while accounting for replication. All the lines were genotyped using a high-density rice array (HDRA) of 700k SNP markers [107]. A total of 385,118 SNP markers were used for analysis after removing SNP markers with minor allele frequencies less than 0.05.

5.3.2 Metabolic profiling

Five dehusked mature grains of each genotype were taken from the pool of all plant individuals and used for metabolite profiling. The grains were frozen and ground to fine powder by a ball mill (TissuelyzerII, Qiagen, Düsseldorf, Germany) at liquid nitrogen temperature. Around 50 mg of aliquot was weighed and used for the metabolite extraction and profiling using a 7200 GC-QTOF system (Agilent, Santa Clara, CA, USA) according to the protocol previously described [186]. The chromatography peaks were annotated to metabolites according to the retention time and mass spectral information in the Fiehn Metabolomics database (Agilent). The peak heights of representative ions for individual metabolites were normalized by that of internal standard, ribitol (m/z 319), and the fresh weights of materials to determine relative metabolite contents. The retention time and representative ion m/z of each peak and the relative metabolite contents are found in the Supplementary Files. Relative metabolite abundance was corrected for run and experimental batch effects by treating them as random, separately for the control and HNT conditions.

5.3.3 Statistical analyses

A total of 192 and 188 rice lines with phenotypes, genotypes, and metabolites were used for the control and HNT conditions, respectively. The utility of metabolic profiles to classify control and HNT conditions was evaluated. This was followed by a comparison of the predictive abilities of genomic prediction, metabolic prediction, and their combination for grain length, width, and perimeter. Finally, potential auxiliary metabolites that can be used to increase the multi-trait genomic prediction of grain-size phenotypes were explored under control and HNT conditions.

Classification of HNT stress status

The following classification models were used to classify HNT stress status based on the metabolic profiles of 380 (192 + 188) plants. In other words, we evaluated the classification accuracy of distinguishing control and HNT samples from metabolite profiling data without knowing their growing conditions because breeders or growers may be interested in knowing whether the plants had been affected by HNT. Our hypothesis was that there is sufficient differential metabolic abundance between the control and HNT conditions that can be used to classify HNT stress status.

Logistic regression: Logistic regression (LR), which is built on the logit link function, models the probability that the metabolic profile of each plant belongs to the control or HNT stress conditions.

Support vector machine: Support vector machines (SVM) coupled with a radial basis function kernel was used to find the nonlinear separation boundary. The idea behind the SVM is to maximize the margin around the separating hyperplane (control or HNT status) by solving quadratic programming.

Random forest: Random forest (RF) is an ensemble learner based on numerous decision tree classifiers constructed from subsamples of the data. Each tree in the RF predicts the category (control or HNT status) under which a new plant in the testing set belongs. The final category was assigned to a new plant according to the majority vote.

Extreme gradient boosting: Extreme gradient boosting (XGBoost) is an ensemble machine learning framework that uses gradient boosted decision trees. Relative to the gradient boosting machine, XGBoost is faster and delivers higher prediction performance. We implemented LR, SVM, RF, and XGBoost in the caret R package [83].

Metabolic prediction of grain-size phenotypes

In addition to the regression versions of SVM (i.e., support vector regression or SVR), RF, and XGBoost, ordinary least squares (OLS), best linear unbiased prediction (BLUP), and BayesC were used for the metabolic prediction of grain-size phenotypes.

Ordinary least squares: Metabolic OLS (MOLS) was constructed using metabolic abundance as a predictor in the OLS framework.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{W}_m\mathbf{a}_m + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is a vector of phenotypes (grain length, grain width, and grain perimeter); $\mathbf{1}$ is the vector of ones; μ is the overall mean; $\mathbf{W}_m$ is a centered and standardized metabolic abundance matrix; $\mathbf{a}_m$ is a vector of fixed metabolic effect, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\epsilon^2)$, is a vector of residuals. Here σ_ϵ^2 is the residual variance, and $\mathbf{I}$ is an identity matrix. The MOLS model was fitted using the `lm` function in R [144]. This model was only used for metabolic prediction because the number of SNP markers was greater than the number of accessions in genomic prediction.

BayesC: A Bayesian shrinkage and variable selection model, BayesC [81], was used to estimate the metabolic effect using the following model.

$$y_i = \mu + \sum_{j=1}^{m_m} w_{m_{ij}} a_{m_j} + \epsilon_i,$$

where y_i is the vector of phenotypes for the i th accession; m_m is the total number of metabolites; $w_{m_{ij}}$ is the j th metabolic abundance of i th accession; and a_{m_j} is the j th metabolic

abundance effect. The prior of a_{m_j} was assumed to follow a mixture distribution

$$a_{m_j}|\pi, \sigma_a^2 = \begin{cases} 0 & \text{with probability } \pi \\ \sim N(0, \sigma_a^2) & \text{with probability } 1 - \pi, \end{cases}$$

where σ_a^2 is the common metabolic abundance variance and π is a mixture proportion set to 0.99.

Metabolic best linear unbiased prediction: Best linear unbiased prediction regresses the vector of phenotypes on a kernel relationship matrix derived from the biological profiles of individuals [118]. The model considered for the metabolic best linear unbiased prediction (MBLUP) was

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_m\mathbf{u}_m + \boldsymbol{\epsilon},$$

where $\mathbf{Z}_m$ is the incidence matrix relating metabolites to phenotypic records, $\mathbf{u}_m$ is the vector of the random metabolic values of the accessions, and $\boldsymbol{\epsilon}$ is the vector of the residuals. The distributions of random metabolic effect was assumed to follow $\mathbf{u}_m \sim N(\mathbf{0}, \mathbf{M}\sigma_{u_m}^2)$, where $\mathbf{M}$ is the metabolic relationship matrix and $\sigma_{u_m}^2$ is the metabolic variance. The metabolic relationship matrix represents the similarity of metabolic profiles among accessions, which was computed as a function of the metabolic abundance cross-product $\mathbf{M} = \frac{\mathbf{W}_m\mathbf{W}_m'}{m_m}$.

Genomic prediction of grain-size phenotypes

Performance of the metabolic prediction was compared with that of genomic best linear unbiased prediction (GBLUP), which is the most commonly used genomic prediction model [176]. Here, metabolic abundance covariates were replaced with SNP marker covariates. The

GBLUP model used was

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_g\mathbf{u}_g + \boldsymbol{\epsilon},$$

where $\mathbf{Z}_g$ is the incidence matrix relating gene content to phenotypic records and $\mathbf{u}_g$ is the vector of the random additive genetic values of the accessions. We assumed $\mathbf{u}_g \sim N(\mathbf{0}, \mathbf{G}\sigma_{u_g}^2)$, where $\mathbf{G} = \frac{\mathbf{W}_g\mathbf{W}_g'}{m_g}$ is the genomic relationship matrix; $\sigma_{u_g}^2$ is the additive genetic variance; $\mathbf{W}_g$ is a centered and standardized gene content matrix; and m_g is the total number of SNP markers.

Metabolic ($h_m^2 = \frac{\sigma_{u_m}^2}{\sigma_{u_m}^2 + \sigma_\epsilon^2}$) and genomic ($h_g^2 = \frac{\sigma_{u_g}^2}{\sigma_{u_g}^2 + \sigma_\epsilon^2}$) heritability estimates were obtained from MBLUP and GBLUP, respectively. These estimates can be interpreted as the proportion of phenotypic variance explained by metabolic or genomic relationship among lines.

Additionally, the extent of increased performance due to the integration of metabolites and SNP markers was evaluated by extending MBLUP and GBLUP via multiple kernel learning as follows [5].

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_m\mathbf{u}_m + \mathbf{Z}_g\mathbf{u}_g + \boldsymbol{\epsilon},$$

This approach was named integrated genomic metabolic best linear unbiased prediction (GMBLUP). Also, we performed the Mantel test to investigate whether the correlation between the $\mathbf{G}$ and $\mathbf{M}$ matrices is statistically different [101].

The aforementioned BayesC, MBLUP, GBLUP, and GMBLUP were implemented in a Bayesian manner using the BGLR R package [135]. A flat prior was assigned to μ . The variance components, $\sigma_{u_m}^2$, $\sigma_{u_g}^2$, and σ_ϵ^2 were drawn from a scaled inverse χ^2 distribution with the degrees of freedom $\nu = 5$ and scale parameter s such that the prior means of variance components

equal half of the phenotypic variance. A total of 30,000 Markov Chain Monte Carlo samples after 10,000 burn-in with a thinning rate of 10 were used to obtain the posterior means for all the unknowns.

Multi-trait genomic prediction of grain-size phenotypes

We evaluated the gain in genomic prediction performance of grain-size phenotypes by fitting bivariate GBLUP, when metabolites were used as a correlated trait. We hypothesized that some metabolites could enhance the genomic prediction via a correlated response. All possible pairwise combinations of the target grain-size phenotypes (grain length, width, perimeter) and metabolites (auxiliary responses) were investigated. Genetic and residual variances in single-trait GBLUP were extended to the following variance-covariance structure.

$$\Sigma_g = \begin{bmatrix} \sigma_{u_{g1}}^2 & \sigma_{u_{g12}}^2 \\ \sigma_{u_{g21}}^2 & \sigma_{u_{g1}}^2 \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_{12}}^2 \\ \sigma_{\epsilon_{21}}^2 & \sigma_{\epsilon_2}^2 \end{bmatrix},$$

where subscripts 1 and 2 refer to the grain-size phenotype and metabolite, respectively. An inverse Wishart distribution was assigned to Σ_g and Σ_ϵ with degrees of freedom $\nu = 4$ and scale matrix S such that the prior means of Σ_g and Σ_ϵ equal half of the phenotypic variance. The bivariate GBLUP was fitted using 30,000 Markov chain Monte Carlo samples, 10,000 burn-ins, and a thinning rate of 10, implemented in the BGLR R package [137].

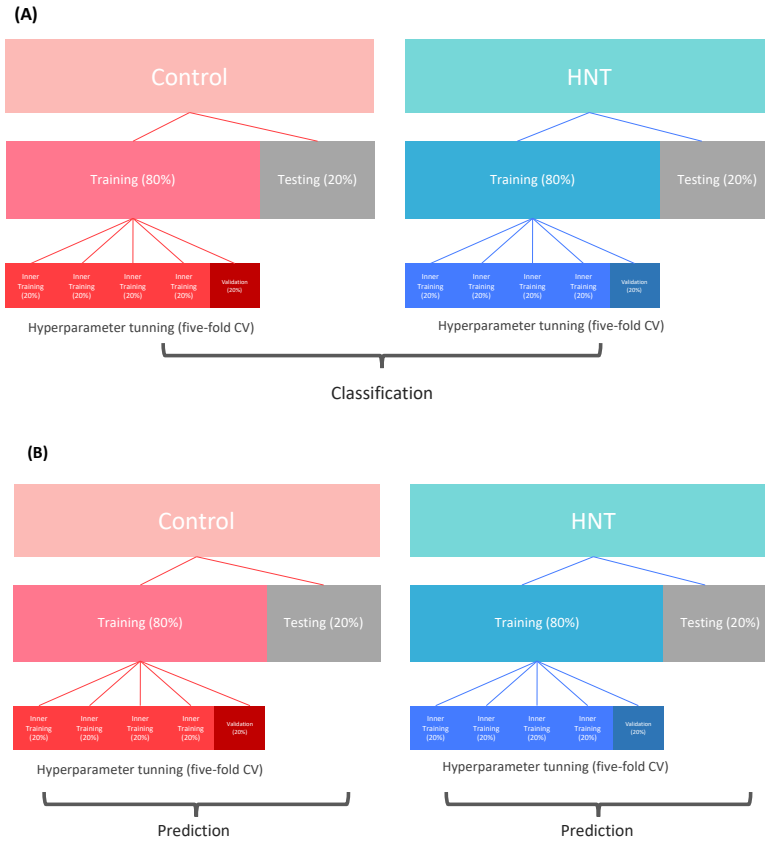


Figure 5.1: Cross-validation (CV) design for binary classification of high night temperature stress conditions (A) and metabolomic and genomic prediction of grain size related phenotypes (B).

Cross-validation strategies

Repeated random subsampling cross-validation (CV) was used to evaluate the classification and prediction model performance. For classification, we first split the accessions into training (80%) and test (20%) sets separately for the control and HNT, so that each condition was represented in the training and testing sets equally (Figure 5.1A). The training set for each condition was further split into inner training and validation sets to fine-tune the hyperparameters. The inner training set was used for hyperparameter tuning using five-fold CV. The training sets from the control and HNT groups were combined to form a unified training set.

The final model performance was evaluated in an independent testing set combined with the control and HNT conditions, which were never used in the model training. Repeated random sub-sampling CV for classification was performed 25 times. The accuracy of classification performance was derived as $\frac{TP+TN}{TP+TN+FN+FP}$, where TP, TN, FN, and FP are the number of accessions in the true positive, true negative, false negative, and false positive classes, respectively. Since the number of accessions in the control and HNT conditions were not exactly the same (192 and 188, respectively), we also evaluated classification performance using the F1 score and the area under a receiver operating characteristic (ROC) curve (AUC). The F1 score is robust to imbalanced data and is defined as the harmonic mean of the precision and recall $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$. The AUC measures the area under the entire ROC curve, which plots the TP rate ($\frac{TP}{TP+FN}$) vs. the FP rate ($\frac{FP}{FP+TN}$). The accuracy and F1 score were derived using the Metrics R package [50] and the AUC was derived using the pROC R package [151]. In addition to evaluating the utility of the whole metabolic profile for classification, we investigated the classification performance of random subsets of 10, 20, 30, 40, 50, and 60 metabolites by randomly reconstructing each subset 20 times.

The performance of metabolic and genomic predictions of grain-size phenotypes was evaluated similar to the classification, except that the predictions were performed separately for the control and HNT conditions (Figure 5.1B). The predictive performance of the models was assessed using Pearson correlation between the predictive values and phenotypes of the accessions. The repeated random subsampling CV for metabolic and genomic prediction was repeated 100 times. In metabolic prediction, we also evaluated whether metabolic effects estimated in one condition could be used to predict phenotypes in another. Specifically, we trained metabolites in the HNT and predicted phenotypes in the control and vice versa. This scenario investigate the transferability of the metabolic signal across stress conditions. Two scenarios were considered for the multi-trait genomic prediction of grain-size phenotypes

(Figure 5.2). Scenario 1 included splitting the accessions into training (80%) and testing (20%) sets. The models were trained in the training sets, and the predictive performance of the genomic prediction was evaluated in the remaining testing sets. Scenario 2 included the metabolic information of all accessions in a training set and assessed the genomic prediction performance of grain-size phenotypes using a testing set. The repeated random sub-sampling CV for multi-trait genomic prediction was repeated 25 times.

5.3.4 Subpopulation structure

Principal component analysis was performed to evaluate whether the known rice diversity panel subpopulation information can be captured by SNP markers or metabolites. The genomic and metabolic relationship matrices were used to extract principal components. The small number of genotypes within subpopulations was not large enough to conduct genomic or metabolic prediction for each subpopulation. For this reason, leave-one-subpopulation-out CV was performed to assess the impact of subpopulation structure on prediction accuracy by predicting each subpopulation from the remaining subpopulations.

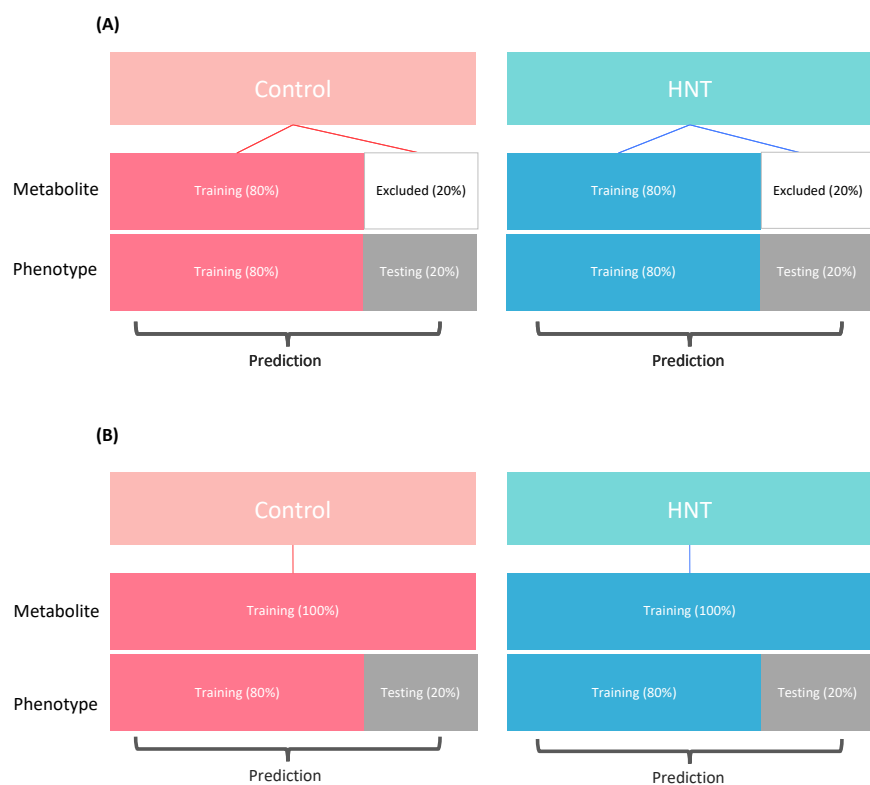


Figure 5.2: Cross-validation design for multi-trait prediction. Scenario 1 (A) and Scenario 2 (B).

5.4 Results

5.4.1 Correlation analysis

Metabolic profiling of rice grains detected 73 metabolites (Table C.1). Pairwise comparisons (r) of metabolic abundance revealed correlated metabolites (Figure C.1). Under control conditions, four metabolites, citraconic acid, arabinose, lyxose, and ribose were positively associated with each other ($|r| > 0.9$) (Figure C.1 A). Pairs of leucine and valine, isoleucine and valine, isoleucine and leucine, and ornithine and citrulline also showed notable positive correlations. Under HNT, leucine was positively associated with isoleucine, arabinose, and ribose ($|r| > 0.9$). Citraconic acid was positively associated with ribose, adenine, and uridine ($|r| > 0.9$). Furthermore, pairs of dihydrouracil and asparagine, glutamine and asparagine, ribose and arabinose, and isomaltose and lyxose showed notable positive correlations ($|r| > 0.9$) (Figure C.1A). When the metabolic abundance was expressed in terms of the ratio of control to HNT, arabinose and citraconic acid, arabinose and ribose appeared as two positively correlated metabolic pairs ($|r| > 0.9$) (Figure C.2). Grain width was associated with many metabolites (Figure 5.3). In particular, proline, serine, aspartic acid, tyrosine, glucose, lysine, tryptophan, galactinol, and sucrose were positively correlated ($r > 0.3$), whereas trehalose was negatively correlated ($r < -0.3$) with grain width under control conditions (Figure 5.3A). In contrast, melibiose was positively correlated ($r > 0.3$), while trehalose was negatively correlated ($r < -0.3$) with grain width under HNT conditions (Figure 5.3B).

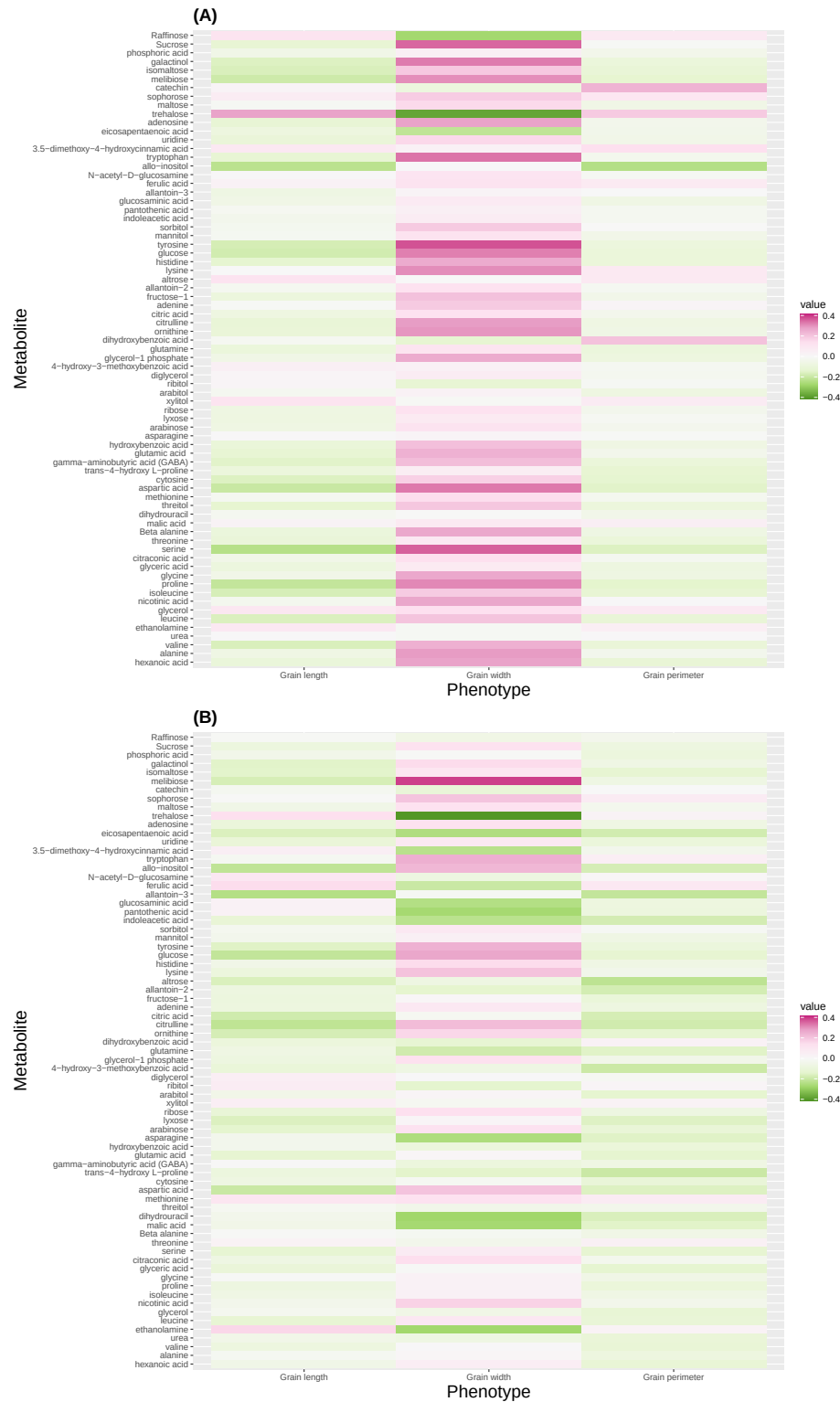


Figure 5.3: Correlation between metabolites and grain size phenotypes in control (A) and high night time temperature stress (B) conditions.

5.4.2 Evaluation of metabolic classification performance

RF and XGBoost delivered the best classification accuracy equally, followed by SVM and LR (Figure 5.4). The means of classification for RF and XGBoost were both 0.98, suggesting that the metabolic profiles of the control and HNT conditions could be accurately used for classification. The mean SVM accuracy decreased moderately to 0.78. However, the LR classification performance was worse than that of a random classifier with a mean accuracy of 0.41 and a large CV uncertainty. Because slightly different number of accessions were used between the control and HNT conditions, the classification performance of the four models was evaluated using alternative measures. The F1 scores (Figure C.3) and AUC (Figure C.4) corroborated the accuracy results, suggesting that the classification accuracy performance obtained was robust. Classification accuracy was proportional to the number of metabolites included in the model (Figure C.5). The opposite was observed in CV uncertainty, which was disproportional to the number of metabolites included in the model. A set of 10 metabolites alone achieved an accuracy above 0.8, albeit with a large CV uncertainty in RF and XGBoost. As the number of metabolites in the model increased, the accuracy increased and the CV uncertainty decreased. The accuracy of the 60 metabolites was slightly lower than that of all the metabolites. A set of 10 metabolites alone achieved an SVM accuracy of above 0.7, which gradually approached the accuracy achieved by the full set of metabolites. However, the LR did not follow this pattern. It consistently performed poorly regardless of the number of metabolites included in the model. The results of the F1 scores and AUC classification measures agreed with the classification accuracy (Figures C.6 and C.7).

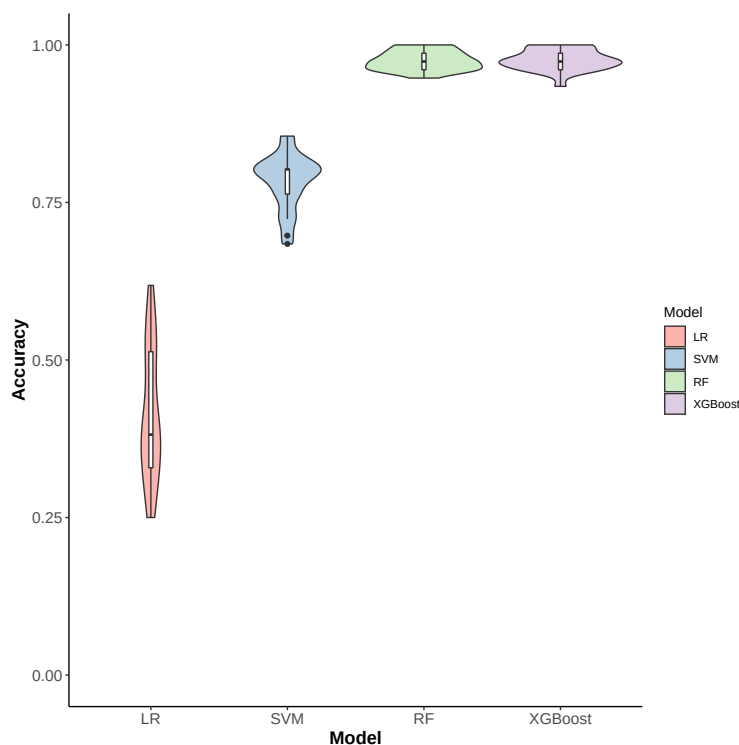


Figure 5.4: Classification accuracy of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; RF: random forest; and XGBoost: extreme gradient boosting.

5.4.3 Evaluation of metabolic prediction performance

The performance of the metabolic prediction of grain-size phenotypes using MOLS, RF, SVM, XGBoost, BayesC, and MBLUP is shown in Figures 5.5, 5.6, and 5.7. Points below the straight line indicate that the model shown on the x-axis performed better, whereas points above the straight line indicate the model shown on the y-axis performed better. BayesC and MBLUP were the equally best metabolic prediction models for grain length and delivered similar predicted values (Figure 5.5). Their mean predictive correlations were 0.35 and 0.33 in control and 0.33 and 0.31 in HNT conditions, respectively. However, although

the means were similar, BayesC was better than MBLUP in 69% (control) and 80% (HNT) of the CV resampling runs. MOLS resulted in the worst performance, with mean predictive correlations of 0.20 (control) and 0.14 (HNT). The prediction performance of BayesC and MBLUP were higher than that of MOLS in more than 75% (control) and 84% (HNT) of the resampling runs. The metabolic prediction performance of the remaining models, RF, SVR, and XGBoost, was between that of BayesC or MBLUP and MOLS. For example, BayesC performed better than RF, SVR, and XGBoost in 84%, 77%, and 76% of the resampling runs in the control and 59%, 61%, and 59% of the resampling runs in the HNT.

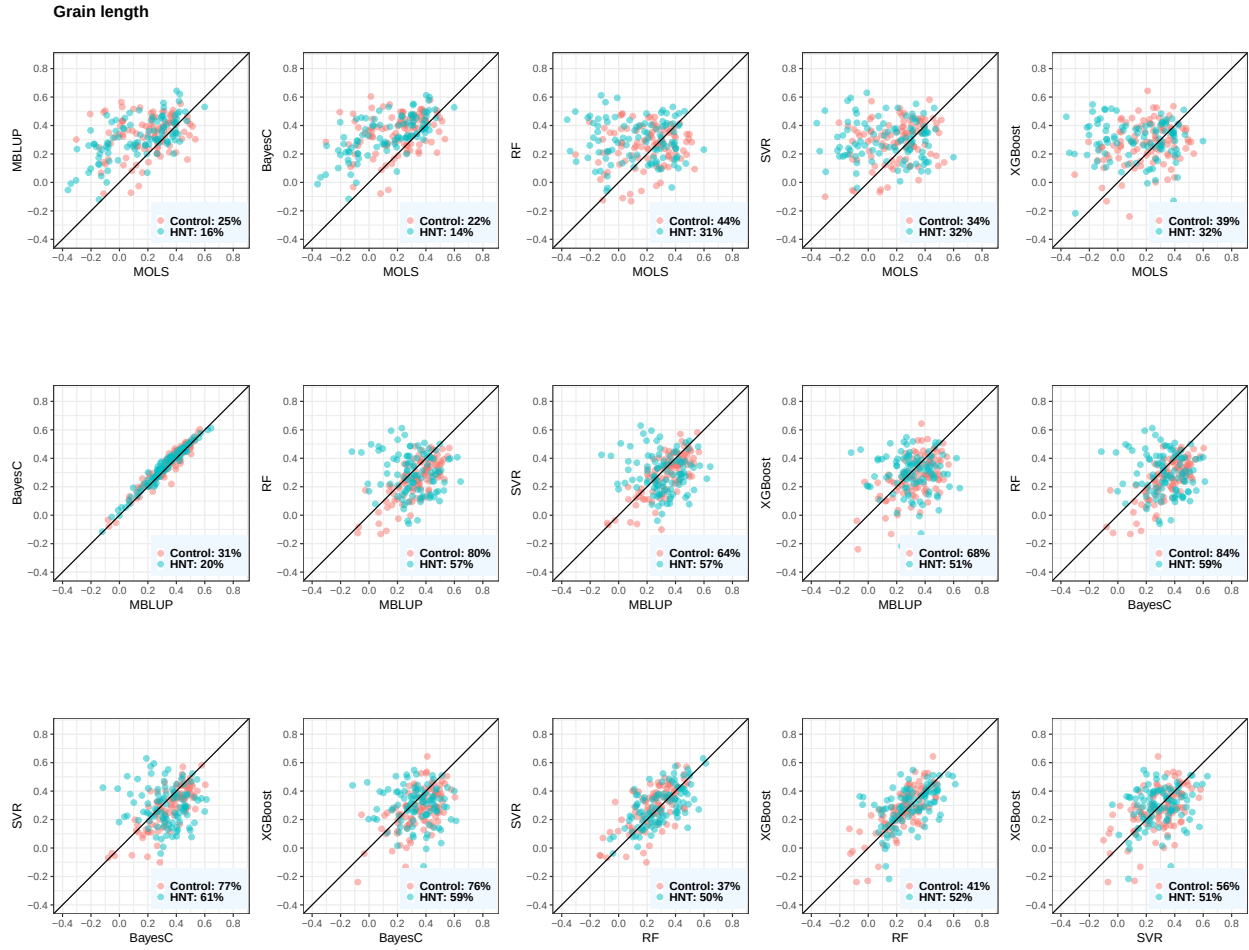


Figure 5.5: Predictive correlations of grain length using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.

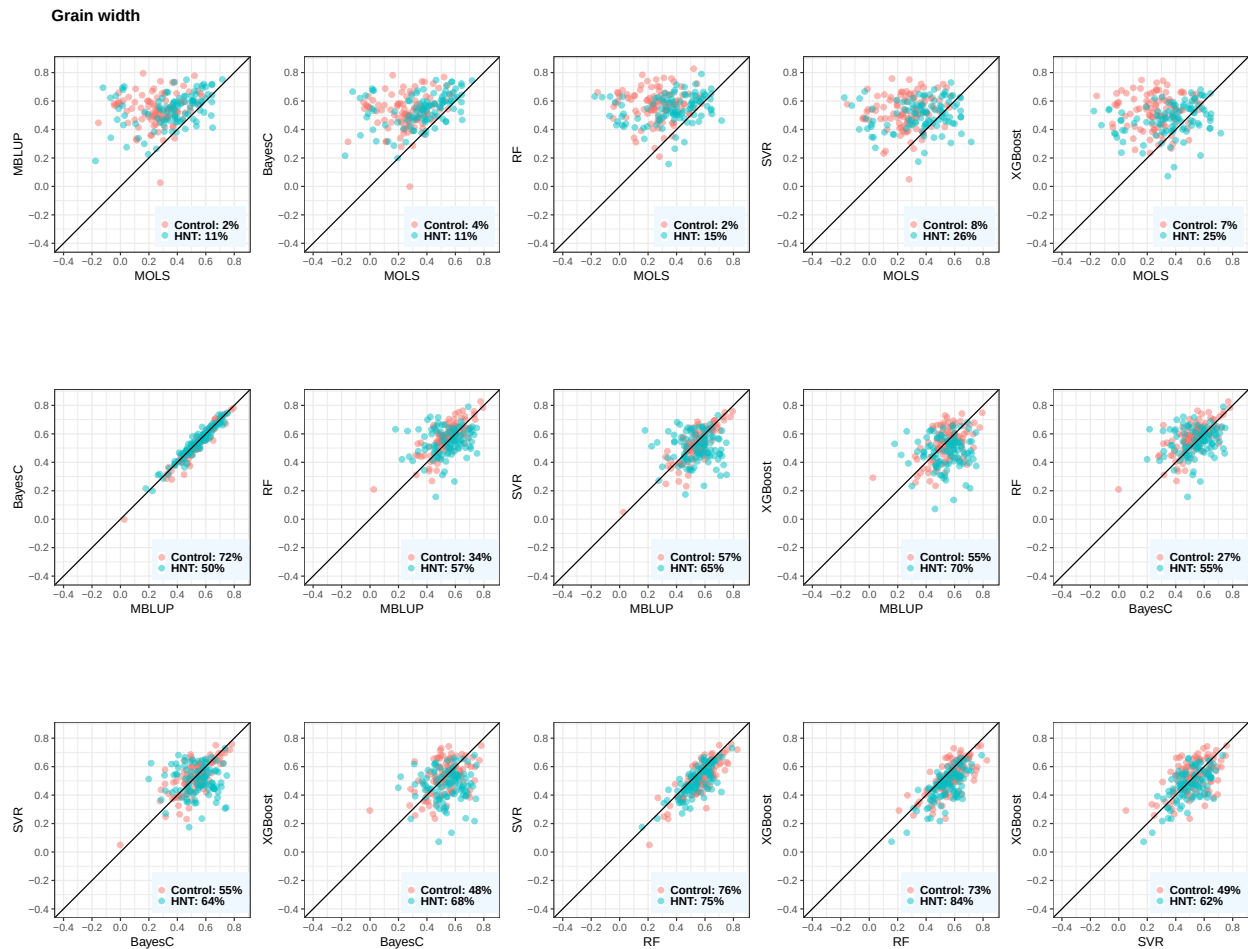


Figure 5.6: Predictive correlations of grain width using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.

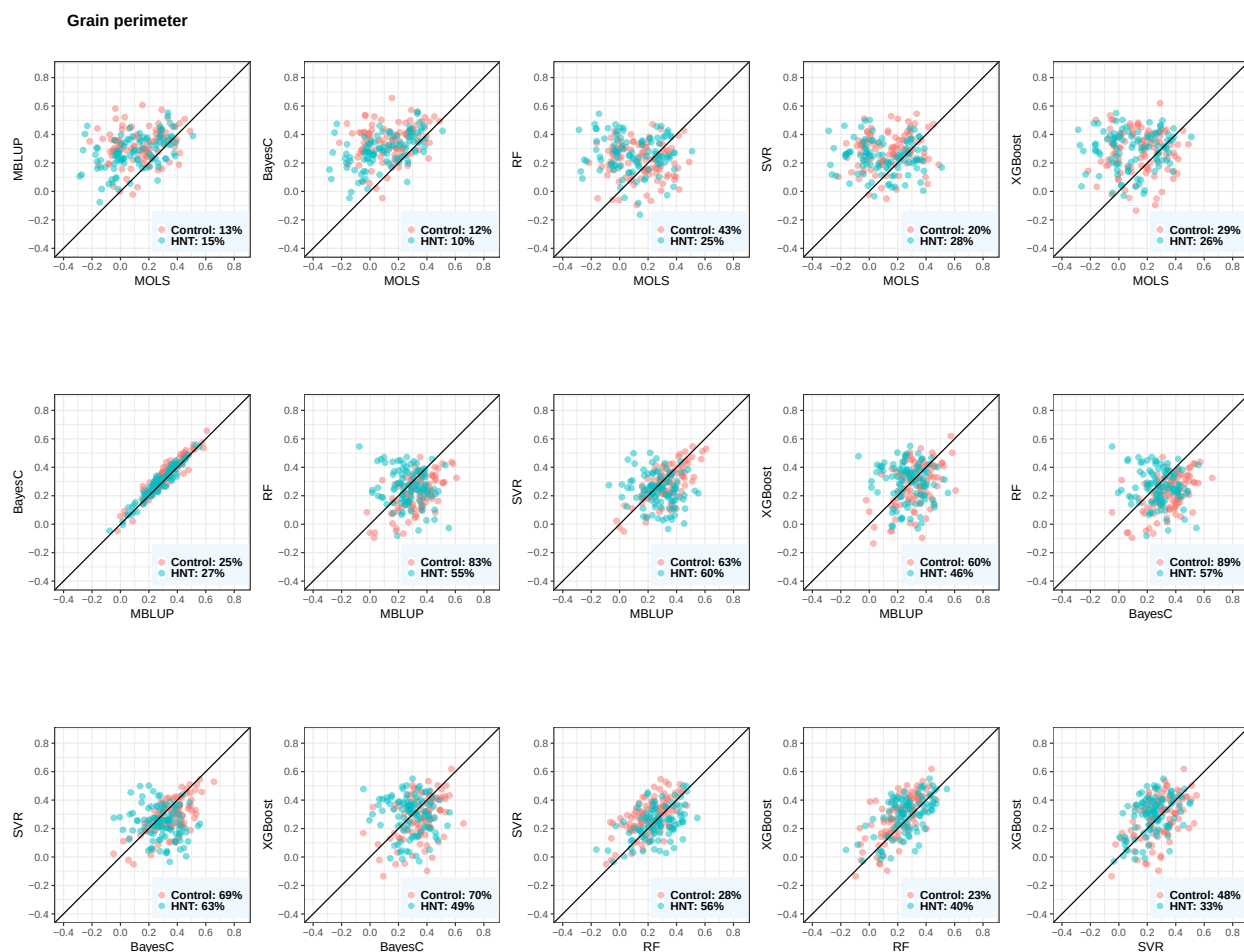


Figure 5.7: Predictive correlations of grain perimeter using metabolic prediction in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MOLS: metabolic ordinary least squares; MBLUP: metabolic best linear unbiased prediction; RF: random forests; SVR: support vector regression; and XGBoost: extreme gradient boosting.

For grain width measured under control conditions, RF was the best metabolic prediction model with a predictive correlation of 0.57, closely followed by MBLUP of 0.54 (Figure 5.6). RF was better than MOLS, BayesC, MBLUP, SVR, and XGBoost in 98%, 73%, 73%, 76%, and 73% of resampling runs, respectively. SVR, XGBoost, and BayesC performed equally well, and their predictive performance was better than that of MOLS. In the case of grain

width measured under HNT conditions, BayesC and MBLUP equally delivered the best predictive correlation of 0.54, followed by SVR and XGBoost. For example, MBLUP showed a higher predictive performance than MOLS, BayesC, SVR, and XGBoost in 89%, 50%, 65%, and 70% of the resampling runs, respectively. Under both conditions, MOLS was the worst prediction machine.

For grain perimeter, BayesC consistently produced the best prediction (Figure 5.7). Its mean predictive correlations were 0.35 and 0.29 in the control and HNT conditions, respectively. BayesC performed better than MOLS, MBLUP, SVR, RF, and XGBoost in 88%, 75%, 69%, 89%, and 70% of the resampling runs in the control, whereas it performed better in 90%, 73%, 63%, 57%, and 51% of the resampling runs in the HNT.

Overall, BayesC and MBLUP produced similar predicted metabolic values across the three phenotypes (Figures 5.5, 5.6, and 5.7). We obtained the highest prediction for grain width, whereas the predictive correlations of grain length and grain perimeter were similar and had slightly lower predictive outcomes. No significant difference was observed between the control and HNT conditions with respect to the predicted results. Additionally, we investigated whether the metabolic abundance obtained in one condition could be used to predict phenotypes under another condition. Overall, we found a decrease in metabolic prediction across HNT stress conditions (Figure C.8). For example, when control phenotypes were predicted from HNT metabolites, we observed 23%, 4%, and 31% decrease in grain length, grain width, and grain perimeter, respectively, whereas when HNT phenotypes were predicted from control metabolites, we observed 18% and 4% decrease in grain width and grain perimeter, respectively. However, no decrease was observed in grain length.

5.4.4 Evaluation of genomic prediction performance

The Mantel test showed that **G** and **M** matrices are statistically different from each other ($p < 0.05$, $r = 0.03$). The performance of GBLUP and GMBLUP for grain-size phenotypes relative to that of MBLUP is shown in Figure 5.8. MBLUP was chosen to represent a metabolic prediction model because it performed well across the three traits under both conditions with a relatively faster computational time than BayesC. Overall, GBLUP consistently provided a better prediction than MBLUP in at least 98%, 87%, and 95% of CV resampling runs under both conditions for grain length, width, and perimeter, respectively. The mean predictive correlations were 0.64, 0.73, and 0.57 for grain length, width, and perimeter in control conditions, respectively, whereas 0.64, 0.67, 0.63 for grain length, width, and perimeter in HNT conditions, respectively. We observed mixed results for a genomics and metabolite integration model. In the case of grain length, GMBLUP did not improve the prediction when compared with that by GBLUP. GMBLUP performed better than GBLUP in only 48% (control) and 18% (HNT) of the resampling runs. However, in the case of grain width, GMBLUP performed better than GBLUP in 69% (control) and 72% (HNT) of the resampling runs. The results obtained for grain perimeter were mixed. Although the prediction performance of GMBLUP was better than that of GBLUP in 66% of the resampling runs under control conditions, GMBLUP performed better than GBLUP in only 37% of the resampling runs under HNT conditions. Overall, GMBLUP achieved a marginal gain in prediction than that achieved by GBLUP, with an average increase of 1.5%.

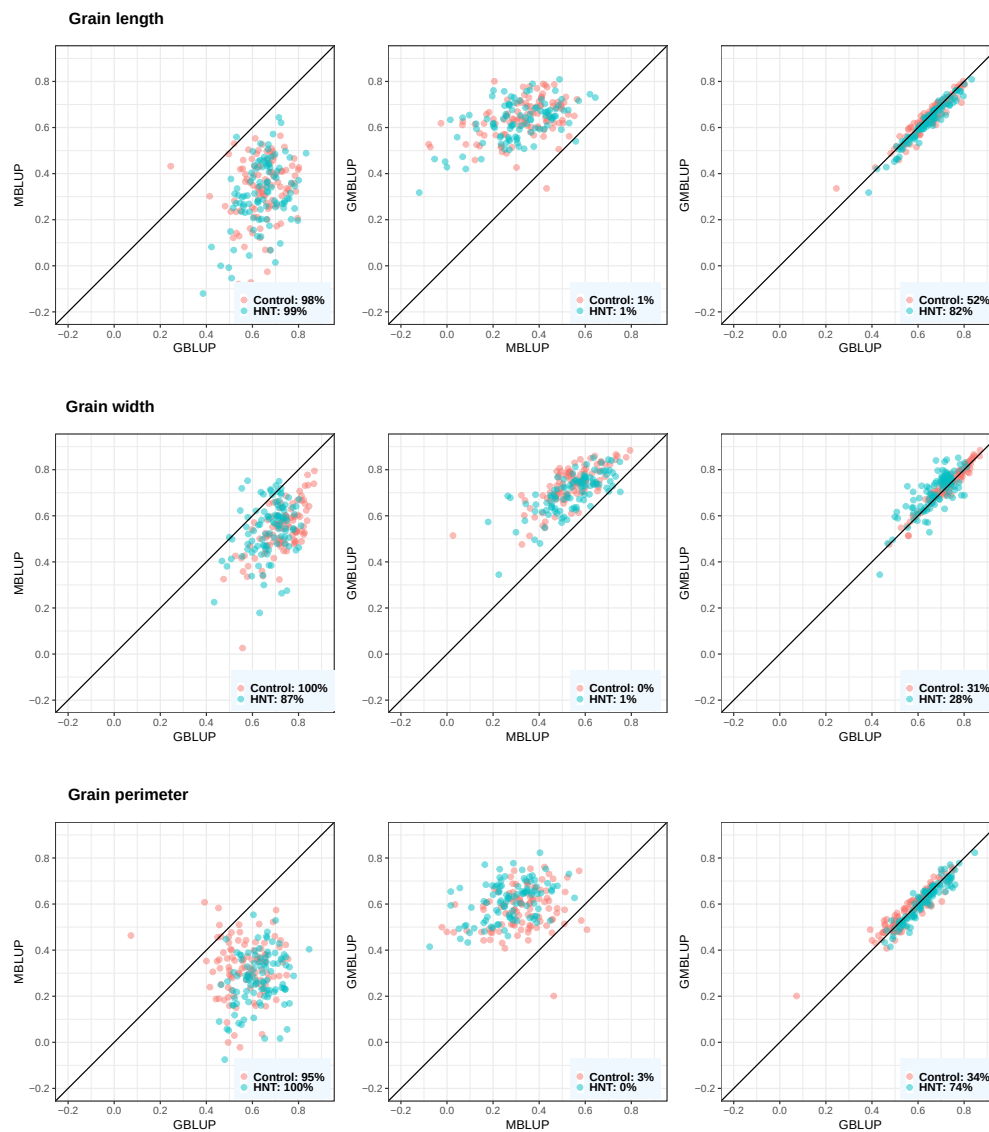


Figure 5.8: Predictive correlations of grain length, grain width, and grain perimeter using metabolic, genomic, and multi-omic prediction models in control and high night time temperature stress conditions. The percentages on the bottom right show the number of cross-validation resampling runs that the model on the x-axis performed better than the model on the y-axis. MBLUP: metabolic best linear unbiased prediction; GBLUP: genomic best linear unbiased prediction; and GMBLUP: genomic and metabolic best linear unbiased prediction.

5.4.5 Estimates of metabolic and genomic heritability

The metabolic and genomic heritabilities of grain-size phenotypes were estimated using MBLUP, GBLUP, and GMBLUP (Table 5.1). Under control conditions, genomic heritability estimates of grain length, width, and perimeter were similar, and explained at least 75% of the phenotypic variance. Grain width showed the highest metabolic heritability estimate, reaching more than half of the estimated genomic heritability. Grain length and grain perimeter showed lower metabolic heritability estimates than grain width. When the metabolites and SNP markers were fitted together, the majority of variations were captured by genomics. The estimates obtained from the HNT conditions were similar to those obtained from the control conditions. Grain width showed larger metabolic heritability estimates than grain length and perimeter. Genomics captured a large proportion of the variation when SNP markers and metabolites were simultaneously included in the model. However, the grain perimeter genomic heritability estimate was slightly lower than that of grain length and width. Conventional heritability estimates of grain-size phenotypes were lower than those of SNP markers (Table C.2).

5.4.6 Evaluation of multi-trait genomic prediction performance

The utility of metabolites as an auxiliary phenotype for multi-trait genomic prediction of grain-size phenotypes under control and HNT conditions was investigated in CV Scenarios 1 and 2. Given that there were three phenotypes and 73 metabolites for each condition, the bivariate GBLUP was fitted 438 times. In Scenario 1, multi-trait GBLUP consistently produced a greater predictive correlation for grain width than single-trait GBLUP under the control conditions (Figure 5.9). All the metabolites included in the analyses contributed to increased prediction. We did not observe any increase in the prediction of grain length

Table 5.1: Metabolic heritability (h_M^2) and genomic heritability (h_G^2) of grain size phenotypes in control and high night temperature stress conditions.

Method	Grain length		Grain width		Grain perimeter	
	h_G^2	h_M^2	h_G^2	h_M^2	h_G^2	h_M^2
Control						
GBLUP ¹	0.78	-	0.76	-	0.75	-
MBLUP ²	-	0.33	-	0.44	-	0.33
GMBLUP ³	0.70	0.09	0.63	0.14	0.63	0.12
High night temperature						
GBLUP	0.76	-	0.76	-	0.70	-
MBLUP	-	0.30	-	0.43	-	0.26
GMBLUP	0.67	0.08	0.60	0.15	0.61	0.08

¹ Genomic best linear unbiased prediction

² Metabolic best linear unbiased prediction

³ Genomic metabolic best linear unbiased prediction

and perimeter. In Scenario 2, we identified at least one metabolite that increased the multi-trait genomic prediction performance for each trait (Figure 5.10). Three metabolites, glutamic acid, allantoin-2, and allantoin-3, increased multi-trait GBLUP prediction more than single-trait GBLUP for grain length under HNT conditions. No metabolites improved grain length predictions under control conditions. A total of 46 metabolites improved predictions of grain width under control conditions. In particular, the gain in multi-trait GBLUP achieved by trehalose was statistically significant compared to that by single-trait GBLUP based on the paired one-sided t-test and paired one-sided Wilcoxon signed-rank test. Under HNT conditions, 11 metabolites, including ethanolamine, malic acid, dihydrouracil, asparagine, glutamine, allantoin-2, pantothenic acid, glucosaminic acid, ferulic acid, 3,5-dimethoxy-4-hydroxycinnamic acid, and trehalose, increased the genomic prediction performance for grain width. These metabolites were also identified in control conditions except for allantoin-2 and glucosaminic acid. Two metabolites, dihydroxybenzoic acid and catechin, increased the multi-trait GBLUP prediction for grain perimeter under control conditions. These values were statistically different from those of single-trait GBLUP. No

metabolites improved grain perimeter predictions under HNT. All the metabolic, genomic, and multi-omic prediction results are also summarized in Table C.3.

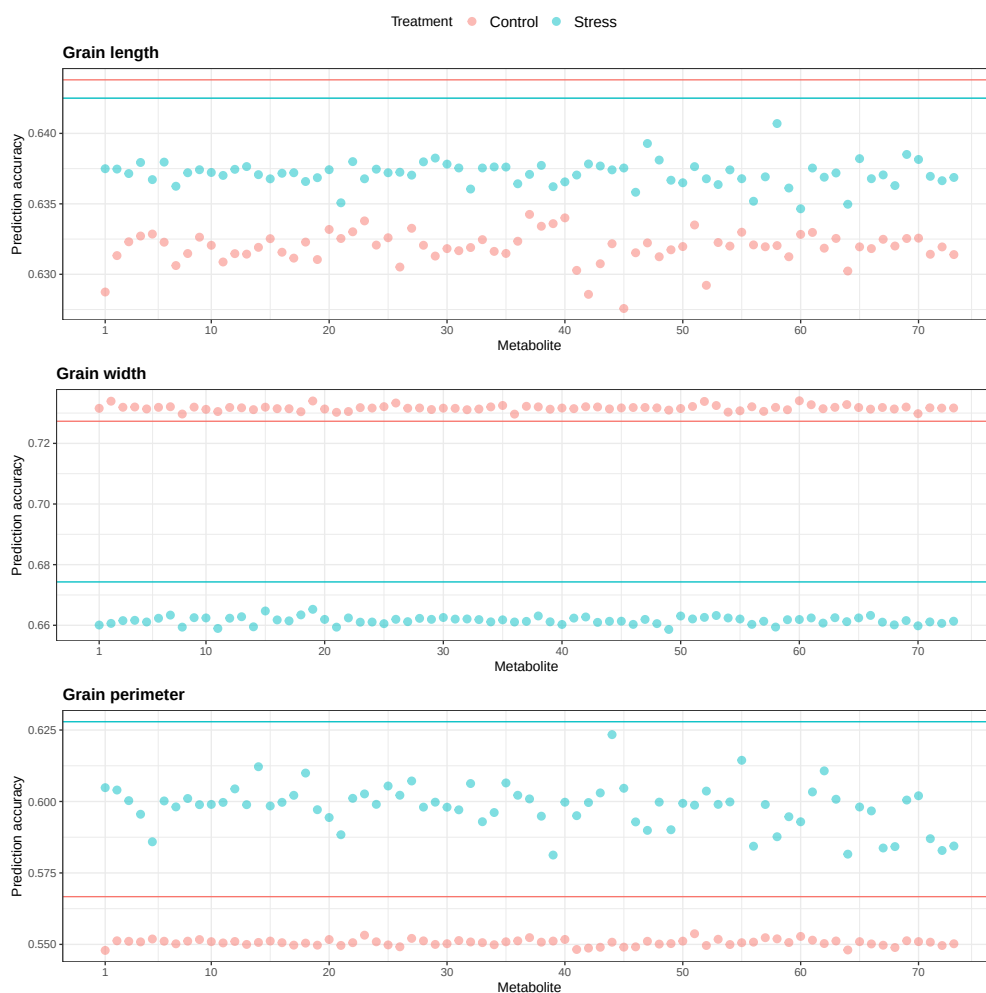


Figure 5.9: Predictive correlations of Scenario 1 multi-trait (bivariate) genomic prediction for grain length, grain width, and grain perimeter when metabolites were used as a secondary phenotype under control and high night time temperature stress conditions. The horizontal lines indicate the predictive correlations of single-trait genomic prediction.

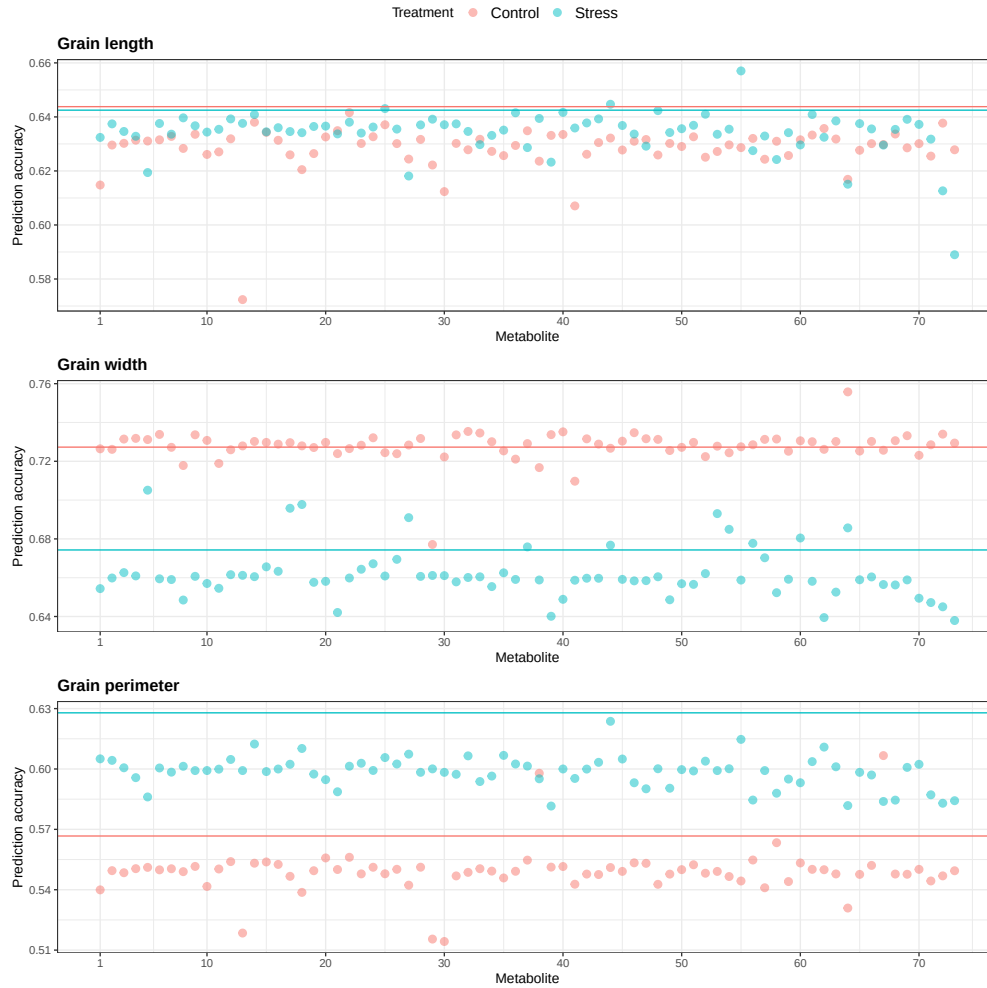


Figure 5.10: Predictive correlations of Scenario 2 multi-trait (bivariate) genomic prediction for grain length, grain width, and grain perimeter when metabolites were used as a secondary phenotype under control and high night time temperature stress conditions. The horizontal lines indicate the predictive correlations of single-trait genomic prediction.

5.4.7 The impact of subpopulation structure

Principal component analysis plots based on single nucleotide polymorphisms and metabolites are shown in Figures S9 and S10. The rice diversity panel subpopulation structure was more visible when SNP markers were used to derive principal components. Overall, the predictive correlations from leave-one-subpopulation-out CV were much lower than those of repeated random subsampling CV (Table C.4), suggesting that predicting a specific subpopulation that was not used in the training data is challenging.

5.5 Discussion

Advances in genomics and metabolic profiling have provided a new resource for studying HNT responses in rice. In this study, we evaluated the utility of metabolites for classifying HNT stress conditions and predicting grain-size-related phenotypes. In regression modeling, the performance of metabolic prediction and usefulness of metabolites as auxiliary phenotypes were evaluated in the context of genomic prediction. We found that several pairs of metabolites were correlated under the control and HNT conditions. Among the three phenotypes investigated, metabolite abundance was strongly associated with grain width. Mostly amino acids and sugars were correlated with grain width under the control condition (Figure 5.3), indicating the relationship between grain shapes and carbohydrate and protein metabolism. The correlation between tryptophan and seed width has been also reported [145]. Under the HNT condition, oligosaccharides showed correlations with grain width, which may indicate that carbohydrate metabolism plays a crucial role in determining grain shape. Interestingly, carbohydrates are a major class of metabolites which are affected by HNT in our previous studies of cereals grains [26, 69]. The correlation may reflect the changes in carbohydrate metabolism in rice grains under HNT. However, it should also be considered that differences in grain shapes among cultivars could affect metabolite composition due to the difference in the ratios of cell types with varying metabolite compositions. The relationships between metabolite accumulation and grain shapes must be carefully assessed in future experiments.

Grain length and width are prominent grain size factors that substantially impact the rice grain yield parameters [67, 128, 194]. Grain length is known to be directed by the elongation of the pericarp tissue, defined at the early stages of grain development [99, 141]. In contrast, the grain width is largely determined by cell division and proliferation of the endosperm tissue (after the fertilization event) and is a driving force for the sucrose allocation to be used for endosperm development and grain storage reserves production [12, 103]. The endosperm

cell number and proliferation (determinant of grain width) are affected by the supply of photoassimilates (sucrose) during the active grain filling stage, thus, substantially influencing the final grain weight parameters [12]. Likewise, another study reported that final grain weight is more highly correlated with grain width than grain length in the diverse winter wheat population [140]. The improvement in grain width is stated to be one of the major causes leading to increment in the final grain weight parameters for the elite wheat varieties [140], signifying the importance of this phenotypic trait for the enhancement of crop yields. Additionally, previous studies have reported that deviation from optimal temperature conditions alters the endosperm cellularization timing in rice, causing a detrimental impact on the final grain size parameters [16, 37]. Therefore, it is likely that the prolonged occurrence of HNT during grain filling stages impairs the sucrose allocation in the endosperm cells, leading to a more negative impact on the grain width than grain length, which is established before grain width. Furthermore, the genetic determinants regulating these two grain size traits (grain length and width) have distinct responses to temperature abnormalities within rice diversity panel 1 accessions [27]. Only a few accessions of rice diversity panel 1 retained both grain length and width under HNT, signifying unique genetic regulation for these grain size traits in rice [27].

5.5.1 Utility of metabolites for classification

Using all available metabolites resulted in the appropriate classification of HNT conditions with high accuracy when suitable classification models, such as RF or XGBoost, were used. This suggests that there is a differential metabolic abundance between control and HNT conditions. Logistic regression, which is a simpler classification model, was not sufficient to distinguish the signals between the control and HNT conditions. A random subset of only 10 metabolites produced moderate classification accuracy. However, this accuracy was unstable

with high uncertainty. Increasing the number of metabolites contributed to making the classification more robust than increasing its accuracy (30 metabolites provided the accuracy close to the accuracy achieved by the full set of metabolites). The accuracy achieved from a random subset of 60 metabolites was similar to that achieved from a full set of metabolites. This implies that most metabolites are altered during HNT and contributes to increasing the classification power. Although there are no previous reports investigating the classification performance of metabolites to distinguish between control and HNT conditions, our results showed that we can obtain reasonable classification accuracy.

5.5.2 Utility of metabolites for prediction and heritability analysis compared to SNP markers

Overall, BayesC and MBLUP showed relatively high and stable predictive correlations for grain length, width, and perimeter, suggesting that prediction models commonly used in genomic prediction are equally applicable to metabolic abundance data. In particular, BLUP appeared to be the most efficient method in terms of predictability and computational time. The extent of predictive correlations ranged from low to moderate. As expected from the correlation analysis, we observed the greatest predictive correlation for grain width. However, genomic prediction performance based on a 700k array was consistently better than that of metabolites. Our results agree with those of previous studies in oats, which found that metabolic prediction was not superior to genomic prediction for agronomic traits, including seed length and width [65]. Similarly, no improvement was observed in 1000-grain weight and grain number per plant in hybrid rice [182, 196, 198]. In contrast, metabolic prediction has been reported to be better than genomic prediction for oat fatty acids [65]. In addition, other related studies reported lower or comparable predictive performance of metabolites relative to SNP markers for agronomic traits related to maturity, morphology,

bioenergy, or yield-related in inbred maize [49, 197], hybrid maize [149, 161, 189], and hybrid wheat [206]. Simultaneous integration of metabolites and SNP markers in a single model moderately improved predictions for grain width in control and HNT conditions and grain perimeter in control conditions. This is partly in line with Hu et al. [65], who reported a gain in GMBLUP prediction for seed length and width in oats. On an average, the estimates of metabolic heritability were approximately half of those of genomic heritability. A smaller number of available metabolites may lead to lower estimates. Another reason could be that the metabolite-phenotype relationship is affected by the time of measurement in metabolite profiling. Unlike SNP markers, metabolite abundance varies with plant developmental stages. Nevertheless, metabolites explained a greater proportion of variability when compared at the per metabolite or per SNP level. For example, in grain width, each metabolite explained 0.6% of the metabolic heritability, whereas each SNP explained 0.0001% of the genomic heritability.

5.5.3 Utility of metabolites as auxiliary phenotypes in genomic prediction

We identified several metabolites that contributed to increased multi-trait genomic prediction of grain-size-related phenotypes. In particular, several metabolites aided the genomic prediction of grain width. Trehalose is a metabolite whose levels in grains were correlated with grain width (Figures 5.3, C.1, and C.2) and significantly contributed to the grain width prediction (Figures 5.9 and 5.10) in both control and HNT conditions. Trehalose is closely related to trehalose-6-phosphate (T6P), a signaling metabolite involved in the regulations of photosynthesis, carbon partitioning, and reproductive development [130, 142, 166]. T6P accumulation correlates with the levels of sucrose and other major sugars in wheat spikes [102]. Therefore, it is likely that the differences in trehalose and T6P metabolism that appeared

as the accumulation of trehalose in grains among rice cultivars affected grain development and shape. It is also interesting to note that trehalose accumulation is affected by HNT in developing rice and wheat grains [26, 69]. The correlation between trehalose and grain width and the trehalose contribution to the prediction may suggest the influences of trehalose metabolism in determining grain shape under HNT conditions.

5.6 Conclusions

This study showed that the metabolic profiles of rice genotypes could be used solely to classify control and HNT conditions with high accuracy. Although the metabolic prediction of grain-size phenotypes was low to moderate, they were the most effective for grain width. Genomic prediction delivered better results than metabolic prediction. The simultaneous integration of metabolites and genomics in a single statistical model yielded a minor improvement in prediction. No significant differences in metabolic or genomic predictions were observed between the control and HNT conditions. We identified several metabolites that can be used to enhance multi-trait genomic prediction as auxiliary phenotypes. These metabolites could be candidates for further studies to understand their role in the tolerance to HNT. Taken together, this study demonstrates the usefulness of rice metabolites for classification and prediction tasks under control and HNT conditions.

5.6.1 Data availability

Phenotypic and metabolic data used herein are available in the Supplementary Files at Figshare. Genotypic data regarding the rice accessions are available at the rice diversity panel website (<http://www.ricediversity.org/>). Scripts used in this work are publicly available in GitHub (https://github.com/yebigithub/VTUNL_Rice).

Chapter 6

Genomic prediction of metabolic content in rice grain in response to warmer night conditions

6.1 Abstract

It has been argued that metabolic content can be used as a selection marker to accelerate crop improvement because metabolic profiles in crops are often under genetic control. Evaluating the role of genetics in metabolic variation is a long-standing challenge. Rice, one of the world's most important staple crops, is known to be sensitive to recent increases in nighttime temperatures. Quantification of metabolic levels can help measure rice responses to high night temperature (HNT) stress. However, the extent of metabolic variation that can be explained by regression on whole-genome molecular markers remains to be evaluated. In the current study, we generated metabolic profiles for mature grains from a subset of rice diversity panel accessions grown under optimal and HNT conditions. Metabolite accumulation was low to moderately heritable, and genomic prediction accuracies of metabolite accumulation were within the expected upper limit set by their genomic heritability estimates. Genomic heritability estimates were slightly higher in the control group than in the HNT group. Genomic correlation estimates for the same metabolite accumulation between

the control and HNT conditions indicated the presence of genotype-by-environment interactions. Reproducing kernel Hilbert spaces regression and image-based deep learning improved prediction accuracy, suggesting that some metabolite levels are under non-additive genetic control. Joint analysis of multiple metabolite accumulation simultaneously was effective in improving prediction accuracy by exploiting correlations among metabolites. The current study serves as an important first step in evaluating the cumulative effect of markers in influencing metabolic variation under control and HNT conditions.

6.2 Introduction

It has long been recognized that metabolic profiles in plants are under genetic control [154, 155]. It has also been argued that the use of metabolite accumulation as quantitative traits can be an important addition to the resources that can be used to accelerate crop improvement through genome-assisted breeding [34, 82]. Plant metabolites play essential roles in growth, development, agronomic performance, quality, flavor, and stress resistance. In particular, metabolites are considered intermediates of biochemical reactions that link agronomic traits to the genome. Thus, metabolite profiles provide an opportunity to understand the physiological state of a crop. Due to recent advances in metabolite profiling techniques, the combination of genetic and metabolic approaches is increasingly being used to study end-use quality traits in crop plants [35].

Rice (*Oryza sativa* L.) is one of the most important staple crops worldwide, providing energy and nutrients to meet the needs of a growing population. Several studies have characterized the metabolic profiling of diverse rice accessions and mature grain metabolites are considered important for yield and grain quality [36, 127]. Seed quality of rice is of agronomic importance, which is influenced by the chemical composition [85]. Rice produces many kinds of bioactive compounds. Previous studies have characterized metabolic differences between accessions or subpopulations [64, 80, 84]. One strategy for integrating metabolite accumulation in rice into quantitative genetic modeling is to treat metabolite accumulation as predictors or auxiliary phenotypes to improve the prediction accuracy of target traits [11]. Genotype-metabolite association studies in rice have also been reported for primary and secondary (i.e., specialized) metabolites using QTL mapping [41, 106, 200] and genome-wide association studies [18, 19, 105] to dissect the biochemical basis of the rice metabolome. However, the investigation of the extent of metabolic variation that can be explained by whole-genome molecular markers (e.g., genomic prediction) remains fragmentary. The avail-

ability of high-coverage metabolic and single nucleotide polymorphism (SNP) data provides an opportunity to perform quantitative genetic analysis of rice metabolite accumulation, such as estimation of genomic heritability and genomic correlation, or prediction of genetic values using whole-genome regression analysis [112, 199].

Although sustained increases in rice production and quality improvement are essential to meet the demands of a growing population, rice is susceptible to abiotic stresses such as heat, cold, drought, flooding, and salinity that result in yield losses [190, 204]. Of these, a greater increase in nighttime temperatures than daytime temperatures is of particular concern [29, 177, 192]. Previous studies have reported that high night temperatures (HNT) negatively affect grain yield and quality traits [27, 68, 71, 117, 134, 156, 159, 168, 178, 180, 188]. Because changes in biochemical or physiological levels under abiotic stress are likely to be reflected in the metabolic profiles of plants [126], quantification of metabolic levels can help measure rice responses to HNT stress. A recent study showed differential metabolic abundance in rice grains between control and HNT conditions [26].

With the increasing impact of climate change on crop productivity, evaluating the role of genetic variation in regulating the abundance of rice metabolites under control and HNT conditions can be useful for understanding rice responses to HNT stress. Therefore, the objectives of this study were 1) to quantify the extent of variation in rice metabolite content explained by all SNP and 2) to evaluate the accuracy of genomic prediction of metabolite accumulation under control and HNT conditions.

6.3 Materials and Methods

6.3.1 Plant materials and growth conditions

Rice diversity panel 1 lines [205] were used in this study. These lines (accessions) represent the following subpopulation groups: tropical japonica (25.11%), temperate japonica (22.37%), indica (18.72%), aus (17.35%), admixed japonica (9.13%), aromatic (3.20%), admixed indica (2.74%), and admixed (1.38%). Six seedlings per accession were transplanted into 4-inch pots containing natural soil. The pots were arranged in a randomized complete block design. The HNT experiment was conducted as previously described [11, 27]. Briefly, all plants were grown under controlled conditions until flowering. When approximately 50% of the primary panicle had completed fertilization, half of the plants from each accession were transferred to HNT conditions until maturity. All plants were harvested at physiological maturity. A total of 192 and 188 rice lines were obtained from the control and HNT conditions, respectively. All lines were genotyped using a high-density rice array of 700k SNP markers [107]. A total of 388,219 SNP markers were used for analysis after removing SNP markers with minor allele frequencies less than 0.05.

6.3.2 Metabolic profiling

The metabolic profiling procedures based on gas chromatography-mass spectrometry are described elsewhere [11]. Briefly, five dehusked mature grains of each genotype were taken from the pool of all plant individuals and used for metabolite profiling. Grains were frozen and ground to a fine powder using a ball mill (TissuelyzerII, Qiagen, Düsseldorf, Germany) at liquid nitrogen temperature. An aliquot of approximately 50 mg was weighed and used for metabolite extraction and profiling using a 7200 GC-QTOF system (Agilent, Santa Clara,

CA, USA) according to the previously described protocol [186]. Chromatography peaks were annotated to metabolites according to retention time and mass spectral information in the Fiehn Metabolomics database (Agilent). Peak heights of representative ions for individual metabolites were normalized to that of the internal standard, ribitol (m/z 319), and to the fresh weights of the materials to determine relative metabolite content. Retention time and representative ion m/z of each peak and relative metabolite content are provided in the Supplementary files. The logarithmic transformation was applied to the metabolite abundance values as the majority of the metabolites followed a right-skewed distribution. They are further corrected for run and experimental batch effects by treating them as random, separately for the control and HNT conditions. A total of 66 metabolite abundance values were used as response variables for the subsequent genetic analysis. The list of metabolite contents analyzed in the current study is presented in Table C.1.

6.3.3 Genomic heritability of metabolite accumulation

A total of 162 rice lines with metabolite accumulation and SNP marker data in both control and HNT conditions were used for quantitative genetic analysis. The amount of metabolite content variability explained by SNP was evaluated by estimating genomic heritability. A single-trait genomic best linear unbiased prediction (GBLUP) model was

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is a vector of metabolic abundance values; $\mathbf{1}$ is a vector of ones; μ is the overall mean; $\mathbf{Z}$ is the incidence matrix relating gene content to metabolic abundance values; $\mathbf{u}$ is a vector of random additive genetic values of the accessions; and $\boldsymbol{\epsilon}$ is a vector of residuals. We assumed $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}\sigma_u^2)$, where $\mathbf{G} = \frac{\mathbf{W}\mathbf{W}'}{m}$ is the genomic relationship matrix between

the accessions; σ_u^2 is the additive genetic variance; $\mathbf{W}$ is the centered and standardized gene content matrix; and m is the number of SNP markers. Further, we assumed $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\epsilon^2)$, where $\mathbf{I}$ is the identity matrix and σ_ϵ^2 is the residual variance. Given the restricted maximum likelihood estimates of the variance components from the GBLUP model, genomic heritability estimates were calculated as $h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}$. Genomic heritability estimates were inferred using the sommer R package [23] for each metabolite accumulation individually.

6.3.4 Single-trait genomic prediction of metabolites

The single-trait GBLUP model described above was used to evaluate the accuracy of genomic prediction of metabolite abundance values. Predictive performance was evaluated by repeated random subsampling cross-validation (CV), repeated 100 times. We split the accessions so that 80% and 20% of the accessions were included in the training and testing sets, respectively. The predictive performance of the model was assessed using the Pearson correlation between the observed metabolic abundance values and the predicted metabolic abundance values.

We also evaluated whether capturing non-additive genetic effects would result in improved prediction of metabolite abundance values using reproducing kernel Hilbert spaces regression (RKHS) [40]. RKHS uses the Gaussian kernel ($\mathbf{GK}$) instead of $\mathbf{G}$. The Gaussian kernel matrix is equivalent to modeling additive by additive epistatic gene action up to infinite order by taking the Hadamard product between $\mathbf{G}$ matrices [73]. The $\mathbf{GK}$ is also known as a space continuous version of the diffusion kernel, which is deployed on graphs [119]. The $\mathbf{GK}$ between a pair of accessions i and j with their SNP vectors $\mathbf{w}_i \in (0, 1, 2)$ and

$\mathbf{w}_j \in (0, 1, 2)$ is given by:

$$\begin{aligned}\mathbf{GK}(\mathbf{w}_i, \mathbf{w}_j) &= \exp(-\theta d_{ij}^2) \\ &= \prod_{k=1}^m \exp(-\theta (w_{ik} - w_{jk})^2),\end{aligned}$$

where $d_{ij} = \sqrt{(w_{i1} - w_{j1})^2 + \dots + (w_{ik} - w_{jk})^2 + \dots + (w_{im} - w_{jm})^2}$ is the Euclidean distance and θ is the smoothing parameter. The role of θ is to smooth the relationship between the accessions. A large θ results in **GK** entries closer to 0 and a smaller θ results in entries closer to 1. Thus, θ controls the degree of genomic similarity between accessions. We applied kernel averaging (i.e., multiple kernel learning) by creating two **GK** matrices using extreme values of the smoothing parameters such that the means of the average off-diagonal elements of the corresponding kernels were 0.25 and 0.78, respectively.

The single-trait GBLUP and RKHS models were fit using the BGLR R package [135]. A flat prior was assigned to μ . The variance components, σ_u^2 and σ_e^2 , were drawn from a scaled inverse χ^2 distribution with degrees of freedom $\nu = 5$ and scale parameter s such that the prior means of the variance components equal half the phenotypic variance. A total of 50,000 Markov Chain Monte Carlo samples were used after 10,000 burn-ins with a thinning rate of 5 to obtain the posterior means for all unknowns.

6.3.5 Genomic correlation between the same metabolite accumulation in different treatments

Genomic correlations between the same metabolite measured in control and HNT conditions were estimated by fitting bivariate GBLUP. A multi-trait model postulates phenotypic values of a genotype in different treatments (environments) as different but correlated traits [54].

Genetic correlations between treatments provide information about the relative magnitudes of genotypic values among accessions that change between treatments. A greater deviation from the genetic correlation of 1 indicates the presence of genotype by treatment (i.e., genotype by environment) interactions for metabolite accumulation. Genetic and residual variances in the single-trait GBLUP model were extended to the following variance-covariance structure.

$$\Sigma_u = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_{12}}^2 \\ \sigma_{u_2}^2 & \sigma_{u_{21}}^2 \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_{12}} \\ \sigma_{\epsilon_{21}} & \sigma_{\epsilon_2}^2 \end{bmatrix},$$

where subscripts 1 and 2 refer to the metabolite abundance values in the control and HNT conditions, respectively. In this parameterization, $\mathbf{u}$ and $\boldsymbol{\epsilon}$ were assumed to follow a multivariate Gaussian distribution of $\mathbf{u} \sim N(0, \Sigma_u \otimes \mathbf{G})$ and $\boldsymbol{\epsilon} \sim N(0, \Sigma_\epsilon \otimes \mathbf{I})$, respectively, where $\otimes$ is the Kronecker product. An inverse Wishart distribution was assigned to Σ_u and Σ_ϵ with degrees of freedom $\nu = 4$ and scale matrix S such that the prior means of Σ_u and Σ_ϵ equal half of the phenotypic variance. The bivariate GBLUP was fit using 50,000 Markov chain Monte Carlo samples, 10,000 burn-ins, and a thinning rate of 5, implemented in the BGLR R package [137]. Since there were 66 metabolites, the above bivariate model was fitted 66 times.

6.3.6 Exploratory factor analysis

To facilitate interpretation of the genomic heritability, genomic correlation, and genomic prediction results, exploratory factor analysis [115] was performed to identify underlying latent factors controlling $t = 66$ metabolites. All metabolites were modeled as

$$\mathbf{Y} = \mathbf{A}\mathbf{F} + \mathbf{U}, \tag{6.1}$$

where $\mathbf{Y}$ is the $t \times n$ matrix of metabolic abundance values; $\mathbf{\Lambda}$ is the $t \times q$ matrix of factor loadings relating metabolite contents and latent common factors; $\mathbf{F}$ is the $q \times n$ matrix of latent factor scores; and $\mathbf{U}$ is the $t \times n$ vector of unique effects not explained by the q underlying latent common factors. Here, n is the number of accessions, q is the number of latent factors, and t is the number of metabolites. The variance-covariance matrix of $\mathbf{Y}$ was modeled as

$$Var(\mathbf{Y}) = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}, \quad (6.2)$$

where $Var(\mathbf{Y})$ is the $t \times t$ variance-covariance matrix of the metabolites; $\mathbf{\Phi}$ is the variance of the factor scores; and $\mathbf{\Psi}$ is a $t \times t$ diagonal matrix of unique variance. Here, the unknown parameters are $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$, which must be estimated from the data. We assumed $\mathbf{\Phi} = \mathbf{I}$, which yields factors with unit variance [3, 75]. Assuming $\mathbf{F} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ were estimated by maximizing the log-likelihood of $\mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi} | \mathbf{Y})$ coupled with the varimax rotation [78] using the psych R package [147]. A threshold of $\lambda > |0.4|$ was used to screen out factor loadings, followed by assigning each metabolite to only one of the factors based on its largest loading. The optimal number of factors was determined by parallel analysis [55, 61]. First, simulated data were generated from the observed data, and then the eigenvalues were extracted until the observed data had a smaller eigenvalue than the simulated data. The number of eigenvalues was considered as the number of optimal factors based on 20 simulated analyses.

6.3.7 Simultaneous regression modeling of metabolites

Evaluation of the extent of correlations between different metabolite accumulation at the genetic level and the gain in genomic prediction by leveraging correlations between metabolites are of interest in quantitative genetics. Inference of genetic correlations between different

metabolites within the same treatment and genomic prediction of metabolites considering all metabolites were performed separately in the control and HNT conditions using the MegaLLM R package [152]. MegaLLM is a multi-trait mixed effect factor model that exploits the factorization of genetic covariances between traits to speed up computation. The MegaLLM model was

$$\begin{aligned}\mathbf{Y} &= \mathbf{XB} + \mathbf{F}\mathbf{\Lambda} + \mathbf{ZU}_R + \mathbf{E}_R \\ \mathbf{F} &= \mathbf{ZU}_F + \mathbf{E}_F,\end{aligned}$$

where $\mathbf{X}$ is the design matrix including the intercepts for each trait with the effect size matrix $\mathbf{B}$; $\mathbf{F}$ is an $n \times q$ matrix of latent factors; $\mathbf{\Lambda}$ is a $q \times t$ factor loading matrix; and $\mathbf{U}_R$ and $\mathbf{U}_F$ are coefficients matrices. The following distributions were specified for the random effects.

$$\begin{aligned}\mathbf{U}_R &\sim \mathcal{N}(0, \mathbf{K}, \mathbf{\Phi}_R) \\ \mathbf{E}_R &\sim \mathcal{N}(0, \mathbf{I}_n, \mathbf{\Psi}_R) \\ \mathbf{U}_F &\sim \mathcal{N}(0, \mathbf{K}, \mathbf{\Phi}_F) \\ \mathbf{E}_F &\sim \mathcal{N}(0, \mathbf{I}_n, \mathbf{\Psi}_F),\end{aligned}$$

where $\mathcal{N}$ is the matrix normal distribution with mean, covariance between rows, and covariance between columns; $\mathbf{K}$ is a kernel matrix between accessions; $\mathbf{\Phi}_R$ and $\mathbf{\Psi}_R$ are diagonal matrices of dimension t ; and $\mathbf{\Phi}_F$ and $\mathbf{\Psi}_F$ are diagonal matrices of dimension q . A total of 10,000 Markov Chain Monte Carlo samples after 6,000 burn-ins and 7,000 Markov Chain Monte Carlo samples after 4,000 burn-ins were used for genetic correlation and multi-trait genomic prediction analyses, respectively, with a thinning rate of 5 to obtain the posterior means for all unknowns. For MegaLMM-based genomic prediction of metabolite abundance values, we evaluated the importance of additive genetic variation and non-additive genetic

variation by setting $\mathbf{K} = \mathbf{G}$ (MegaLMM-G) and $\mathbf{K} = \mathbf{GK}$ (MegaLMM-GK). The factorization implemented in MegaLMM allowed simultaneous fitting of all 66 metabolites to estimate their genetic correlations and perform multi-trait genomic prediction. The same CV scheme described in the single-trait genomic prediction was used.

6.3.8 Deep learning models

Deep learning (DL), specifically convolutional neural networks (CNN), has proven effective for analyzing image data [42]. It has the ability to approximate any function given enough data to learn from. However, previous studies applying CNN in the context of genomic prediction have directly used a SNP marker matrix, $\mathbf{G}$ matrix, or its Cholesky decomposition as input predictors, which may not fully exploit the strength of CNN [116]. To take advantage of CNN capabilities, we transformed SNP data into images following the modified version of the DeepInsight framework [163] and neural network modules. Briefly, the framework consists of rearranging similar SNPs into clusters, extracting image features, and fitting them into a regression model. An overview of these three steps is summarized in Figure 6.1. The first step transposes the original SNP matrix and splits it by chromosomes to form 12 sub-SNP matrices, then transforms each sub-SNP matrix into a feature matrix in the 2D Cartesian plane by taking the top two components from the t-SNE [175], where a point represents a SNP marker. Here, a nonlinear dimensional reduction technique, t-SNE, was used to embed high-dimensional SNP data into a low-dimensional space of two dimensions. When two SNP markers are in high linkage disequilibrium, they are expected to have similar Cartesian coordinates. The convex hull algorithm was then applied to find the smallest rectangle containing all points and rotate it since CNN assumes that the input images are arranged in a horizontal or vertical form. This serves as a conversion from Cartesian coordinates to pixels. The pixel frame size is fixed to 277×277 . Finally, the SNP codes (0 and 2) of each

genotype were mapped to these pixel locations representing different shades of gray; 0 was represented by a lighter color, while 2 was depicted in a darker color. Chromosome-specific images were created for each genotype to capture small differences in SNP codes. These steps to transform tabular SNP data into images allow the full capabilities of CNN to be applied to genomic predictions. Since each genotype had 12 images corresponding to 12 chromosomes, we used multi-branch CNN regression. In particular, we considered the following CNN architectures: VGG16 [165], ResNet50 [58], EfficientNetB7 [172], InceptionV3 [171], MobileNetV2 [160], and DenseNet201 [66]. Briefly, VGG16 emphasizes simplicity with deep convolutional layers, ResNet50 introduces skip connections for efficient training of deep networks, EfficientNetB7 uses balanced scaling of network dimensions, InceptionV3 uses inception modules for diverse receptive fields, MobileNetV2 focuses on lightweight architecture with inverted residuals, and DenseNet201 maximizes feature reuse through densely connected blocks. Due to the relatively small number of genotypes in our dataset, which typically requires thousands of images, we took advantage of transfer learning. Specifically, we used pre-training weights for the layers (excluding the top layers) from CNN models trained on the ImageNet dataset [24]. Applying these weights to each SNP image facilitated feature extraction, with the same weights shared across 12 branches. We then used the optimal ridge regression model, selected from regularization strengths of 10, 100, 1,000, and 10,000, to predict metabolite abundance values. The same CV scheme described for single-trait genomic predictions was used. We implemented these models using Python version 3.9.0 and the deep learning framework TensorFlow version 2.12.0 with the Keras API version 2.6.0.

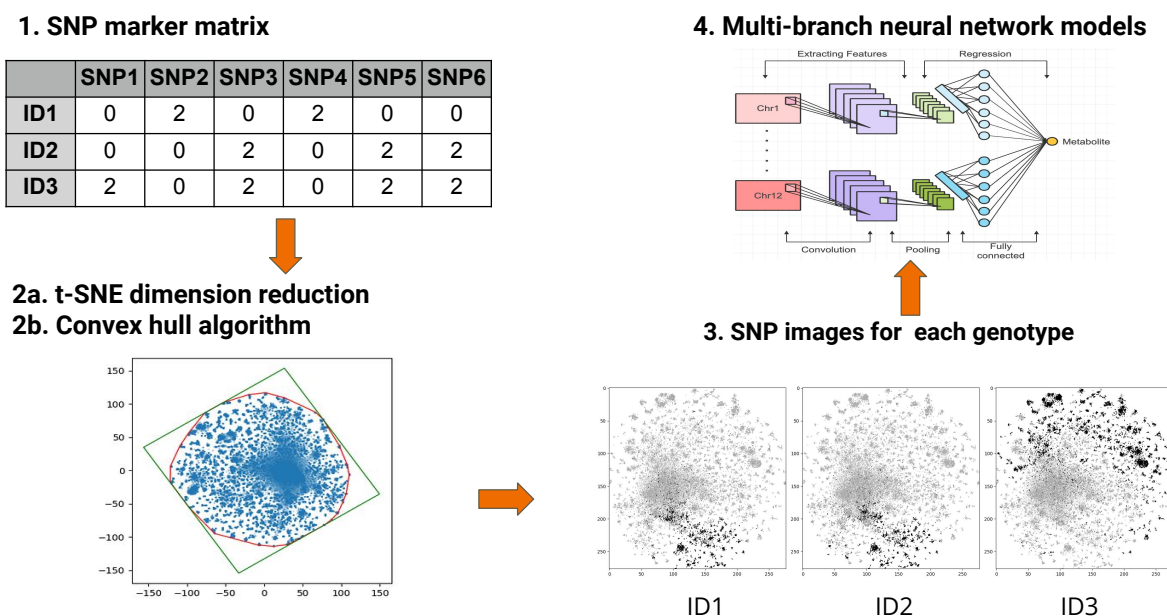


Figure 6.1: Flowchart of converting single nucleotide polymorphisms to image data.

6.4 Results

6.4.1 Single-trait analysis of metabolites

Pairwise Pearson correlations between metabolic profiles are shown in Figure D.1. Pairs between leucine-valine, isoleucine-valine, isoleucine-leucine, asparagine-dihydrouacil, pantothenic acid-glyceric acid, ornithine-citrulline, and adenosine-uridine in the control condition, as well as leucine-valine, isoleucine-valine, isoleucine-leucine, asparagine-dihydrouacil, adenine-glyceric acid, and 4-hydroxy-3-methoxybenzoic acid-hydroxybenzoic acid in the HNT condition showed positive correlation coefficients greater than 0.8. Genomic heritability estimates for metabolite accumulation are shown in Figure 6.2. The majority of metabolite abundance values were low to moderately heritable (Figure 6.2A). Genomic heritability estimates ranged from 0 to 0.80 and 0 to 0.70 in control and HNT conditions, respectively.

Overall, metabolite accumulation showed larger genomic heritability estimates in control compared to HNT conditions (Figure 6.2B). Their mean (median) values were 0.27 (0.24) and 0.22 (0.19) in control and HNT conditions, respectively. The Pearson correlation coefficient of genomic heritability estimates between control and HNT conditions was 0.50, indicating that the two treatments have a modestly similar pattern. Although most metabolites showed agreement of heritability estimates between control and HNT conditions, some metabolites deviated from this trend (Figure 6.2C). A total of 24 and 29 metabolites, respectively, showed small (< 0.05) and large (> 0.1) heritability differences between control and HNT conditions. Genomic correlation estimates of the same metabolites between the two treatments are shown in Figure 6.3. All metabolites showed genomic correlation estimates less than 1, indicating genotype by environment interactions to some extent (Figure 6.3A). The genomic correlation estimates ranged from 0.12 to 0.86. Their mean and median values were 0.48 and 0.49, respectively (Figure 6.3B). Genomic prediction accuracy of metabolite accumulation in control and HNT conditions is shown in Figure 6.4. The results varied widely between metabolites (Figure 6.4A). The accuracy of genomic predictions ranged from 0.01 to 0.66 and 0.01 to 0.63 in control and HNT conditions, respectively. Consistent with genomic heritability estimates, genomic prediction accuracies were higher in control compared to HNT conditions (Figure 6.4B). Their mean (median) values were 0.32 (0.32) and 0.28 (0.29) in control and HNT conditions, respectively. Similar to the genomic heritability estimates obtained, although most metabolites showed an agreement of genomic prediction accuracy between control and HNT conditions, there were some metabolites that deviated from this trend (Figure 6.4C). A total of 24 and 20 metabolites, respectively, showed small (< 0.05) and large (> 0.1) differences in genomic prediction accuracy between control and HNT conditions.

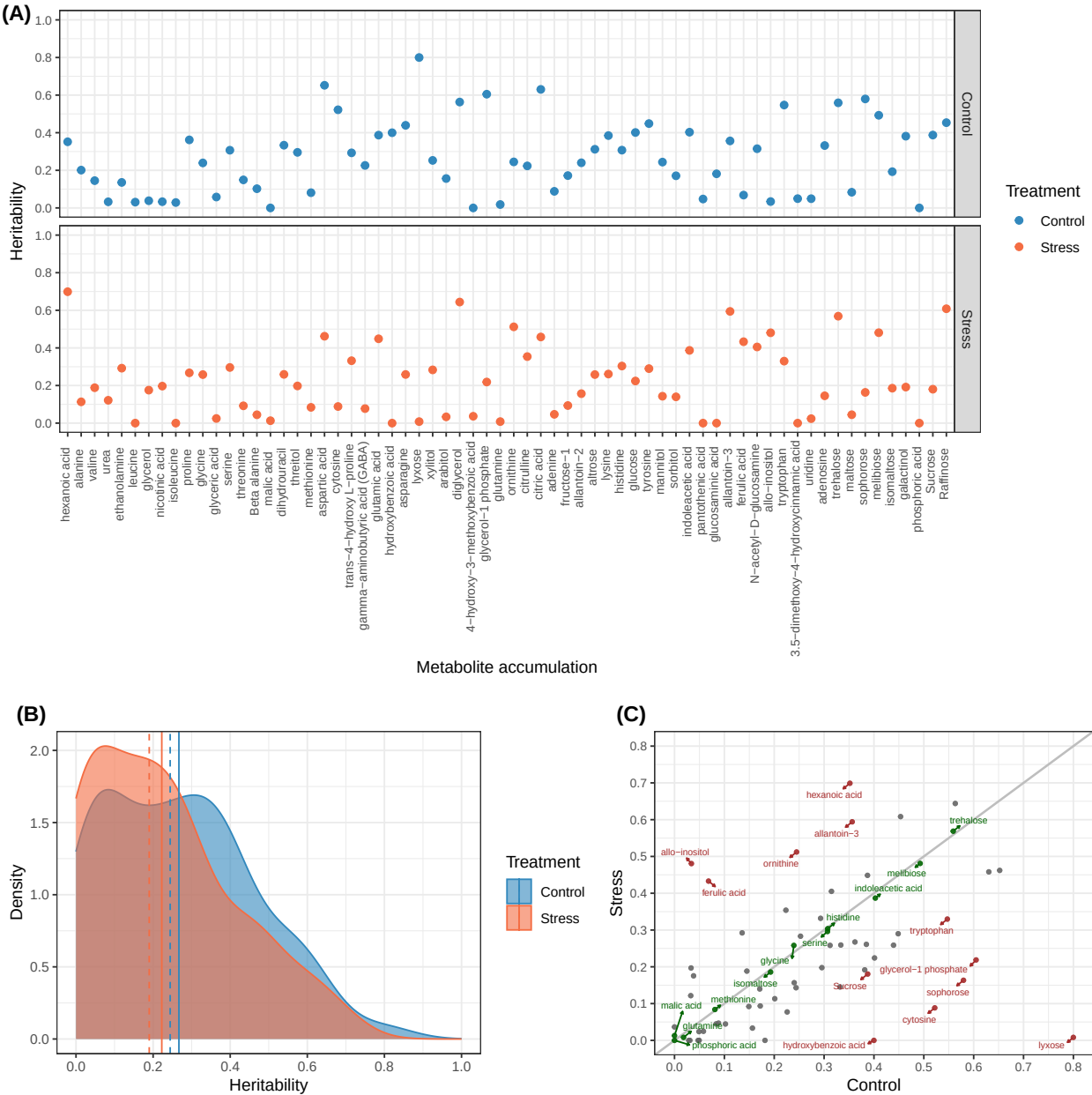


Figure 6.2: Genomic heritability estimates of metabolite accumulation in control and high night temperature stress conditions. A) Scatter plot. B) Density plot. Solid and dashed lines indicate mean and median, respectively. C) Agreement of heritability estimates between control and high night temperature stress conditions. Metabolite accumulation in green and red colors indicate that the heritability difference between control and high night temperature stress conditions was small (< 0.05) and large (> 0.1).

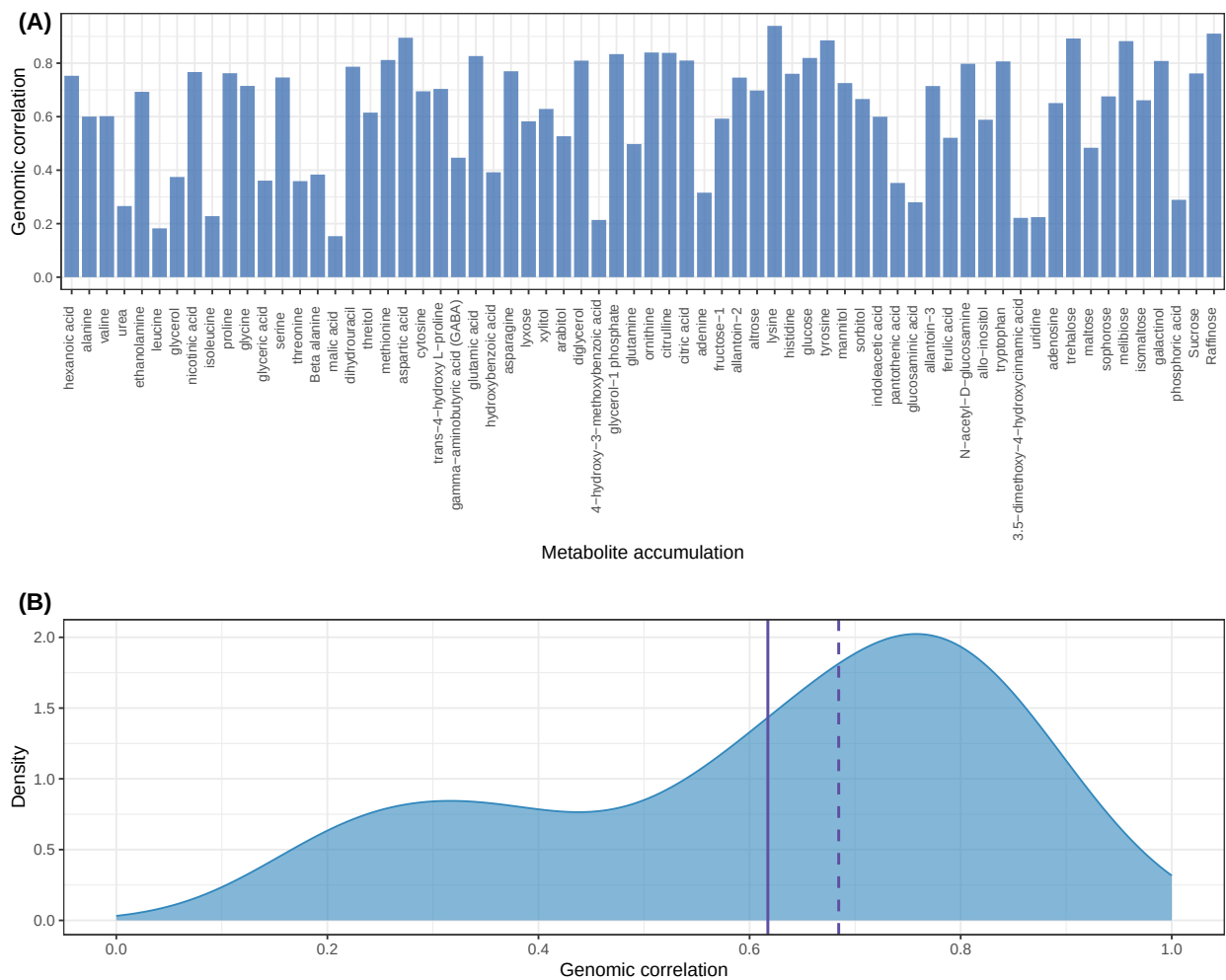


Figure 6.3: Genomic correlation estimates between the same metabolite accumulation measured under control and high night temperature stress conditions. A) Scatter plot. B) Bar chart. Solid and dashed lines indicate mean and median, respectively.

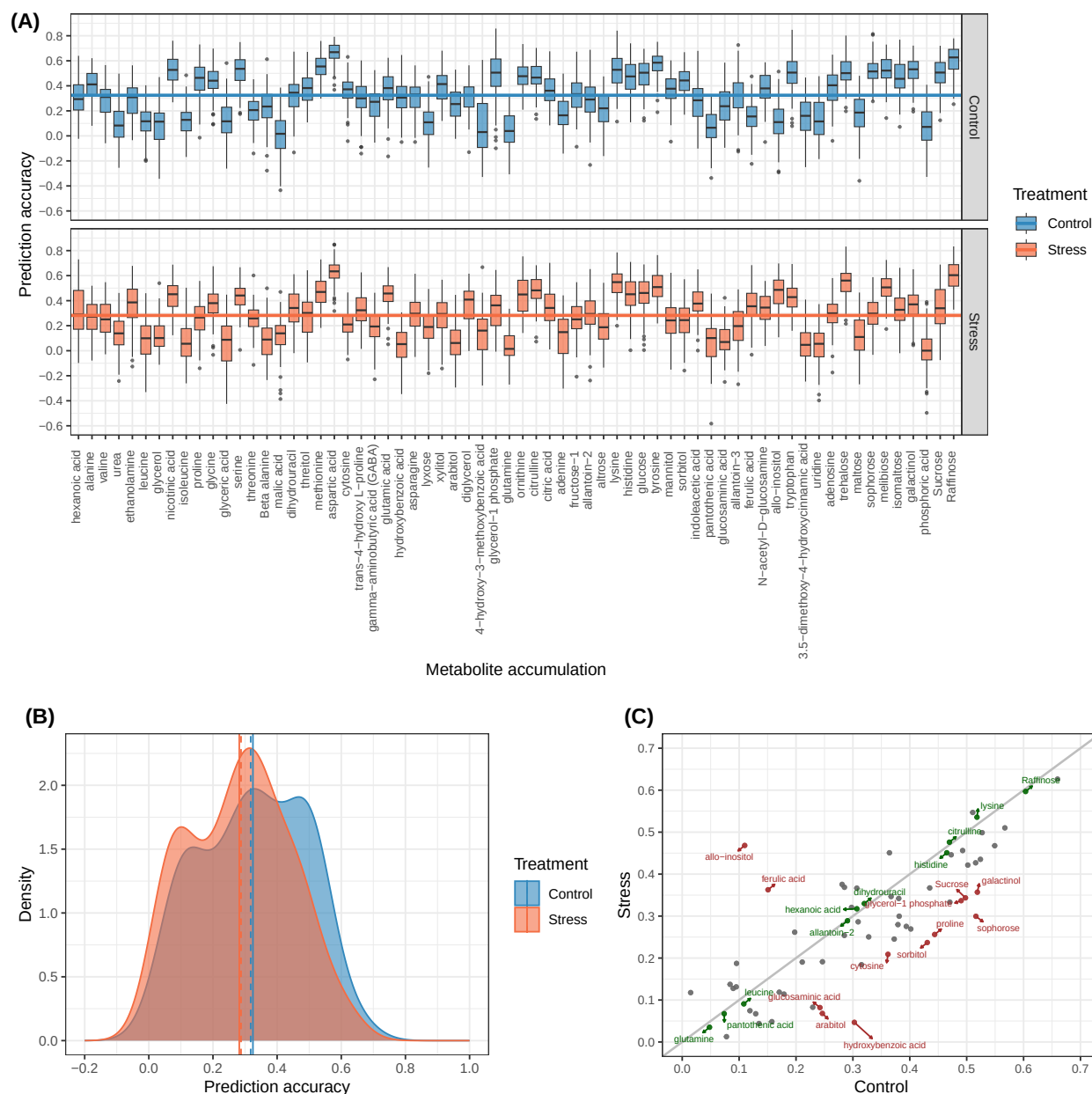


Figure 6.4: Genomic prediction accuracy of metabolite accumulation in control and high night temperature stress conditions. A) Box plot. The horizontal line indicates the mean value. B) Density plot. The solid and dashed lines indicate the mean and median, respectively. C) Agreement of genomic prediction accuracy between control and high night temperature stress conditions. Metabolite accumulation in green and red colors indicate that the genomic prediction difference between control and high night temperature stress conditions was small (< 0.05) and large (> 0.1).

There was also a linear relationship between genomic heritability estimates and genomic prediction accuracies. The coefficients of determination obtained from ordinary least squares between genomic heritability estimates and genomic prediction accuracies squared were 0.89 and 0.88 for the control and HNT conditions, respectively.

6.4.2 Latent factors behind metabolites

Parallel analysis was used to find the optimal number of latent factors. In both the control and HNT conditions, the first five eigenvalues from the real data were larger than the first five eigenvalues from the simulated random data. Therefore, five latent factors were investigated in the next step. Exploratory factor analysis was performed to understand the biological significance of the five latent factors by examining the covariation among the measured metabolite accumulation. The magnitude of the contributions of the latent factors to metabolite accumulation in the control and HNT conditions is summarized in Figure D.2. Because explanatory factor analysis allows for cross-loading of metabolites, the largest loading for each metabolite accumulation was selected based on a threshold of $> |0.4|$. This resulted in each metabolite loading on only one factor, except that urea, glycine, methionine, arabitol, diglycerol, citric acid, allo-inositol, trehalose, maltose, and sophorose, and hexanoic acid, urea, threitol, lyxose, arabitol, diglycerol, and citric acid did not load on any factors in the control and HNT conditions, respectively.

Density plots of genomic heritability, genomic correlation, and genomic prediction accuracy categorized into 10 latent factors are shown in Figure D.3. Here, the first five (F1-F5) and last five (F6-F10) latent factors denote factors controlling the control and HNT conditions, respectively. In the control condition, F2 (0.43) and F4 (0.30) had higher mean estimates of genomic heritability, and F2 (0.51), F5 (0.39), and F4 (0.27) had higher means of genomic

prediction accuracy. In the HNT condition, F10 (0.33), F7 (0.32), and F8 (0.30) had higher mean estimates of genomic heritability, and F7 (0.45), F8 (0.28), and F10 (0.27) had higher means of genomic prediction accuracy.

6.4.3 Single-trait analysis using non-additive models

The gain in prediction accuracy provided by single-trait RKHS or DL-based on different architectures relative to single-trait GBLUP is shown in Figure 6.5. The inclusion of epistasis increased the prediction accuracy of some metabolite accumulation in RKHS. For example, a gain in prediction was observed for allo-inositol, glyceric acid, uridine, trans-4-hydroxy L-proline, and allantoin-2 in the control conditions as well as for urea, hexanoic acid, glycerol, maltose, allantoin-2, and N-acetyl-D-glucosamine in the HNT condition. Capturing epistasis using RKHS resulted in an average gain of 0.34% and -2.78% for the control and HNT conditions, respectively. Overall, the use of DL by converting SNP into image data increased prediction accuracy more than RKHS. An example of a set of SNP transformed into image data for a randomly selected genotype is shown in Figure 6.6. EfficientNetB7, ResNet50, and VGG16 showed noticeable gains in the control conditions. In particular, in the control conditions, VGG16, EfficientNetB7, and ResNet50 improved prediction accuracy in 30, 26, and 8 metabolite contents, respectively. In the HNT conditions, VGG16 and MobileNetV2 showed a gain in prediction accuracy in 23 and 13 metabolite contents, respectively.

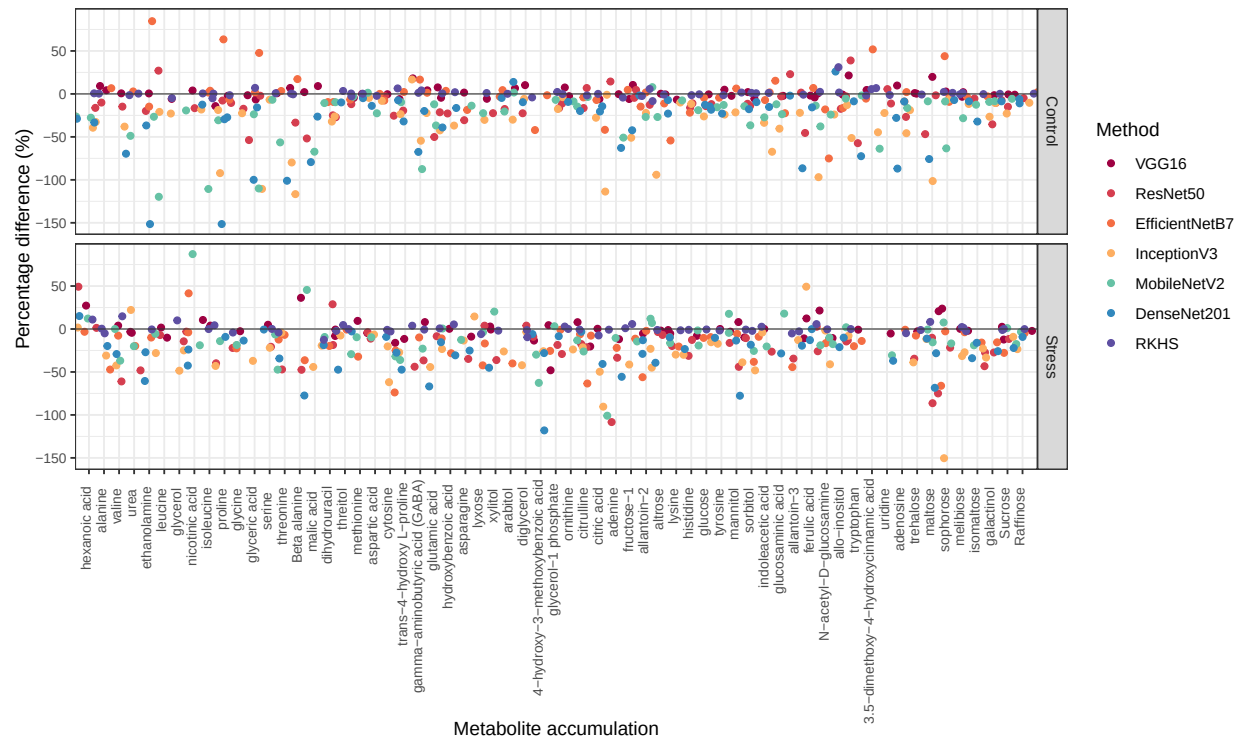


Figure 6.5: Percentage difference of gain in prediction accuracy for single-trait reproducing kernel Hilbert spaces regression (RKHS), VGG16, ResNet50, EfficientNetB7, InceptionV3, MobileNetV2, and DenseNet201 relative to single-trait genomic best linear unbiased prediction.

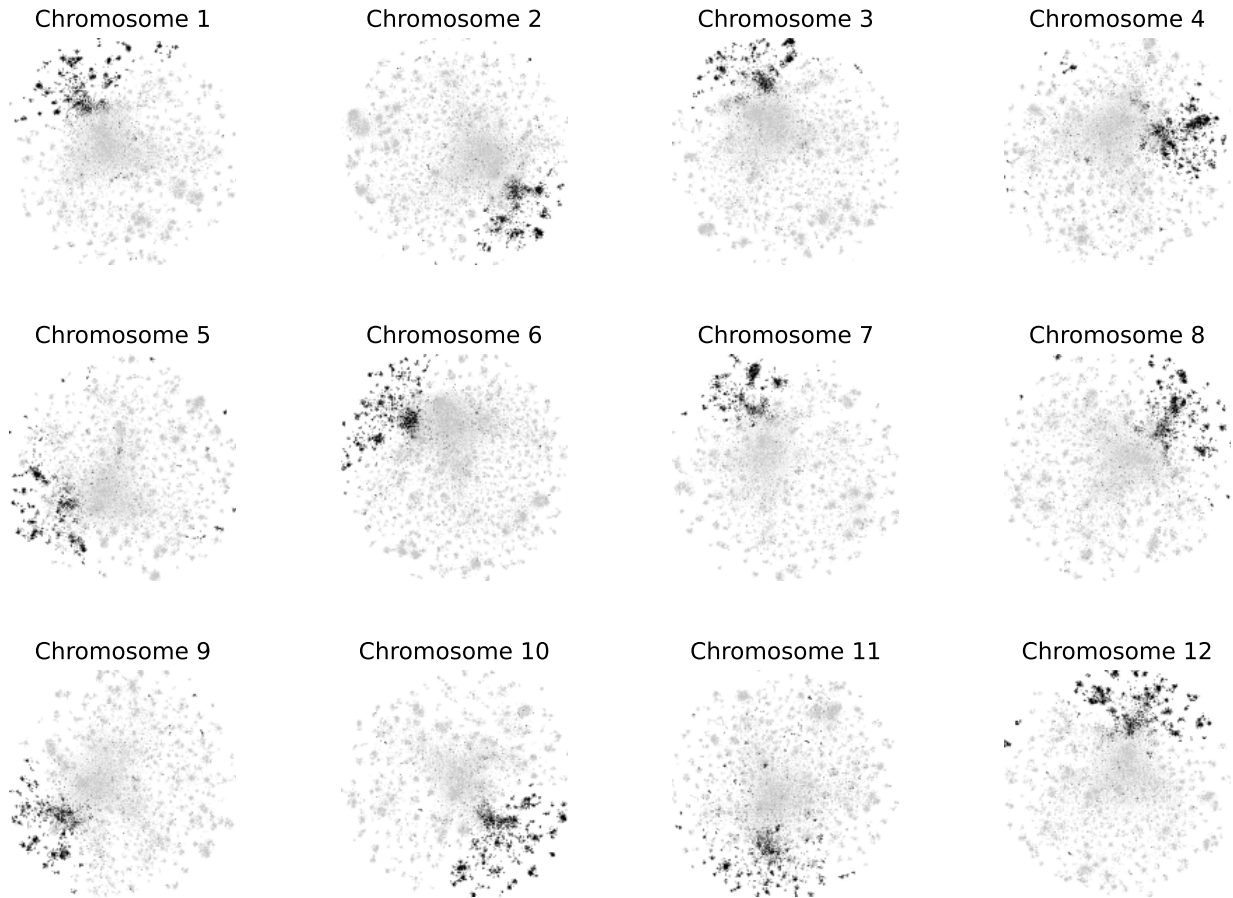


Figure 6.6: Example of a set of single nucleotide polymorphisms transformed into image data for a randomly selected genotype. Images of 12 chromosomes were processed in the multi-channel convolutional neural networks.

6.4.4 Joint analysis of multiple metabolites

Simultaneous regression modeling of all metabolite abundance values was performed using MegaLMM. Estimates of genomic correlations between different metabolite abundance values are shown in Figure D.4. Estimates ranged from -0.28 to 0.79 and -0.20 to 0.74 for the control and HNT conditions, respectively. Similar to the results for genomic heritability and single-trait genomic prediction, genomic correlation estimates were larger in the control than in the HNT conditions (Figure 6.7). Their mean (median) values were 0.14 (0.17) and 0.12 (0.16)

in the control and HNT conditions, respectively.

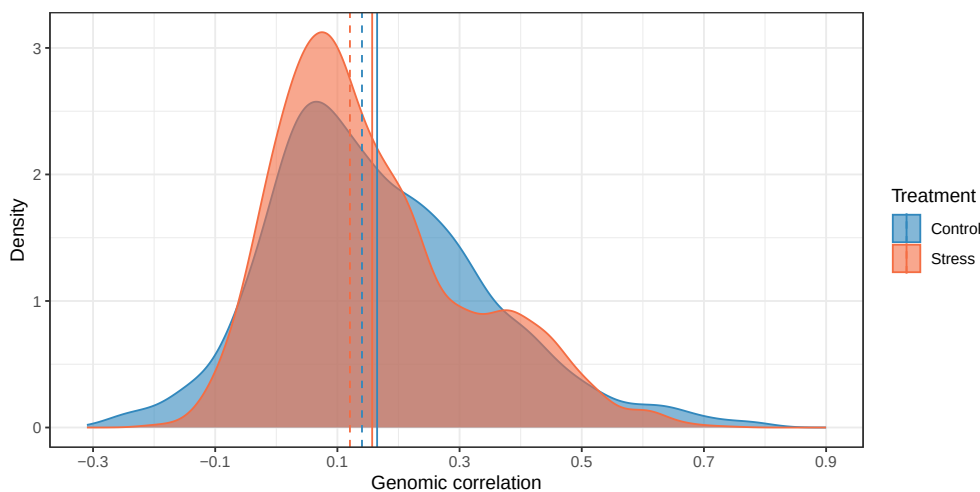


Figure 6.7: Genomic correlation estimates between different metabolite accumulation in control and high night temperature stress conditions. The solid and dashed lines indicate mean and median, respectively.

The gain in prediction accuracy provided by multi-trait MegaLMM-G and multi-trait MegaLMM-GK relative to single-trait GBLUP is shown in Figure 6.8. Using the correlation between metabolites through simultaneous fitting increased the genomic prediction accuracy of some metabolites, including threonine, gamma-aminobutyric acid, allo-inositol, uridine, and maltose in the control condition as well as proline, malic acid, xylitol, 4-hydroxy-3-methoxybenzoic acid, and sucrose in the HNT condition. When the total metabolites were fitted together with **G** using MegaLMM-G, the average gain in prediction accuracy was +0.48% and -0.45% for the control and HNT conditions, respectively (Figure 6.8B). Finally, combining MegaLMM with **GK** (MegaLMM-GK) resulted in the largest average gain of 4.80% and 1.79% for the control and HNT conditions, respectively (Figure 6.8C). Overall, the use of MegaLMM was not successful for all metabolites, as some metabolites showed lower prediction accuracies than those of GBLUP.

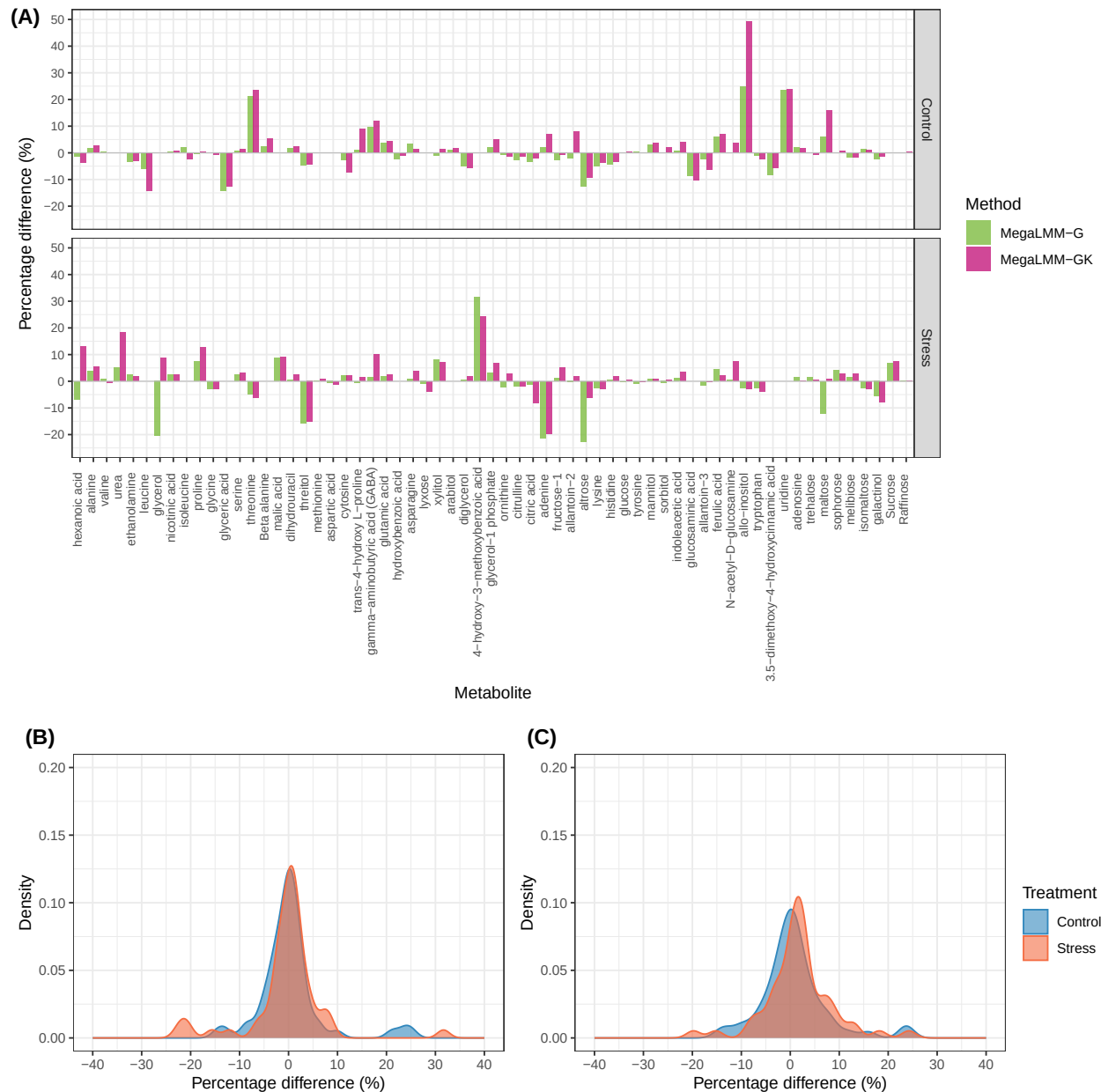


Figure 6.8: Percentage difference of gain in prediction accuracy for multi-trait genomic best linear unbiased prediction (MegaLMM-G) and multi-trait reproducing kernel Hilbert spaces regression (MegaLMM-GK) relative to single-trait genomic best linear unbiased prediction (A). Density plots of percentage difference are shown for MegaLMM-G (B) and MegaLMM-GK (C).

6.5 Discussion

In rice, HNT stress negatively affects grain yield and quality. It has been reported that an increase in 1°C average minimum temperature during the growing season results in a 10% yield loss [133]. Therefore, there is a need to better understand the response of rice to HNT stress and to improve the resilience of rice to HNT stress. Studies of metabolites as phenotypes using a quantitative genetic approach via whole-genome data have recently increased in plants [13, 14, 47, 129]. Although metabolic profiles of plants can be used to measure their HNT responses [126], the use of whole-genome regression analysis to analyze metabolite accumulation has been limited in rice. In this study, we 1) quantified the extent of genetic control of variation in metabolite levels and 2) evaluated the accuracy of genomic prediction of metabolite accumulation under control and HNT conditions.

Overall, our study revealed low to moderate genomic heritability estimates for the metabolic profiles of rice accessions, consistent with previous findings of higher broad-sense heritability for secondary metabolites compared to primary metabolites in rice grains as assessed using line replications [106]. Our results show that the observed pattern of low to moderate heritability extends to genome-based heritability estimates. Although a genome-wide association study was beyond the scope of this study, previous research suggests that secondary metabolites tend to be controlled by a limited number of loci with large effects, whereas primary metabolites tend to be controlled by loci with smaller effects [18].

The accuracy of genomic prediction results provided by GBLUP for metabolite accumulation was consistent with the upper bound established by genomic heritability estimates. This observation was supported by a linear relationship between genomic heritability estimates and squared genomic prediction accuracies. The control conditions yielded higher genomic heritability estimates and genomic prediction accuracies compared to the HNT conditions,

suggesting that accurate estimation of SNP effects was more difficult under the HNT condition or the presence of differential genetic control of metabolite accumulation between the two treatments. Several metabolites showed notable agreement or difference in genomic heritability estimates or genomic prediction accuracies between the control and HNT conditions. For example, allo-inositol and ferulic acid showed differences between the two conditions in both genomic heritability and genomic prediction analyses. These two metabolites are often related, although they are not related in metabolic pathways. They were more heritable and predictable under HNT than under control conditions. Ornithine and allantoin-3 were also found to differ between the two conditions in genomic heritability estimates related to nitrogen metabolism. Sucrose, proline, sorbitol, and galactinol showed higher prediction accuracy in genomic predictions under the control condition. Overall, amino acids tended to be enriched in both conditions.

The degree of genotype by environment interactions was assessed using a bivariate model in which the traits measured in the different environments (i.e., treatments) were treated as distinct but correlated traits. Estimates of genetic correlation between the same trait in different environments obtained from the multi-trait model reflect the degree of re-ranking of genotypes between environments. In general, genetic correlations of less than 0.8 are usually considered to represent different environments [150]. Since many of the genetic correlation estimates of the metabolite were less than 0.8, the extent of genotype by environment interactions in rice grain metabolites is not weak and indicates that re-ranking among accessions occurs between the control and HNT conditions.

To facilitate interpretation of the large number of metabolites analyzed in this study, we used exploratory factor analysis. Exploratory factor analysis is a statistical technique that can reveal the underlying structure among phenotypes when no prior assumptions can be made about the nature of the underlying factor. Factors F3 and F9 contained branched-chain

amino acids, or BCAA, and factors F4 and F8 contained N-rich compounds. Of these, F4 and F8 were identified as having higher mean estimates of genomic heritability and mean accuracies of genomic prediction in the control and HNT conditions, respectively, in single-trait metabolite analysis.

The comparison between GBLUP and RKHS or DL suggests that the prediction accuracy of metabolite accumulation can be improved by capturing non-additive genetic effects for some, but not all, metabolites. In particular, the prediction gain provided by DL was remarkable for some metabolites, suggesting that transforming SNP data into image data is a potentially useful approach for genomic prediction by improving feature and structure extraction from non-image data. Our results are consistent with a recent study that reported encouraging results using the original version of DeepInsight for genomic prediction in maize [38]. The modified DL approach used in the current study follows that of the original version in the image transformation step [163], but differs in the adapted CNN architectures. It takes advantage of the recent development of CNN architectures coupled with multi-channel CNN. Furthermore, our results indicate that subsets of rice grain metabolite contents are controlled by non-linear genetic effects. However, these metabolites were not enriched in specific metabolic categories. Overall, the use of RKHS or DL was not useful for all metabolites because their prediction accuracies were lower than those of GBLUP.

Using MegaLMM allowed us to fit all metabolite accumulation simultaneously, taking advantage of the correlations between them. Otherwise, assuming a bivariate analysis was performed, $(66 \times 65) / 2 = 2,145$ pairs of models had to be fitted, which is computationally demanding. Joint quantitative genetic analysis of multiple metabolite accumulation, including genomic correlation and genomic prediction analyses, was provided by MegaLMM. Consistent with the single-trait metabolite analysis, the control conditions yielded higher genomic correlation estimates between metabolites and greater genomic prediction accura-

cies compared to the HNT conditions when MegaLMM was used. This further supports the hypothesis that accurate estimation of SNP effects was more difficult in the HNT condition or that there was differential genetic control of metabolites between the two treatments. Across all metabolites, MegaLMM-GK provided the best prediction on average compared to single-trait GBLUP. This result suggests that incorporating the correlation between metabolite accumulation and epistasis was an important factor in improving genomic predictions of metabolite accumulation.

6.6 Conclusions

Primary metabolite accumulation from rice grains was low to moderately heritable. Genomic heritability estimates were slightly higher in control than in HNT. Genomic prediction accuracies of metabolite accumulation obtained from GBLUP were within the expected upper limit set by their genomic heritability estimates. Genomic correlation estimates for the same metabolite between the control and HNT conditions deviated from 1, indicating the presence of genotype by environment interactions. There was variation among metabolites in how well they agreed between the control and HNT conditions in terms of genomic heritability estimates and genomic prediction accuracies. Genomic prediction models such as RKHS and image-based DL for single-trait metabolite analysis were effective in improving prediction accuracy by capturing non-additive genetic effects for some, but not all, metabolites. Joint analysis of many metabolites via MegaLMM was also shown to be useful for improving genomic prediction accuracy by exploiting correlations among metabolite accumulation. The current study serves as a first important step to evaluate the genomic prediction accuracy of metabolites under control and HNT conditions.

6.7 Data availability

Genotypic data regarding the rice accessions are available at the rice diversity panel website (<http://www.ricediversity.org/>). Scripts used in this work are publicly available in GitHub (<https://github.com/yebigithub/GBLUP4Met>).

Chapter 7

Conclusion

The research presented in this dissertation highlights the transformative potential of advanced computational techniques to address complex challenges in livestock management and crop resilience. By investigating the application of deep learning and machine learning to digital image analysis and the integration of multi-omic data for genomic prediction, this work contributes to the development of novel approaches for computer vision-based livestock monitoring and quantitative genetic modeling of rice. The results demonstrate the potential of these approaches to achieve more sustainable and resilient farming systems.

In Chapter 3, the investigation of the use of depth video data for predicting dairy cow body weight highlighted the value of longitudinal monitoring in capturing the dynamic nature of weight changes. This study found that, among four image-derived features, volume followed by width, correlated best with scale-based body weight. The comparison between thresholding and Mask R-CNN segmentation algorithms revealed that deep learning-based segmentation techniques can substantially improve prediction accuracy. The combination of Mask R-CNN and ordinary least squares or Mask R-CNN and linear mixed model provided the best predictive performance when the model did not include cow variability or when the model included cow variability, respectively. This image-based method has the potential to reduce reliance on more labor-intensive and stress-inducing traditional methods.

Chapter 4 applied deep learning methods to commercial pig data to address the challenges of industry-scale body weight prediction. Using efficient convolutional neural networks and

vision transformers, the study showed that these neural networks can automatically extract essential, untargeted features from depth images, enabling one-step body weight prediction. Vision transformer-based models consistently delivered superior prediction results. This research demonstrated the feasibility of accurate, large-scale predictions in a commercial environment. The success of these models in improving prediction accuracy demonstrates the potential for integrating advanced computer vision techniques into routine farm operations to improve growth monitoring and phenotyping with minimal disruption to the animals.

The limitation of body weight prediction models is their focus on single farms and a limited number of animal breeds. To enhance model robustness in complex real-farm situations, it is essential to extend these models to encompass a broader range of farms and breeds. The application of these models in farming must also address societal aspects, such as ethical implications and prioritize data privacy to safeguard sensitive information and build trust among stakeholders.

The quantitative genetic analysis of rice presented in Chapters 5 and 6 addressed the pressing challenge of climate change's impact on grain size, focusing on the detrimental effects of elevated nighttime temperatures (HNT). In Chapter 5, the study demonstrated that metabolic profiles could accurately classify rice genotypes under control and HNT conditions. While metabolic predictions for grain size phenotypes were moderate, grain width emerged as the most reliably predicted trait. Genomic prediction outperformed metabolic prediction, and integration of metabolites and genomics into a single model provided a slight improvement. Importantly, certain metabolites were identified as potential candidates for improving multi-trait genomic prediction, providing insight into their role in HNT tolerance.

In Chapter 6, the study further investigated the genomic heritability and genomic prediction of metabolite accumulation in rice grains under control and HNT conditions. The study found that genomic heritability estimates were low to moderately heritable and genomic

prediction accuracies were consistent with expected genomic heritability estimates. The observed genomic correlations between control and HNT conditions deviated from 1, indicating genotype-by-environment interactions. The effectiveness of genomic prediction models, such as reproducing kernel Hilbert spaces regression and image-based deep learning, varied across metabolites, with some benefiting from the ability to capture non-additive genetic effects. Additionally, joint analysis of multiple metabolites via MegaLMM improved prediction accuracy by leveraging correlations between metabolite accumulation.

However, a main limitation of these crop studies is that the rice plants were grown under greenhouse conditions designed to simulate high nighttime temperatures. While greenhouse environments provide controlled settings for experiments, they may not fully replicate the complexity and variability of field environmental conditions, such as abiotic stress, soil type, moisture levels, and pest pressure. Consequently, findings from these studies require further validation in diverse field environments to ensure their applicability and reliability for future rice breeding programs.

Collectively, the research chapters in this dissertation contribute to a growing body of knowledge that harnesses the power of machine learning and deep learning in agriculture. The research not only advances the technical capabilities of phenotyping and genomic prediction but also highlights the importance of integrating multiple types of data to address the multifaceted challenges of modern agriculture. The methods and findings presented here have broad implications for the future of precision livestock farming and crop breeding programs. The continued refinement and application of these techniques will be essential in developing more resilient agricultural systems capable of withstanding environmental pressures while meeting the demands of a growing global population. As digital technologies and computational models evolve, their integration into agricultural practices will play a pivotal role in shaping the future of food production, ensuring both sustainability and productivity in the

years to come.

This dissertation, therefore, serves as a foundation for future research that will further explore and refine the use of advanced computational techniques in agriculture. The work has demonstrated the practicality and potential of these approaches and laid the groundwork for ongoing innovations that will continue to push the boundaries of what is possible in livestock management and quantitative genetic modeling of crop resilience.

Appendices

Appendix A

Depth video data-enabled predictions
of longitudinal dairy cow body weight
using thresholding and Mask R-CNN
algorithms

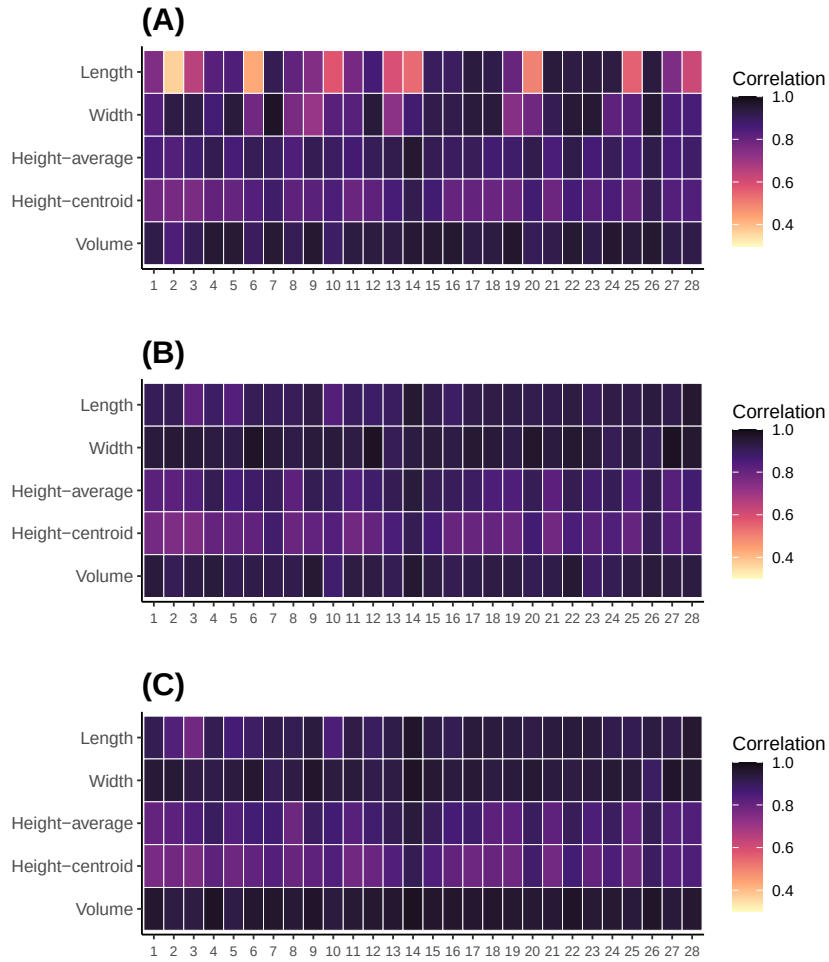


Figure A.1: Per day Pearson correlation heat map between scaled-based body weight and biometric features (length, width, centroid height, average height, and volume). A) single-thresholding, B) adaptive-thresholding, and C) Mask R-CNN.

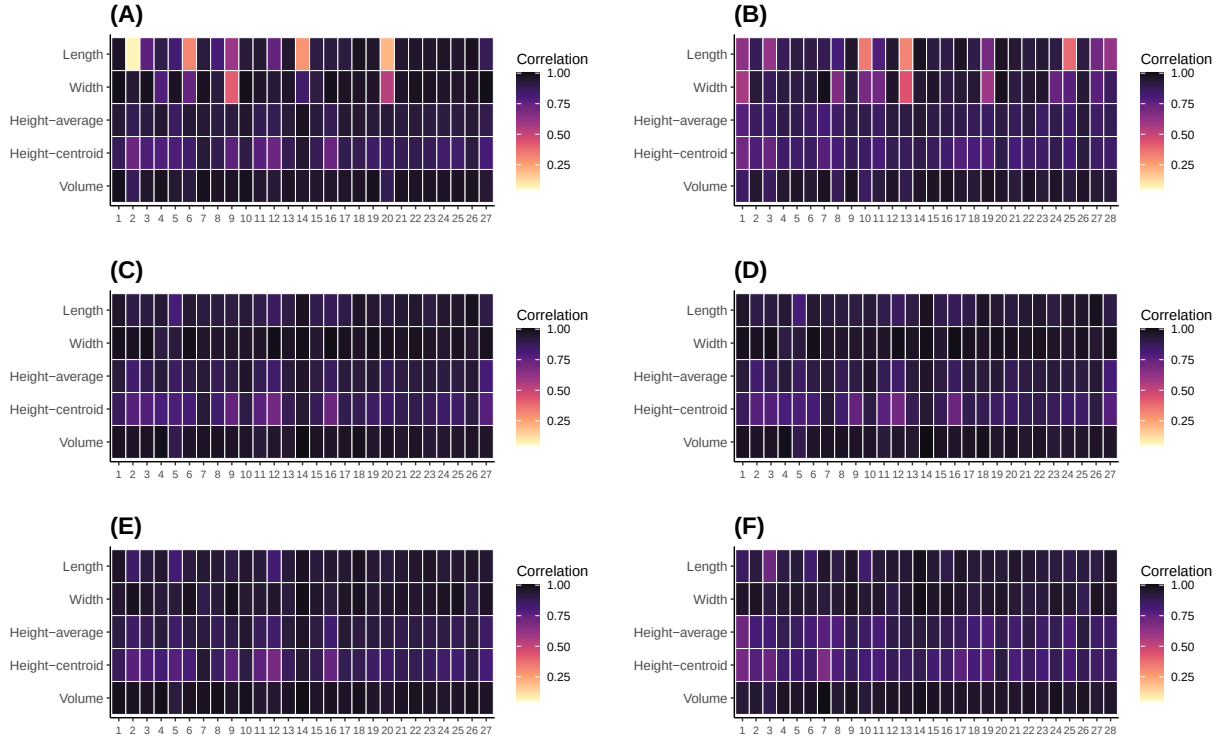


Figure A.2: Per day per AM/PM Pearson correlation heat map between scaled-based body weight and biometric features (length, width, centroid height, average height, and volume). A) AM data using single-thresholding, B) PM data using single-thresholding C) AM data using adaptive-thresholding, D) PM data using adaptive-thresholding, E) AM data using Mask R-CNN, and F) PM data using Mask R-CNN.

Appendix B

Industry-scale prediction of video-derived pig body weight using efficient convolutional neural networks and vision transformers

Table B.1: Single time point cross-validation results for deep learning models.

Model	Visit	RMSE (kg) ¹	R ²²	MAPE ³	MAE (kg) ⁴	Image Count
MobileViT_XXS	T1	4.59	0.76	4.71%	3.03	9,872
MobileViT_S	T1	4.79	0.73	4.95%	3.23	9,872
MobileViT_V2_100	T1	4.74	0.74	4.97%	3.20	9,872
MobileViT_V2_200	T1	4.76	0.74	5.17%	3.36	9,872
MobileViT_V2_150	T1	4.80	0.73	5.29%	3.39	9,872
ResNet50	T1	4.85	0.73	5.43%	3.50	9,872
MobileNetV3Large100	T1	4.99	0.71	5.59%	3.60	9,872
MobileViT_V2_050	T1	5.17	0.69	5.89%	3.80	9,872
MobileNet100	T1	5.16	0.69	5.91%	3.78	9,872

Table B.1: Single time point cross-validation results for deep learning models.

Model	Visit	RMSE (kg) ¹	R ²²	MAPE ³	MAE (kg) ⁴	Image Count
MobileNetV3Large075	T1	5.09	0.70	5.95%	3.83	9,872
MobileNet075	T1	5.15	0.69	6.01%	3.85	9,872
MobileNet050	T1	5.23	0.68	6.03%	3.85	9,872
MobileNetV3Small050	T1	5.60	0.64	6.63%	4.24	9,872
MobileViT_S	T2	3.81	0.84	3.80%	2.91	11,286
MobileViT_V2_200	T2	3.80	0.85	3.80%	2.91	11,286
MobileNet100	T2	3.89	0.84	3.84%	2.96	11,286
MobileViT_XXS	T2	3.99	0.83	3.86%	2.98	11,286
MobileViT_V2_150	T2	3.94	0.83	3.88%	2.98	11,286
MobileViT_V2_050	T2	4.01	0.83	3.96%	3.06	11,286
MobileViT_V2_100	T2	3.96	0.83	4.00%	3.08	11,286
MobileNetV3Large075	T2	4.08	0.82	4.09%	3.16	11,286
MobileNet050	T2	4.11	0.82	4.12%	3.17	11,286
MobileNetV3Large100	T2	4.24	0.81	4.20%	3.22	11,286
ResNet50	T2	4.19	0.81	4.23%	3.23	11,286
MobileNet075	T2	4.28	0.80	4.30%	3.28	11,286
MobileNetV3Small050	T2	4.83	0.75	4.88%	3.77	11,286
MobileViT_S	T3	3.81	0.89	3.08%	2.95	11,248
MobileViT_V2_050	T3	3.85	0.88	3.14%	3.01	11,248
MobileViT_V2_200	T3	3.91	0.88	3.16%	3.01	11,248
MobileViT_V2_100	T3	4.05	0.87	3.17%	3.05	11,248

Table B.1: Single time point cross-validation results for deep learning models.

Model	Visit	RMSE (kg)¹	R²²	MAPE³	MAE (kg)⁴	Image Count
MobileNetV3Large100	T3	4.05	0.87	3.19%	3.07	11,248
MobileViT_XXS	T3	4.29	0.86	3.21%	3.11	11,248
MobileNet075	T3	4.09	0.87	3.24%	3.12	11,248
MobileNetV3Large075	T3	4.14	0.87	3.26%	3.13	11,248
MobileViT_V2_150	T3	4.07	0.87	3.26%	3.15	11,248
ResNet50	T3	4.06	0.87	3.31%	3.17	11,248
MobileNet050	T3	4.21	0.86	3.40%	3.28	11,248
MobileNet100	T3	4.32	0.85	3.43%	3.33	11,248
MobileNetV3Small050	T3	4.46	0.84	3.71%	3.54	11,248
MobileViT_S	T4	8.19	0.56	5.60%	5.79	11,007
MobileViT_V2_150	T4	8.25	0.56	5.65%	5.90	11,007
MobileViT_V2_100	T4	8.21	0.56	5.72%	5.94	11,007
MobileNetV3Large075	T4	8.24	0.56	5.79%	6.08	11,007
MobileViT_V2_050	T4	8.42	0.54	5.84%	6.16	11,007
MobileNet075	T4	8.25	0.56	5.87%	6.10	11,007
ResNet50	T4	8.25	0.56	5.87%	6.08	11,007
MobileNetV3Large100	T4	8.48	0.53	5.95%	6.20	11,007
MobileViT_XXS	T4	8.63	0.52	6.01%	6.27	11,007
MobileNet100	T4	8.56	0.52	6.06%	6.31	11,007
MobileNet050	T4	8.62	0.52	6.11%	6.45	11,007
MobileViT_V2_200	T4	8.82	0.50	6.24%	6.46	11,007

Table B.1: Single time point cross-validation results for deep learning models.

Model	Visit	RMSE (kg) ¹	R ²²	MAPE ³	MAE (kg) ⁴	Image Count
MobileNetV3Small050	T4	9.43	0.42	6.80%	7.17	11,007
MobileViT_S	T5	5.35	0.80	3.42%	3.90	6,570
MobileViT_V2_200	T5	5.39	0.80	3.51%	4.02	6,570
MobileViT_XXS	T5	5.52	0.79	3.59%	4.13	6,570
MobileViT_V2_100	T5	5.45	0.80	3.63%	4.14	6,570
MobileViT_V2_150	T5	5.81	0.77	3.77%	4.24	6,570
MobileViT_V2_050	T5	5.67	0.78	3.77%	4.24	6,570
MobileNet050	T5	5.88	0.76	3.80%	4.33	6,570
MobileNetV3Large075	T5	5.88	0.76	3.85%	4.41	6,570
MobileNet100	T5	6.20	0.74	3.85%	4.41	6,570
MobileNetV3Large100	T5	5.94	0.76	3.87%	4.35	6,570
ResNet50	T5	5.97	0.76	3.92%	4.43	6,570
MobileNet075	T5	6.27	0.73	4.04%	4.60	6,570
MobileNetV3Small050	T5	6.74	0.69	4.34%	4.97	6,570
MobileViT_V2_200	T6	5.77	0.84	3.77%	4.29	3,205
MobileViT_S	T6	5.85	0.84	3.79%	4.34	3,205
MobileViT_XXS	T6	6.08	0.82	3.88%	4.54	3,205
MobileViT_V2_050	T6	5.95	0.83	3.98%	4.60	3,205
MobileViT_V2_150	T6	6.65	0.79	4.21%	4.72	3,205
MobileViT_V2_100	T6	6.42	0.80	4.32%	4.91	3,205
MobileNetV3Large100	T6	6.66	0.79	4.52%	5.18	3,205

Table B.1: Single time point cross-validation results for deep learning models.

Model	Visit	RMSE (kg)¹	R²²	MAPE³	MAE (kg)⁴	Image Count
MobileNet100	T6	7.07	0.73	4.57%	5.86	3,205
ResNet50	T6	7.03	0.73	4.57%	5.84	3,205
MobileNet050	T6	7.12	0.73	4.68%	5.99	3,205
MobileNet075	T6	7.69	0.68	4.94%	6.32	3,205
MobileNetV3Large075	T6	8.38	0.66	5.46%	6.08	3,205
MobileNetV3Small050	T6	8.15	0.68	5.61%	6.30	3,205

¹ Root mean square error

² Coefficients of determination

³ Mean absolute percentage error

⁴ Mean absolute error

Table B.2: Time series forecasting cross-validation results for deep learning.

Model	Scenario ¹	CV	Design	RMSE (kg) ²	R ²³	MAPE ⁴	MAE (kg) ⁵
MobileViT_S	A	0.75	Sparse	4.21	0.87	4.27%	3.13
MobileViT_S	A	0.5	Full	4.21	0.86	4.29%	3.21
MobileViT_S	A	0.75	Full	4.30	0.86	4.29%	3.16
ResNet50	A	0.75	Sparse	4.21	0.87	4.38%	3.22
MobileViT_V2_200	A	0.75	Sparse	4.35	0.86	4.42%	3.27
MobileViT_V2_200	A	0.5	Sparse	4.41	0.85	4.49%	3.38
MobileViT_V2_200	A	0.25	Full	4.44	0.84	4.50%	3.34
ResNet50	A	0.75	Full	4.37	0.86	4.57%	3.34
MobileViT_S	A	0.25	Sparse	4.64	0.83	4.58%	3.43
ResNet50	A	0.5	Full	4.46	0.84	4.60%	3.43
MobileViT_S	A	0.25	Full	4.69	0.83	4.72%	3.55
ResNet50	A	0.5	Sparse	4.62	0.83	4.79%	3.61
MobileViT_V2_200	A	0.75	Full	4.70	0.83	4.79%	3.56
MobileViT_V2_200	A	0.5	Full	4.61	0.83	4.80%	3.52
ResNet50	A	0.25	Full	4.86	0.81	4.88%	3.69
ResNet50	A	0.25	Sparse	4.95	0.81	4.95%	3.71
MobileViT_V2_200	A	0.25	Sparse	5.05	0.80	5.01%	3.79
MobileViT_S	A	0.5	Sparse	4.77	0.82	5.04%	3.75
MobileViT_V2_200	B	0.75	Full	4.06	0.90	3.40%	3.13
MobileViT_S	B	0.75	Sparse	3.95	0.90	3.40%	3.10
MobileViT_V2_200	B	0.75	Sparse	4.04	0.90	3.42%	3.10
MobileViT_S	B	0.75	Full	4.06	0.90	3.47%	3.17

Table B.2: Time series forecasting cross-validation results for deep learning.

Model	Scenario ¹	CV	Design	RMSE (kg) ²	R ²³	MAPE ⁴	MAE (kg) ⁵
MobileViT_S	B	0.25	Sparse	4.45	0.87	3.48%	3.22
MobileViT_V2_200	B	0.5	Full	4.24	0.88	3.51%	3.23
MobileViT_S	B	0.5	Full	4.20	0.88	3.52%	3.24
MobileViT_S	B	0.5	Sparse	4.16	0.89	3.55%	3.22
MobileViT_S	B	0.25	Full	4.47	0.87	3.63%	3.34
ResNet50	B	0.75	Full	4.22	0.89	3.65%	3.31
MobileViT_V2_200	B	0.5	Sparse	4.38	0.87	3.76%	3.42
ResNet50	B	0.75	Sparse	4.47	0.88	3.82%	3.47
ResNet50	B	0.5	Full	4.62	0.86	3.86%	3.55
ResNet50	B	0.5	Sparse	4.53	0.86	3.88%	3.51
MobileViT_V2_200	B	0.25	Full	4.93	0.84	3.89%	3.61
MobileViT_V2_200	B	0.25	Sparse	5.20	0.83	4.07%	3.79
ResNet50	B	0.25	Full	4.98	0.84	4.07%	3.78
ResNet50	B	0.25	Sparse	5.62	0.80	4.48%	4.22
MobileViT_S	C	0.5	Full	7.27	0.70	5.12%	5.30
MobileViT_S	C	0.5	Sparse	7.25	0.70	5.14%	5.25
MobileViT_V2_200	C	0.75	Sparse	7.63	0.62	5.20%	5.38
MobileViT_V2_200	C	0.5	Sparse	7.50	0.68	5.20%	5.42
MobileViT_S	C	0.75	Sparse	7.69	0.61	5.27%	5.45
MobileViT_V2_200	C	0.5	Full	7.66	0.67	5.33%	5.55
MobileViT_S	C	0.25	Sparse	7.79	0.65	5.45%	5.70
MobileViT_S	C	0.75	Full	7.79	0.60	5.46%	5.63

Table B.2: Time series forecasting cross-validation results for deep learning.

Model	Scenario ¹	CV	Design	RMSE (kg) ²	R ²³	MAPE ⁴	MAE (kg) ⁵
ResNet50	C	0.5	Sparse	7.57	0.68	5.48%	5.64
MobileViT_V2_200	C	0.75	Full	7.96	0.58	5.54%	5.75
MobileViT_S	C	0.25	Full	7.70	0.65	5.63%	5.83
MobileViT_V2_200	C	0.25	Full	7.94	0.63	5.73%	5.95
ResNet50	C	0.75	Full	8.12	0.57	5.84%	6.05
ResNet50	C	0.5	Full	8.17	0.62	6.06%	6.19
ResNet50	C	0.75	Sparse	8.67	0.51	6.38%	6.50
ResNet50	C	0.25	Sparse	8.72	0.56	6.51%	6.68
ResNet50	C	0.25	Full	8.83	0.55	6.58%	6.81
MobileViT_V2_200	C	0.25	Sparse	8.98	0.53	6.61%	6.93
ResNet50	D	0.75	Full	4.59	0.89	3.10%	3.53
MobileViT_S	D	0.75	Sparse	4.68	0.88	3.14%	3.63
MobileViT_V2_200	D	0.75	Sparse	4.58	0.89	3.14%	3.61
MobileViT_S	D	0.75	Full	4.89	0.87	3.40%	3.97
ResNet50	D	0.75	Sparse	5.06	0.87	3.49%	3.98
MobileViT_S	D	0.25	Full	5.35	0.87	3.59%	4.02
MobileViT_V2_200	D	0.75	Full	5.36	0.85	3.83%	4.31
MobileViT_V2_200	D	0.5	Sparse	5.88	0.86	4.03%	4.51
MobileViT_S	D	0.25	Sparse	6.04	0.83	4.06%	4.62
MobileViT_S	D	0.5	Full	5.98	0.85	4.07%	4.57
ResNet50	D	0.5	Full	6.18	0.84	4.33%	4.68
MobileViT_V2_200	D	0.5	Full	6.19	0.84	4.35%	4.80

Table B.2: Time series forecasting cross-validation results for deep learning.

Model	Scenario ¹	CV	Design	RMSE (kg) ²	R ²³	MAPE ⁴	MAE (kg) ⁵
MobileViT_S	D	0.5	Sparse	6.10	0.85	4.39%	4.77
ResNet50	D	0.25	Sparse	6.35	0.82	4.42%	4.87
MobileViT_V2_200	D	0.25	Full	6.37	0.81	4.44%	5.03
ResNet50	D	0.5	Sparse	6.48	0.83	4.72%	4.99
ResNet50	D	0.25	Full	6.58	0.80	4.72%	5.21
MobileViT_V2_200	D	0.25	Sparse	7.18	0.76	4.80%	5.61
MobileViT_S	E	0.75	Full	5.43	0.84	3.24%	3.95
MobileViT_S	E	0.75	Sparse	5.29	0.85	3.28%	4.01
MobileViT_V2_200	E	0.75	Full	5.52	0.84	3.45%	4.18
ResNet50	E	0.75	Full	5.83	0.82	3.60%	4.45
MobileViT_V2_200	E	0.75	Sparse	6.28	0.79	3.93%	4.70
ResNet50	E	0.75	Sparse	6.24	0.79	3.94%	4.69
MobileViT_S	E	0.5	Full	7.68	0.79	4.88%	5.69
MobileViT_S	E	0.5	Sparse	8.01	0.77	5.06%	5.89
MobileViT_V2_200	E	0.5	Full	7.90	0.77	5.29%	6.29
MobileViT_V2_200	E	0.5	Sparse	8.49	0.74	5.55%	6.28
MobileViT_S	E	0.25	Sparse	9.56	0.66	5.78%	6.93
MobileViT_S	E	0.25	Full	9.14	0.69	5.89%	6.86
ResNet50	E	0.5	Sparse	9.11	0.70	5.90%	7.03
ResNet50	E	0.5	Full	8.82	0.72	5.93%	6.79
MobileViT_V2_200	E	0.25	Sparse	9.64	0.66	6.29%	7.15
MobileViT_V2_200	E	0.25	Full	9.83	0.64	6.37%	7.24

Table B.2: Time series forecasting cross-validation results for deep learning.

Model	Scenario ¹	CV	Design	RMSE (kg) ²	R ²³	MAPE ⁴	MAE (kg) ⁵
ResNet50	E	0.25	Full	10.24	0.61	6.65%	7.56
ResNet50	E	0.25	Sparse	12.01	0.47	8.44%	9.49

¹ Scenario 0.25, 0.5, 0.75 represent scenarios for subset ratios 25%:75%, 50%:50%, and 75%:25%. Here, 75%:25% indicates that 75% of the data points from all previous time points were used as the training set to predict 25% of the current data points being tested.

² Root mean square error

³ Coefficients of determination

⁴ Mean absolute percentage error

⁵ Mean absolute error

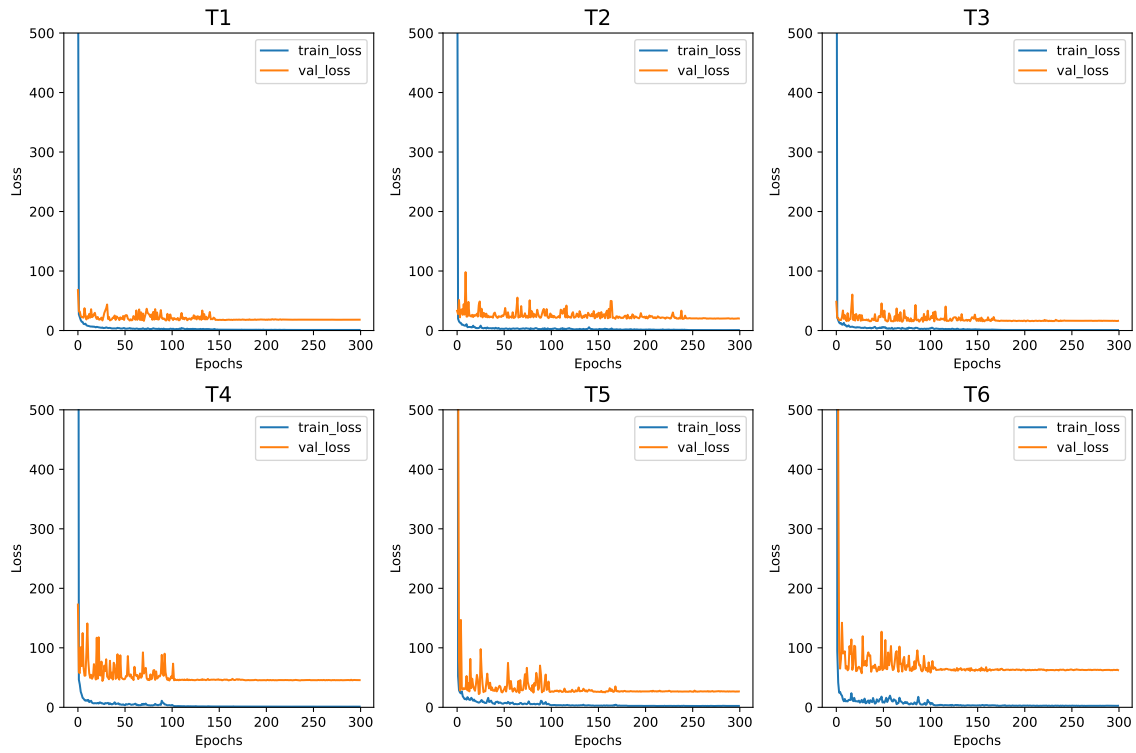


Figure B.1: Single time point cross-validation training history using the best model at each visit. T1 to T6 represent six farm visits. Blue and orange curves represent the mean square error in the training and validation sets, respectively.

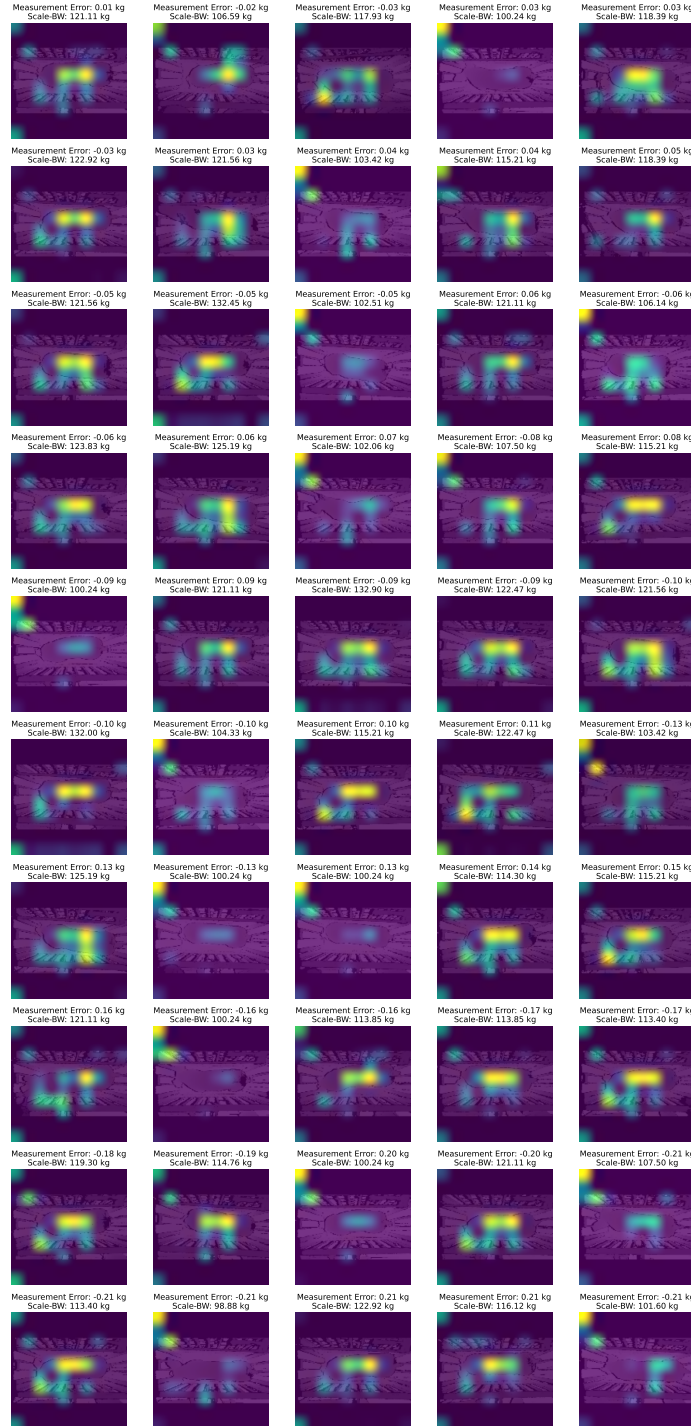


Figure B.2: Gradient-weighted class activation mapping heatmaps for the top 50 prediction results from the T5 visit, generated with MobileViT-S.

Appendix C

Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice

Table C.1: List of metabolites.

Number	Metabolite	Number	Metabolite
1	hexanoic acid	2	alanine
3	valine	4	urea
5	ethanolamine	6	leucine
7	glycerol	8	nicotinic acid
9	isoleucine	10	proline
11	glycine	12	glyceric acid
13	citraconic acid	14	serine
15	threonine	16	Beta alanine
17	malic acid	18	dihydrouracil
19	threitol	20	methionine
21	aspartic acid	22	cytosine
23	trans-4-hydroxy proline	24	gamma-aminobutyric acid (GABA)
25	glutamic acid	26	hydroxybenzoic acid
27	asparagine	28	arabinose
29	lyxose	30	ribose
31	xylitol	32	arabitol
33	ribitol	34	diglycerol
35	4-hydroxy-3-methoxybenzoic acid	36	glycerol-1 phosphate
37	glutamine	38	dihydroxybenzoic acid
39	ornithine	40	citrulline
41	citric acid	42	adenine
43	fructose-1	44	allantoin-2
45	altrose	46	lysine
47	histidine	48	glucose
49	tyrosine	50	mannitol
51	sorbitol	52	indoleacetic acid
53	pantothenic acid	54	glucosaminic acid
55	allantoin-3	56	ferulic acid
57	N-acetyl-D-glucosamine	58	allo-inositol
59	tryptophan	60	3,5-dimethoxy-4-hydroxycinnamic acid
61	uridine	62	eicosapentaenoic acid
63	adenosine	64	trehalose
65	maltose	66	sophorose
67	catechin	68	melibiose
69	isomaltose	70	galactinol
71	phosphoric acid	72	Sucrose
73	Raffinose		

Table C.2: Conventional heritability estimates of grain size phenotypes in control and high night temperature (HNT) stress conditions.

Treatment	Grain length	Grain width	Perimeter
Control	0.71	0.64	0.50
HNT	0.74	0.73	0.54

Table C.3: Predictive correlations of grain-size phenotypes using metabolic, genomic, and multi-omic predictions in control and high night time temperature stress conditions.

Source	Model	Grain length		Grain width		Grain perimeter	
		Control	HNT ⁸	Control	HNT	Control	HNT
Metabolites	MOLS ¹	0.20	0.14	0.26	0.36	0.14	0.10
	MBLUP ²	0.33	0.31	0.54	0.54	0.32	0.27
	BayesC	0.35	0.33	0.52	0.54	0.35	0.29
	RF ³	0.25	0.29	0.57	0.54	0.21	0.26
	SVR ⁴	0.29	0.29	0.52	0.50	0.28	0.24
	XGBoost ⁵	0.27	0.29	0.53	0.47	0.28	0.29
Genetic markers	GBLUP ⁶	0.64	0.64	0.73	0.67	0.57	0.63
Genetic markers & Metabolites	GMBLUP ⁷	0.64	0.63	0.73	0.70	0.58	0.61

¹ Metabolic ordinary least squares

² Metabolic best linear unbiased prediction

³ Random forest

⁴ Support vector regression

⁵ Extreme gradient boosting

⁶ Genomic best linear unbiased prediction

⁷ Genomic metabolic best linear unbiased prediction

⁸ High night time temperature

Table C.4: Predictive correlations of leave-one-subpopulation-group-out cross-validation in control and high night temperature stress conditions.

Testing set ¹	Grain length			Grain width			Grain perimeter		
	M ²	G ³	GM ⁴	M	G	GM	M	G	GM
Control									
tropical-japonica	0.03	0.23	0.22	0.16	0.18	0.29	0.28	0.08	0.54
temperate-japonica	-0.18	0.69	0.38	0.24	0.00	0.10	-0.20	0.62	0.34
indica	0.38	0.68	0.72	-0.09	0.17	0.07	0.36	0.51	0.50
aus	0.21	0.46	0.45	0.29	0.28	0.29	0.32	0.32	0.51
admixed-japonica	0.09	0.31	0.26	0.12	0.39	0.41	0.15	0.04	0.17
High night temperature									
tropical-japonica	0.18	0.31	0.33	0.53	0.43	0.64	0.28	0.27	0.37
temperate-japonica	0.06	0.66	0.53	0.40	0.01	0.03	0.08	0.63	0.46
indica	0.59	0.36	0.40	0.10	0.37	0.14	0.51	0.25	0.45
aus	0.05	0.30	0.29	0.09	-0.26	0.05	0.07	0.33	0.34
admixed-japonica	0.23	0.50	0.57	0.02	0.52	0.57	0.09	0.49	0.45

¹ Aromatic, admixed indica, and admixed were not used as the testing set because the number of genotypes was less than 10

² Metabolic best linear unbiased prediction

³ Genomic best linear unbiased prediction

⁴ Genomic metabolic best linear unbiased prediction

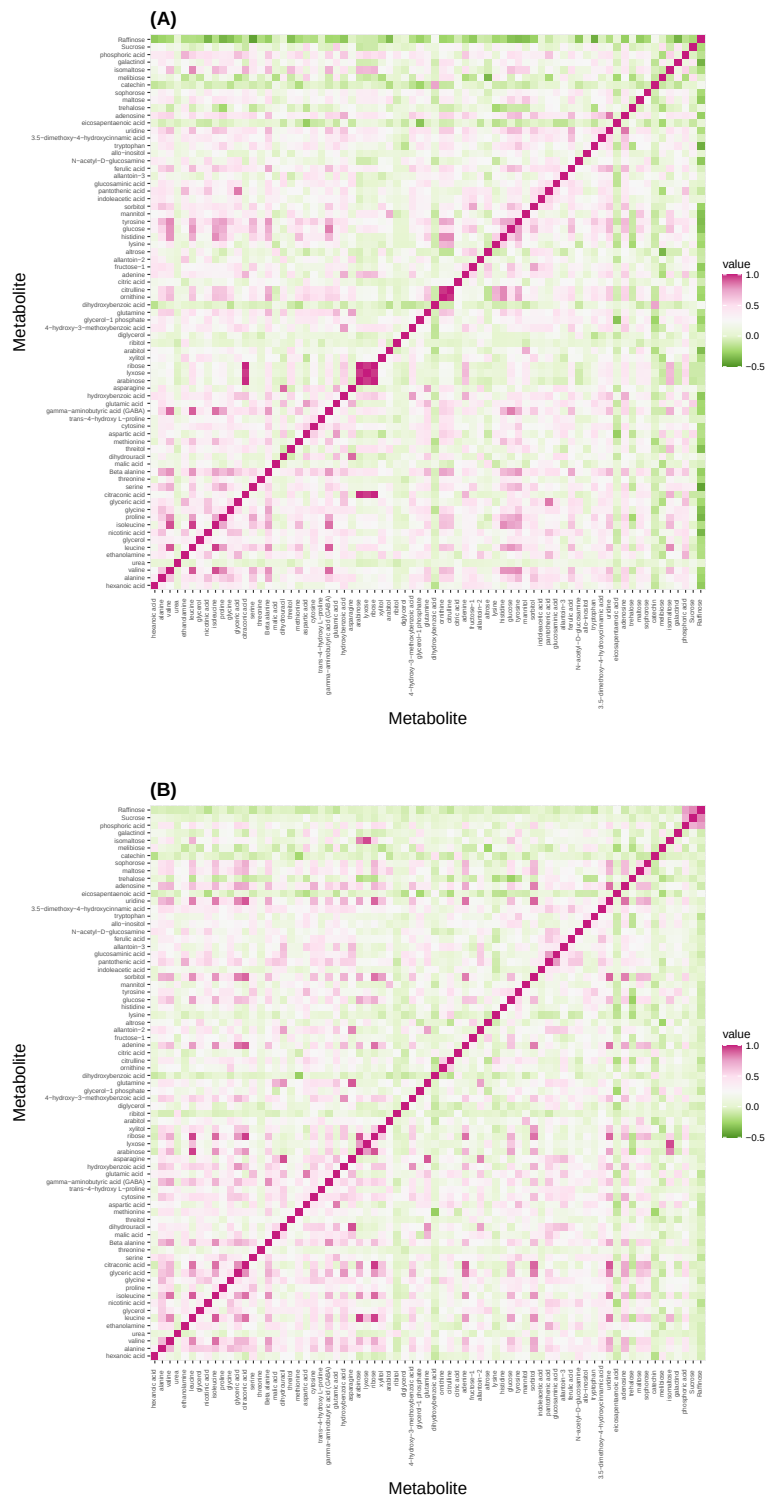


Figure C.1: Pearson correlation heat map among metabolic profiles in control (A) and high night time temperature stress conditions (B).

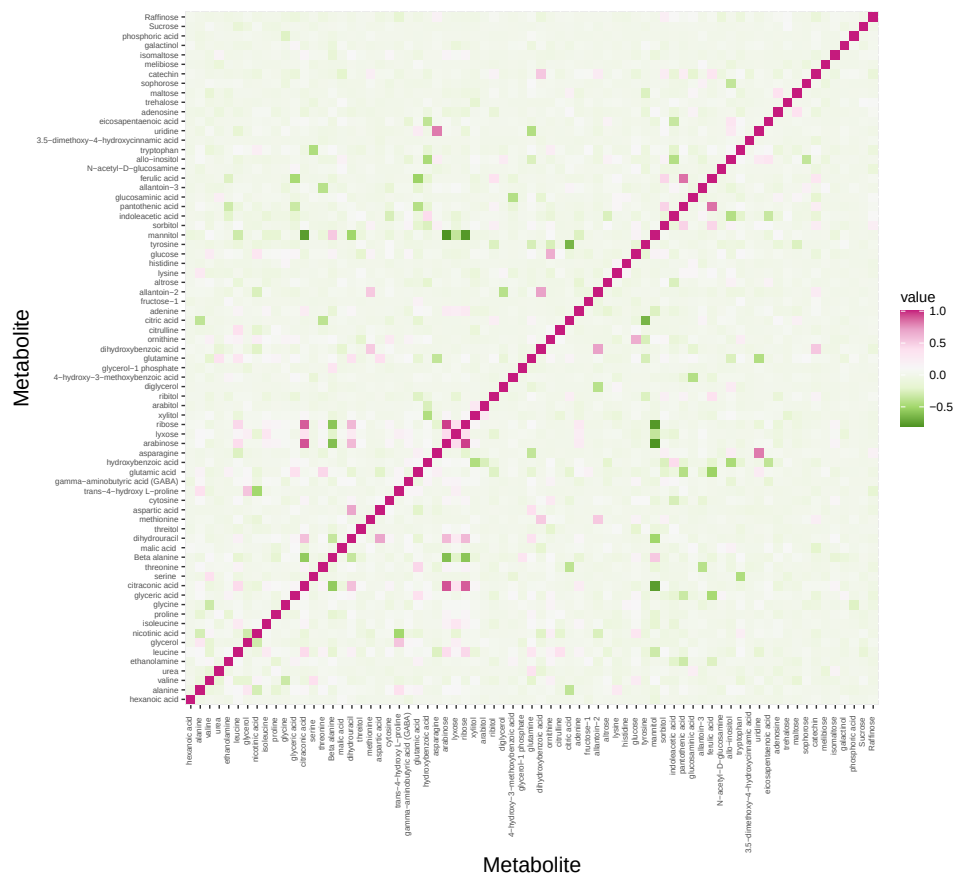


Figure C.2: Pearson correlation heat map among metabolic profiles expressed as the ratio of control and high night time temperature stress.

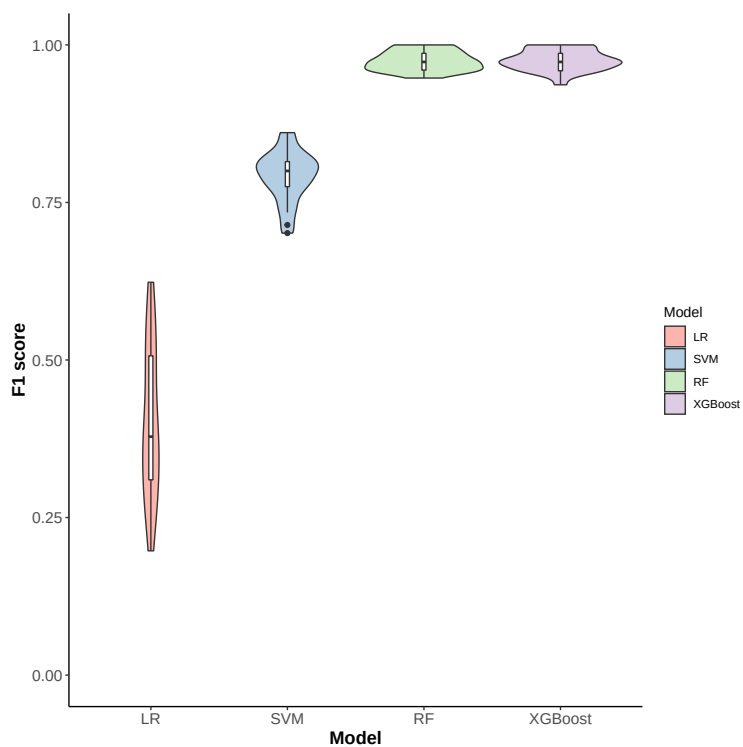


Figure C.3: F1 scores of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; and RF: random forest; XGBoost: extreme gradient boosting.

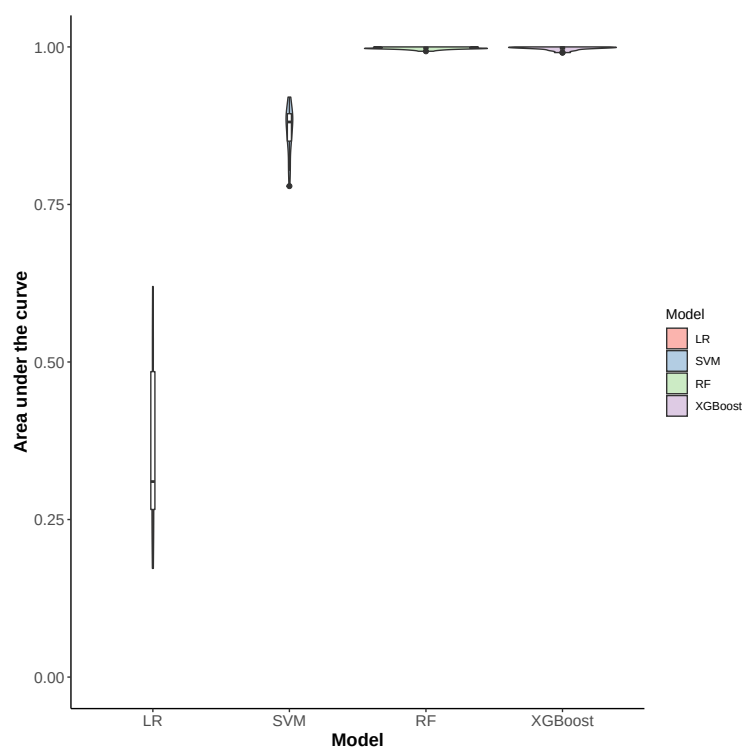


Figure C.4: Area under the curve of high night time temperature conditions (control and stress) using 73 metabolites. LR: logistic regression; SVM: support vector machine; and RF: random forest; XGBoost: extreme gradient boosting.

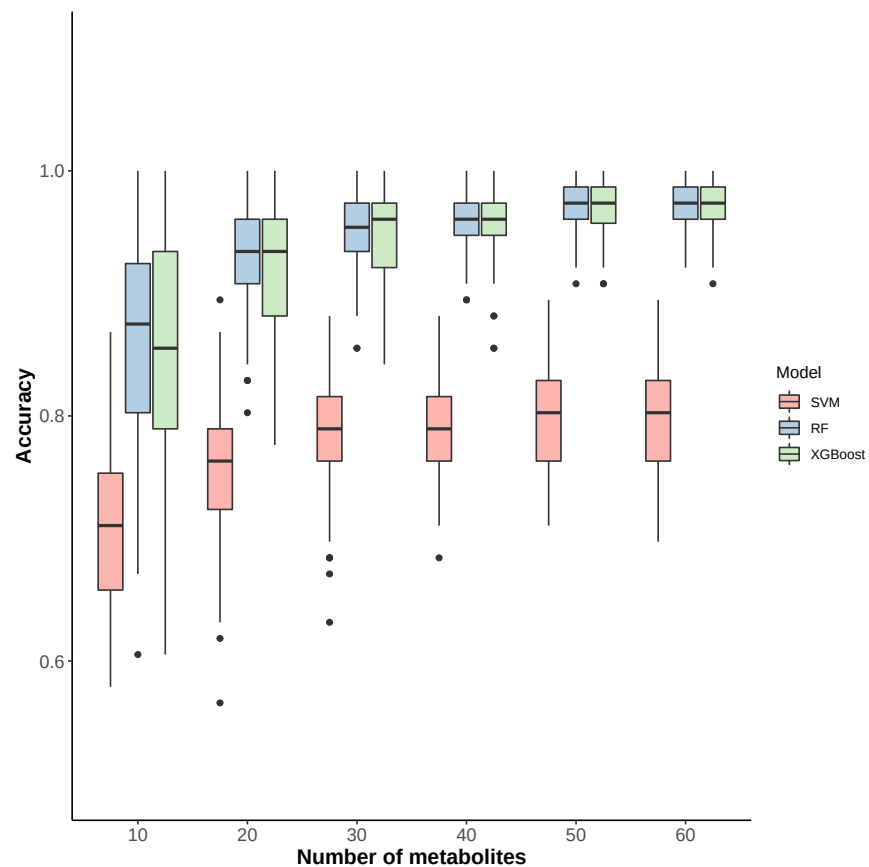


Figure C.5: Classification accuracy of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost).

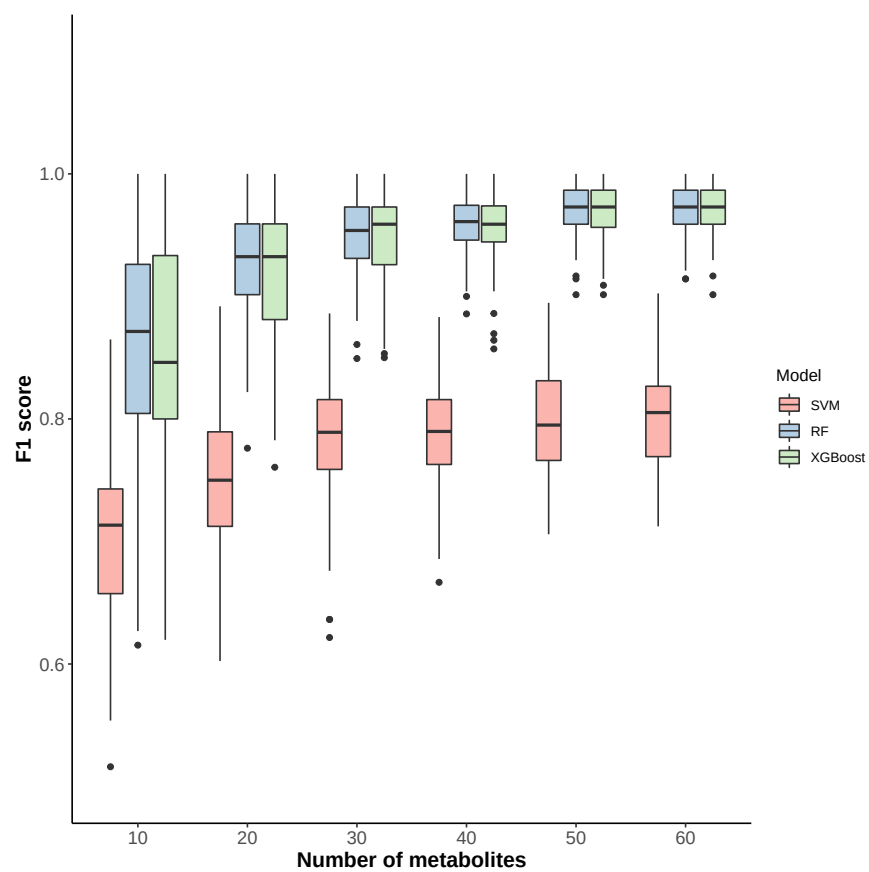


Figure C.6: F1 scores of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost).

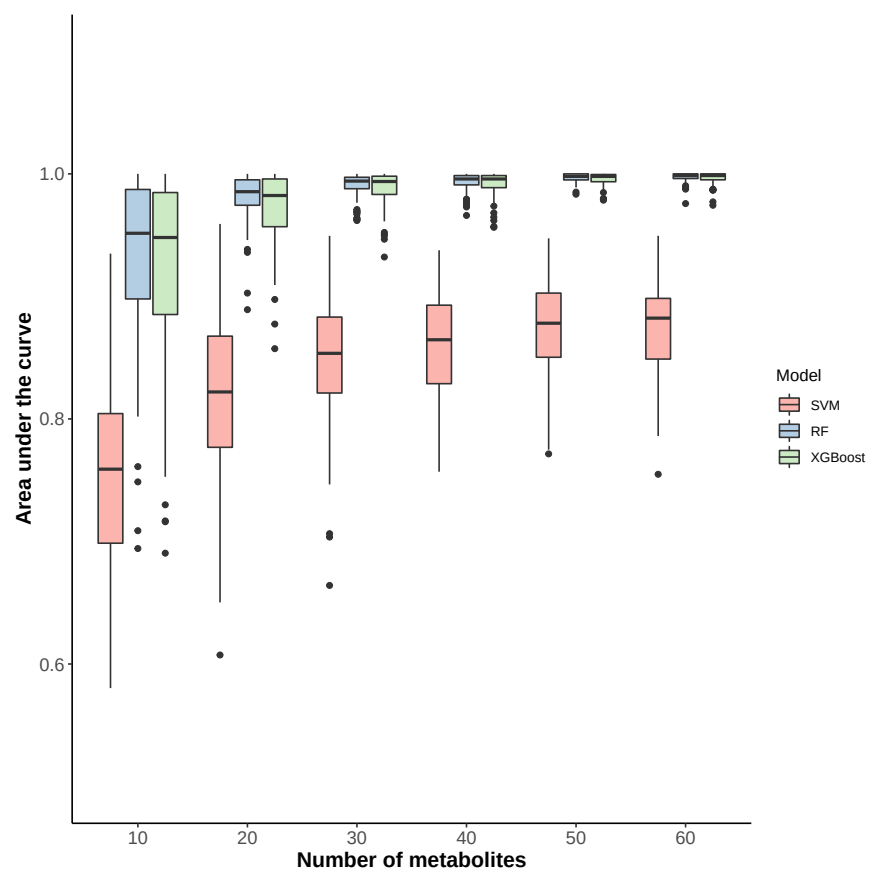


Figure C.7: Area under the curve of high night time temperature conditions (control and stress) using different number of metabolites using support vector machine (SVM), random forests (RF), and extreme gradient boosting (XGBoost).

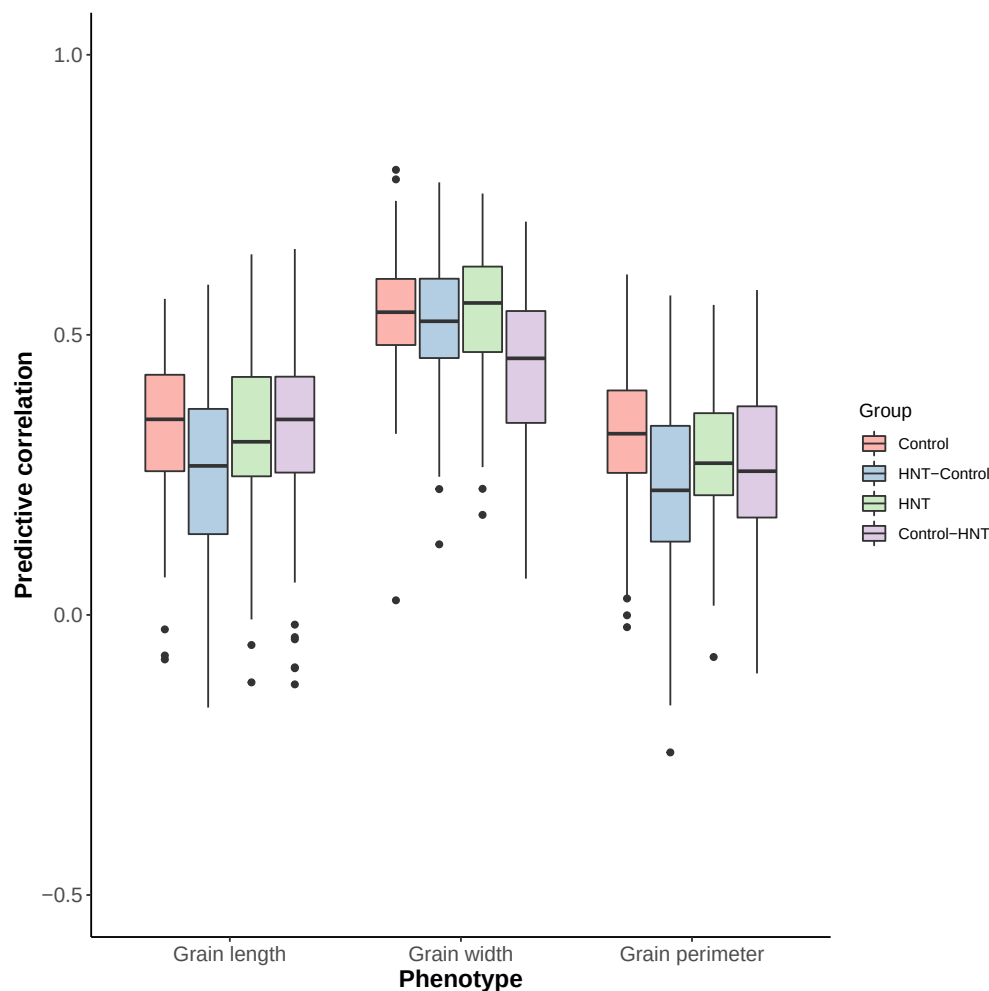


Figure C.8: Metabolic predictions across high night time temperature (HNT) stress conditions. Control: Train the model using phenotypes and metabolites in control and predict control phenotypes from metabolites in control conditions. Control-HNT: Train the model using phenotypes and metabolites in control and predict HNT phenotypes from metabolites in HNT conditions using the metabolic effect estimated in control conditions. HNT: Train the model using phenotypes and metabolites in HNT and predict HNT phenotypes from metabolites in HNT conditions. HNT-Control: Train the model using phenotypes and metabolites in HNT and predict phenotypes from metabolites in control conditions using the metabolic effect estimated in HNT condition.

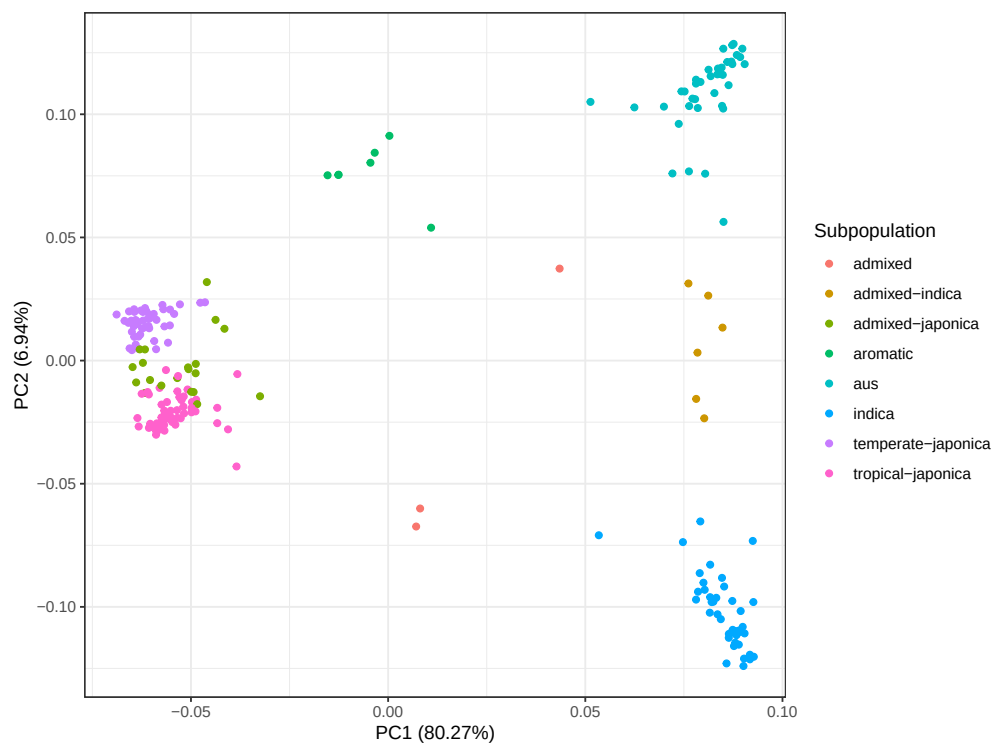


Figure C.9: Principal component analysis plot based on the genomic relationship matrix.

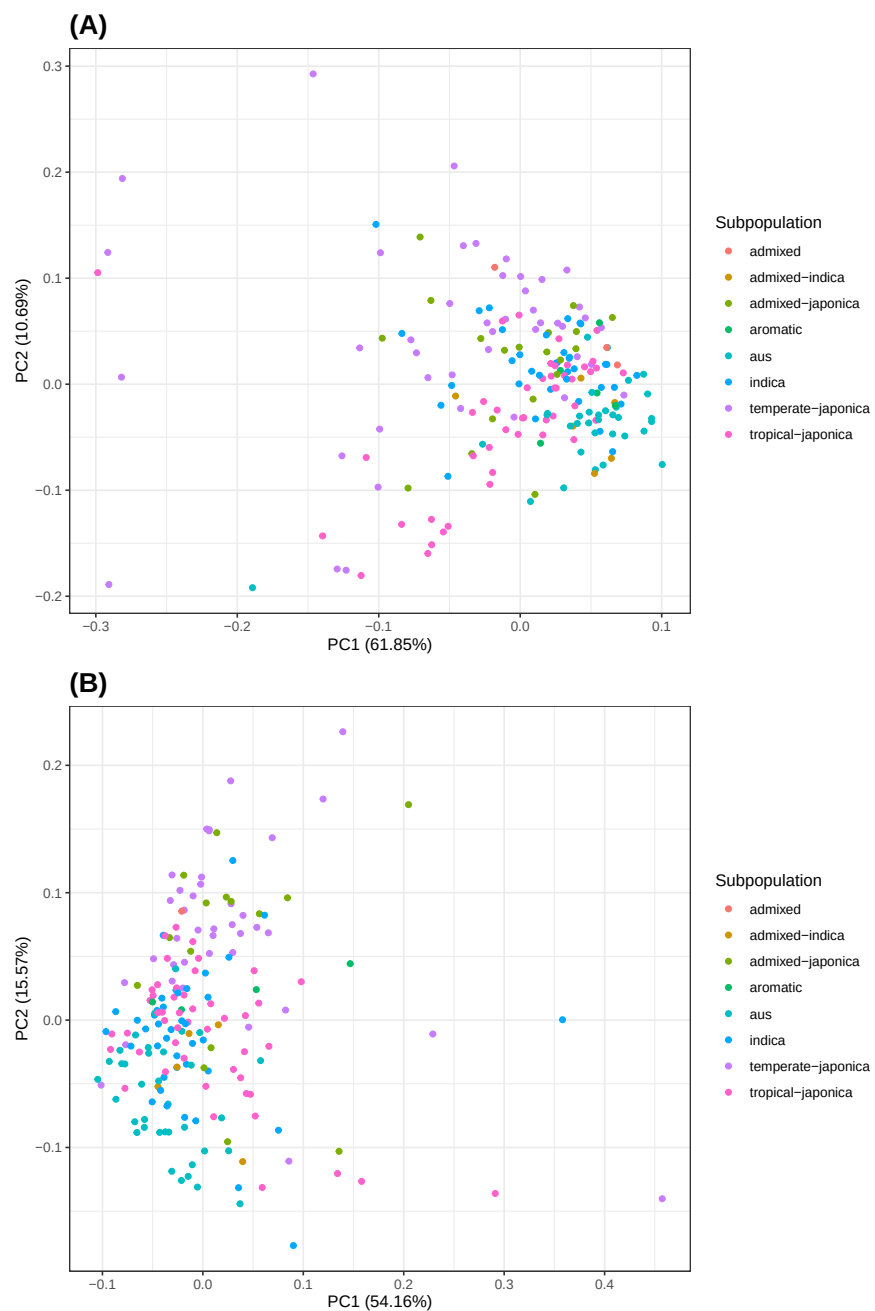


Figure C.10: Principal component analysis plot based on the metabolic relationship matrix in control (A) and high night time temperature stress conditions (B).

Appendix D

Genomic prediction of metabolic
content in rice grain in response to
warmer night conditions

Table D.1: List of metabolites analyzed in the current study.

Number	Metabolite	Number	Metabolite
1	hexanoic acid	2	alanine
3	valine	4	urea
5	ethanolamine	6	leucine
7	glycerol	8	nicotinic acid
9	isoleucine	10	proline
11	glycine	12	glyceric acid
13	serine	14	threonine
15	Beta alanine	16	malic acid
17	dihydrouracil	18	threitol
19	methionine	20	aspartic acid
21	cytosine	22	trans-4-hydroxy L-proline
23	gamma-aminobutyric acid (GABA)	24	glutamic acid
25	hydroxybenzoic acid	26	asparagine
27	lyxose	28	xylitol
29	arabitol	30	diglycerol
31	4-hydroxy-3-methoxybenzoic acid	32	glycerol-1 phosphate
33	glutamine	34	ornithine
35	citrulline	36	citric acid
37	adenine	38	fructose-1
39	allantoin-2	40	altrose
41	lysine	42	histidine
43	glucose	44	tyrosine
45	mannitol	46	sorbitol
47	indoleacetic acid	48	pantothenic acid
49	glucosaminic acid	50	allantoin-3
51	ferulic acid	52	N-acetyl-D-glucosamine
53	allo-inositol	54	tryptophan
55	3.5-dimethoxy-4-hydroxycinnamic acid	56	uridine
57	adenosine	58	trehalose
59	maltose	60	sophorose
61	melibiose	62	isomaltose
63	galactinol	64	phosphoric acid
65	Sucrose	66	Raffinose

APPENDIX D. GENOMIC PREDICTION OF METABOLIC CONTENT IN RICE GRAIN IN RESPONSE TO WARMER NIGHT CONDITIONS

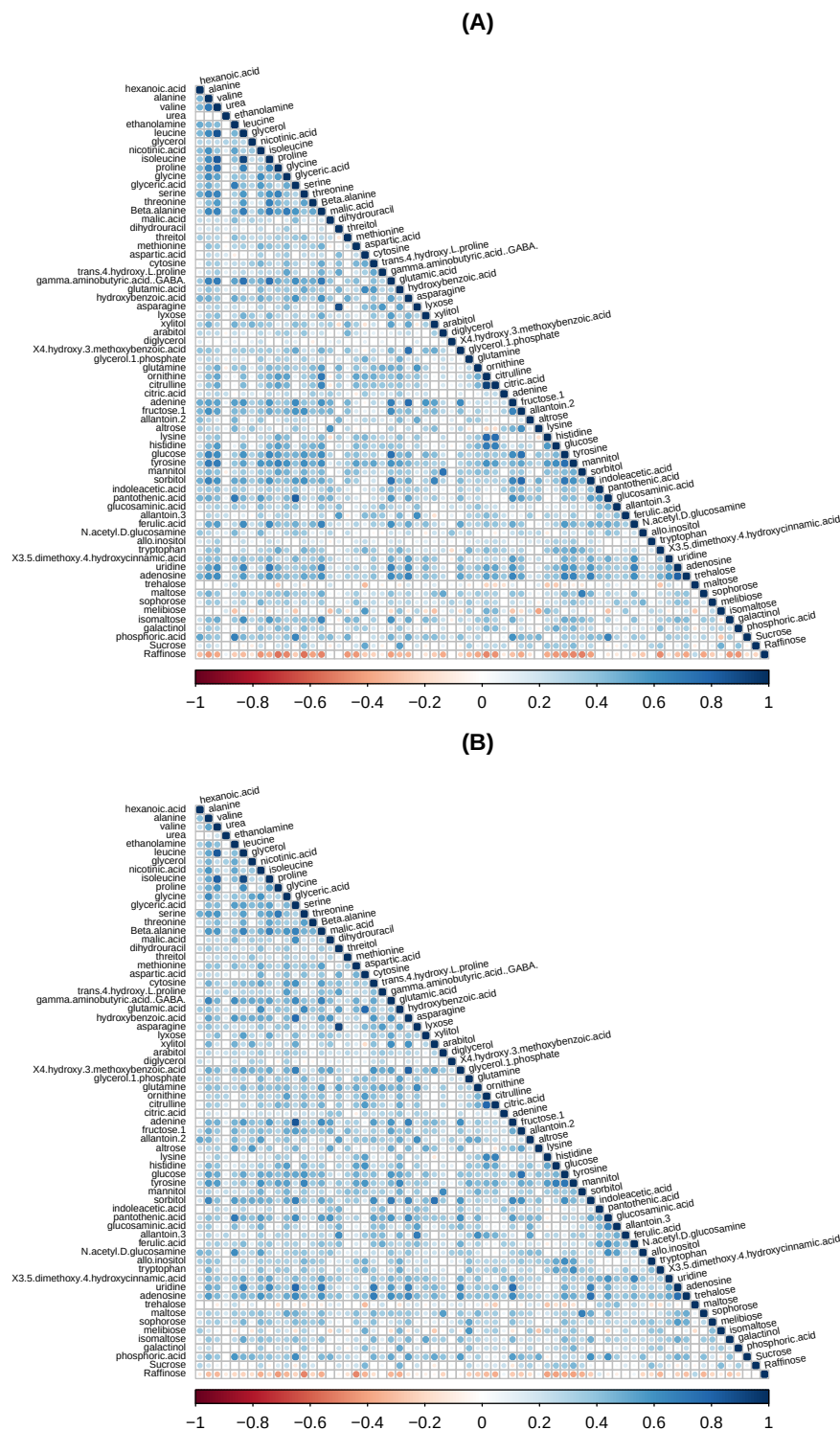


Figure D.1: Pearson correlation heat maps between metabolic profiles in control (A) and high night temperature stress (B).

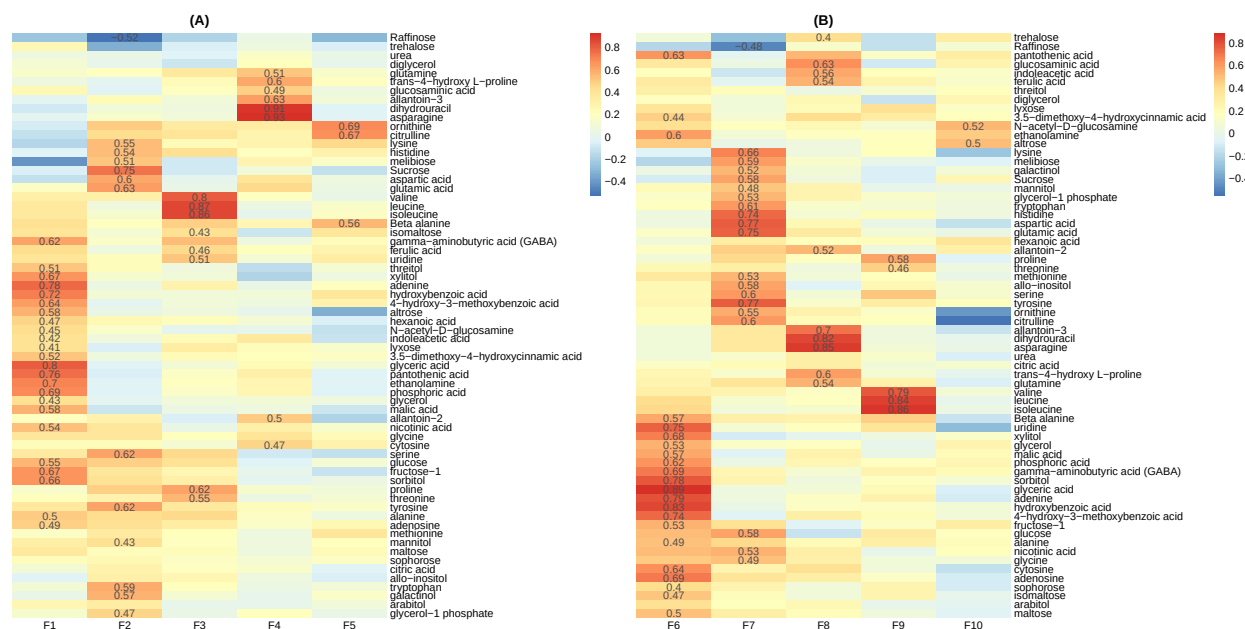


Figure D.2: Heat maps of factor loadings after removing cross-loading by setting a cutoff of $\lambda > |0.4|$. The rows of each panel correspond to the observed metabolite accumulation and the columns correspond to the five latent factors in control (A) and high night temperature stress conditions (B).

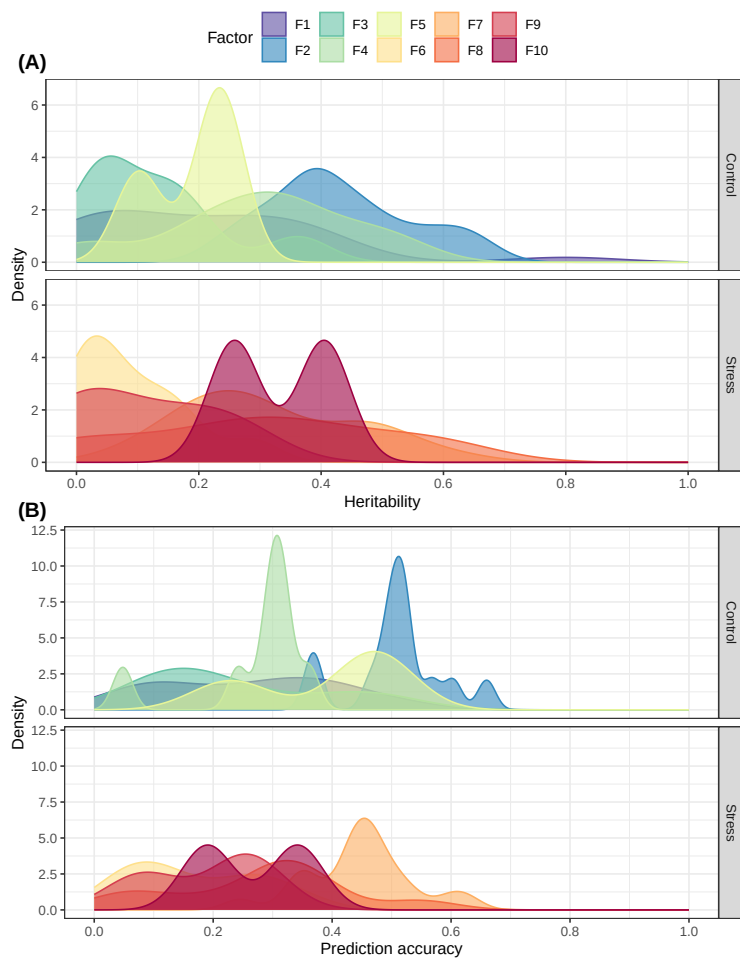


Figure D.3: Genomic heritability estimates of metabolite contents in control and high night temperature stress conditions (A) and genomic prediction accuracy of metabolite contents in control and high night temperature stress conditions, separated by latent factors (B). F1-F5 and F6-F10 are latent factors in control and high night temperature stress conditions, respectively.

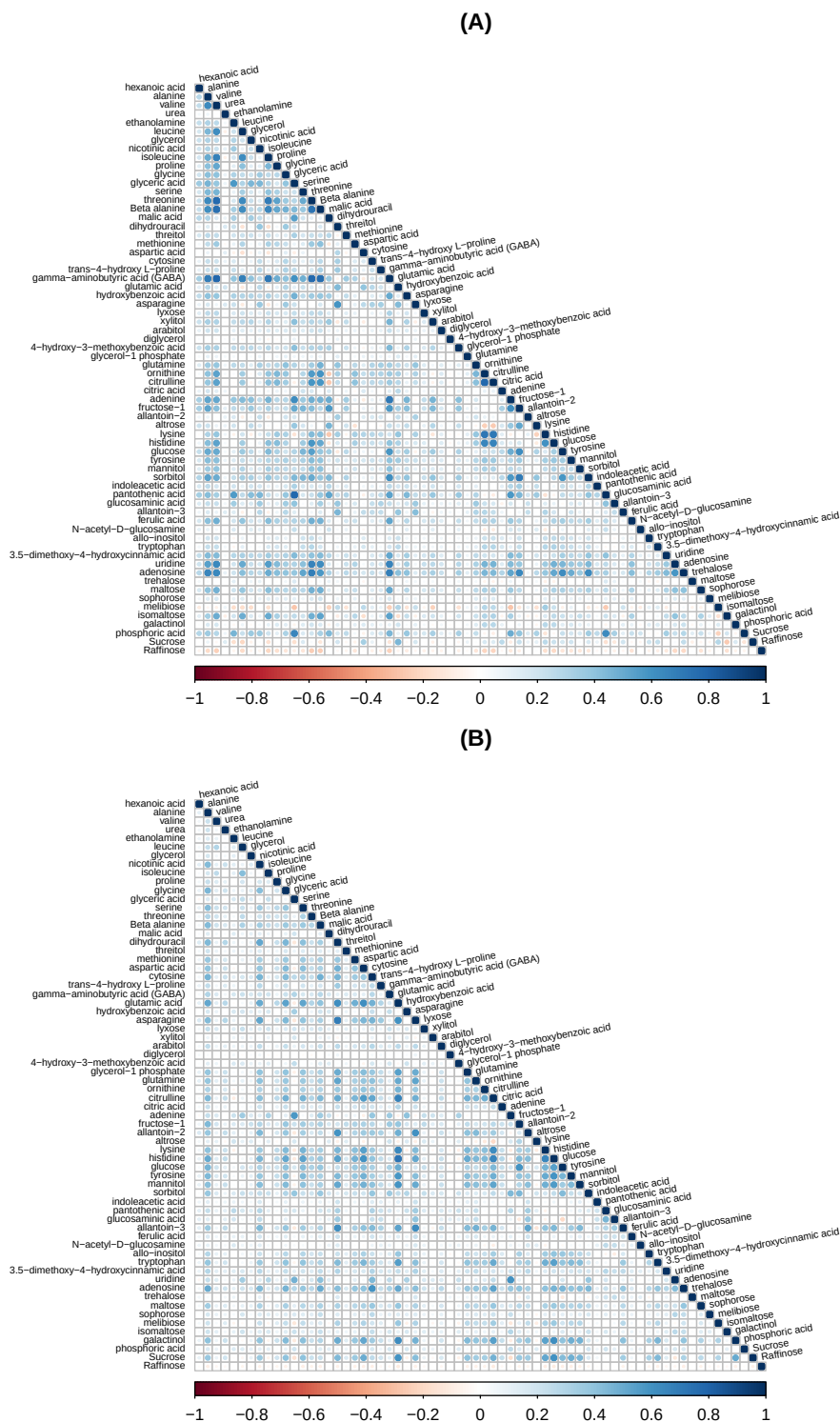


Figure D.4: Genomic correlation estimates between different metabolites in control (A) and high night temperature stress (B).

Bibliography

- [1] Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN.
- [2] Afridi, H., Ullah, M., Nordbø, Ø., Hoff, S. C., Furre, S., Larsgard, A. G., and Cheikh, F. A. (2024). Analyzing data modalities for cattle weight estimation using deep learning models. *Journal of Imaging*, 10(3):72.
- [3] Anderson, T. (2003). An introduction to multivariate statistical analysis (wiley series in probability and statistics). *July 11*.
- [4] Baba, T., Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *Plos one*, 15(2):e0228118.
- [5] Baba, T., Pegolo, S., Mota, L. F., Peñagaricano, F., Bittante, G., Cecchinato, A., and Morota, G. (2021). Integrating genomic and infrared spectral data improves the prediction of milk protein composition in dairy cattle. *Genetics Selection Evolution*, 53(1):1–14.
- [6] Banhazi, T., Tschärke, M., Ferdous, W., Saunders, C., and Lee, S. (2011). Improved image analysis based system to reliably predict the live weight of pigs on farm: Preliminary results. *Australian Journal of Multi-disciplinary Engineering*, 8(2):107–119.
- [7] Bartholomé, J., Prakash, P. T., and Cobb, J. N. (2022). Genomic prediction: progress and perspectives for rice improvement. *Complex Trait Prediction*, pages 569–617.
- [8] Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

- [9] Bello, R.-W., Mohamed, A. S. A., and Talib, A. Z. (2021). Enhanced mask r-cnn for herd segmentation. *International Journal of Agricultural and Biological Engineering*, 14(4):238–244.
- [10] Bi, Y., Campos, L. M., Wang, J., Yu, H., Hanigan, M. D., and Morota, G. (2023a). Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms. *Smart Agricultural Technology*, 6:100352.
- [11] Bi, Y., Yassue, R. M., Paul, P., Dhatt, B. K., Sandhu, J., Do, P. T., Walia, H., Obata, T., and Morota, G. (2023b). Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice. *G3: Genes, Genomes, Genetics*, 13(5):jkad052.
- [12] Brocklehurst, P. (1977). Factors controlling grain weight in wheat. *Nature*, 266(5600):348–349.
- [13] Brzozowski, L. J., Campbell, M. T., Hu, H., Caffè, M., Gutiérrez, L., Smith, K. P., Sorrells, M. E., Gore, M. A., and Jannink, J.-L. (2022). Generalizable approaches for genomic prediction of metabolites in plants. *The Plant Genome*, 15(2):e20205.
- [14] Campbell, M. T., Hu, H., Yeats, T. H., Brzozowski, L. J., Caffè-Treml, M., Gutiérrez, L., Smith, K. P., Sorrells, M. E., Gore, M. A., and Jannink, J.-L. (2021). Improving genomic prediction for seed quality traits in oat (*avena sativa* l.) using trait-specific relationship matrices. *Frontiers in genetics*, 12:643733.
- [15] Cang, Y., He, H., and Qiao, Y. (2019). An intelligent pig weights estimate method based on deep learning in sow stall environments. *IEEE Access*, 7:164867–164875.
- [16] Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., and Walia, H. (2016a).

- Heat stress yields a unique mads box transcription factor in determining seed size and thermal sensitivity. *Plant physiology*, 171(1):606–622.
- [17] Chen, C., Zhu, W., and Norton, T. (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and Electronics in Agriculture*, 187:106255.
- [18] Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., Li, Y., Liu, X., Zhang, H., Dong, H., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature genetics*, 46(7):714–721.
- [19] Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., et al. (2016b). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature communications*, 7(1):12767.
- [20] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1251–1258.
- [21] Condotta, I. C., Brown-Brandl, T. M., Pitla, S. K., Stinn, J. P., and Silva-Miranda, K. O. (2020). Evaluation of low-cost depth cameras for agricultural applications. *Computers and Electronics in Agriculture*, 173:105394.
- [22] Condotta, I. C., Brown-Brandl, T. M., Silva-Miranda, K. O., and Stinn, J. P. (2018). Evaluation of a depth sensor for mass estimation of growing and finishing pigs. *Biosystems Engineering*, 173:11–18.
- [23] Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the r package sommer. *PloS one*, 11(6):e0156744.

- [24] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee.
- [25] Dettmers, T., Lewis, M., Shleifer, S., and Zettlemoyer, L. (2021). 8-bit optimizers via block-wise quantization. *arXiv preprint arXiv:2110.02861*.
- [26] Dhatt, B. K., Abshire, N., Paul, P., Hasanthika, K., Sandhu, J., Zhang, Q., Obata, T., and Walia, H. (2019). Metabolic dynamics of developing rice seeds under high night-time temperature stress. *Frontiers in Plant Science*, 10:1443.
- [27] Dhatt, B. K., Paul, P., Sandhu, J., Hussain, W., Irvin, L., Zhu, F., Adviento-Borbe, M. A., Lorence, A., Staswick, P., Yu, H., et al. (2021). Allelic variation in rice fertilization independent endosperm 1 contributes to grain width under high night temperature stress. *New Phytologist*, 229(1):335–350.
- [28] Dickinson, R., Morton, J., Beggs, D., Anderson, G., Pyman, M., Mansell, P., and Blackwood, C. (2013). An automated walk-over weighing system as a tool for measuring liveweight change in lactating dairy cows. *Journal of Dairy Science*, 96(7):4477–4486.
- [29] Donat, M. G. and Alexander, L. V. (2012). The shifting probability distribution of global daytime and night-time temperatures. *Geophysical Research Letters*, 39(14).
- [30] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [31] Douglas, S., Edwards, S., and Kyriazakis, I. (2014). Management strategies to improve the performance of low birth weight pigs to weaning and their long-term consequences. *Journal of Animal Science*, 92(5):2280–2288.

- [32] Fernandes, A. F., Dórea, J. R., Fitzgerald, R., Herring, W., and 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.
- [33] Fernandes, A. F. A., Dórea, J. R. R., and Rosa, G. J. d. M. (2020). Image analysis and computer vision applications in animal sciences: an overview. *Frontiers in Veterinary Science*, 7:551269.
- [34] Fernie, A. R. and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in genetics*, 25(1):39–48.
- [35] Fernie, A. R. and Tohge, T. (2017). The genetics of plant metabolism. *Annual Review of Genetics*, 51:287–310.
- [36] Fitzgerald, M. A., McCouch, S. R., and Hall, R. D. (2009). Not just a grain of rice: the quest for quality. *Trends in plant science*, 14(3):133–139.
- [37] Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice fertilization-independent endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiology*, 165(1):238–248.
- [38] Galli, G., Sabadin, F., Yassue, R. M., Galves, C., Carvalho, H. F., Crossa, J., Montesinos-López, O. A., and Fritsche-Neto, R. (2022). Automated machine learning: a case study of genomic “image-based” prediction in maize hybrids. *Frontiers in Plant Science*, 13:845524.
- [39] Gan, H., Menegon, F., Sun, A., Scollo, A., Jiang, Q., Xue, Y., and Norton, T. (2024). Peeking into the unseen: Occlusion-resistant segmentation for preweaning piglets under crushing events. *Computers and Electronics in Agriculture*, 219:108683.

- [40] Gianola, D. and Van Kaam, J. B. (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, 178(4):2289–2303.
- [41] Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., and Luo, J. (2013). Genetic analysis of the metabolome exemplified using a rice population. *Proceedings of the National Academy of Sciences*, 110(50):20320–20325.
- [42] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- [43] Grandin, T. and Shivley, C. (2015). How farm animals react and perceive stressful situations such as handling, restraint, and transport. *Animals*, 5(4):1233–1251.
- [44] Groesbeck, C., Lawrence, K., Young, M., Goodband, R. D., DeRouchey, J. M., Tokach, M. D., Nelssen, J. L., and Dritz, S. S. (2002). Using heart girth to determine weight in finishing pigs. *Kansas Agricultural Experiment Station Research Reports*, 10:166–168.
- [45] Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., and Zhu, D. (2017). LSSA_CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs. *Computers and Electronics in Agriculture*, 138:60–68.
- [46] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R. R., Cheng, M.-M., and Hu, S.-M. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8(3):331–368.
- [47] Guo, X., Sarup, P., Jensen, J. D., Orabi, J., Kristensen, N. H., Mulder, F. A., Jahoor, A., and Jensen, J. (2020). Genetic variance of metabolomic features and their relationship with malting quality traits in spring barley. *Frontiers in Plant Science*, 11:575467.
- [48] Guo, Y.-z., Zhu, W.-x., Jiao, P.-p., Ma, C.-h., and Yang, J.-j. (2015). Multi-object

- extraction from topview group-housed pig images based on adaptive partitioning and multilevel thresholding segmentation. *Biosystems Engineering*, 135:54–60.
- [49] Guo, Z., Magwire, M. M., Basten, C. J., Xu, Z., and Wang, D. (2016). Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. *Theoretical and applied genetics*, 129(12):2413–2427.
- [50] Hamner, B. and Frasco, M. (2018). *Metrics: Evaluation Metrics for Machine Learning*. R package version 0.1.4.
- [51] Hansen, M. F., Smith, M. L., Smith, L. N., Jabbar, K. A., and Forbes, D. (2018a). Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device. *Computers in Industry*, 98:14–22.
- [52] Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., and Grieve, B. (2018b). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*, 98:145–152.
- [53] Hao, H., Jincheng, Y., Ling, Y., Gengyuan, C., Sumin, Z., and Huan, Z. (2023). An improved PointNet++ point cloud segmentation model applied to automatic measurement method of pig body size. *Computers and Electronics in Agriculture*, 205:107560.
- [54] Hayes, B. J., Daetwyler, H. D., and Goddard, M. E. (2016). Models for genome $\times$ environment interaction: examples in livestock. *Crop Science*, 56(5):2251–2259.
- [55] Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2):191–205.
- [56] He, K., Girshick, R., and Dollár, P. (2019). Rethinking imagenet pre-training. In

- Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927.
- [57] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- [58] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.
- [59] He, W., Mi, Y., Ding, X., Liu, G., and Li, T. (2023). Two-stream cross-attention vision Transformer based on RGB-D images for pig weight estimation. *Computers and Electronics in Agriculture*, 212:107986.
- [60] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82.
- [61] Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.
- [62] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324.
- [63] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [64] Hu, C., Shi, J., Quan, S., Cui, B., Kleessen, S., Nikoloski, Z., Tohge, T., Alexander, D., Guo, L., Lin, H., et al. (2014). Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. *Scientific reports*, 4(1):5067.

- [65] Hu, H., Campbell, M. T., Yeats, T. H., Zheng, X., Runcie, D. E., Covarrubias-Pazarán, G., Broeckling, C., Yao, L., Caffè-Treml, M., Gutiérrez, L., et al. (2021). Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. *Theoretical and Applied Genetics*, 134(12):4043–4054.
- [66] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708.
- [67] Huang, R., Jiang, L., Zheng, J., Wang, T., Wang, H., Huang, Y., and Hong, Z. (2013). Genetic bases of rice grain shape: so many genes, so little known. *Trends in plant science*, 18(4):218–226.
- [68] Impa, S. M., Raju, B., Hein, N. T., Sandhu, J., Prasad, P. V., Walia, H., and Jagadish, S. K. (2021). High night temperature effects on wheat and rice: Current status and way forward. *Plant, Cell & Environment*, 44(7):2049–2065.
- [69] Impa, S. M., Sunoj, V. J., Krassovskaya, I., Bheemanahalli, R., Obata, T., and Jagadish, S. K. (2019). Carbon balance and source-sink metabolic changes in winter wheat exposed to high night-time temperature. *Plant, Cell & Environment*, 42(4):1233–1246.
- [70] Intel RealSense (Accessed: March 1, 2023.). Intel realsense github. <https://github.com/IntelRealSense>.
- [71] Jagadish, S., Murty, M., and Quick, W. (2015). Rice responses to rising temperatures—challenges, perspectives and future directions. *Plant, cell & environment*, 38(9):1686–1698.
- [72] Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., and He, D. (2019). FLYOLOv3 deep learn-

- ing for key parts of dairy cow body detection. *Computers and Electronics in Agriculture*, 166:104982.
- [73] Jiang, Y. and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics*, 201(2):759–768.
- [74] Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics YOLO.
- [75] Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4):443–482.
- [76] Jun, K., Kim, S. J., and Ji, H. W. (2018). Estimating pig weights from images without constraint on posture and illumination. *Computers and Electronics in Agriculture*, 153:169–176.
- [77] Kadlec, R., Indest, S., Castro, K., Waqar, S., Campos, L. M., Amorim, S. T., Bi, Y., Hanigan, M. D., and Morota, G. (2022). Automated acquisition of top-view dairy cow depth image data using an rgb-d sensor camera. *Translational Animal Science*, 6(4):txac163.
- [78] Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.
- [79] Kashiha, M., Bahr, C., Ott, S., Moons, C. P., Niewold, T. A., Ödberg, F. O., and Berckmans, D. (2014). Automatic weight estimation of individual pigs using image analysis. *Computers and Electronics in Agriculture*, 107:38–44.
- [80] Kim, J. K., Park, S.-Y., Lim, S.-H., Yeo, Y., Cho, H. S., and Ha, S.-H. (2013). Comparative metabolic profiling of pigmented rice (*oryza sativa* l.) cultivars reveals primary metabolites are correlated with secondary metabolites. *Journal of Cereal Science*, 57(1):14–20.

- [81] Kizilkaya, K., Fernando, R., and Garrick, D. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *Journal of animal science*, 88(2):544–551.
- [82] Kliebenstein, D. (2009). Advancing genetic theory and application by metabolic quantitative trait loci analysis. *The plant cell*, 21(6):1637–1646.
- [83] Kuhn, M. (2015). Caret: classification and regression training. *Astrophysics Source Code Library*, pages ascl–1505.
- [84] Kusano, M., Fukushima, A., Kobayashi, M., Hayashi, N., Jonsson, P., Moritz, T., Ebana, K., and Saito, K. (2007). Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *Journal of Chromatography B*, 855(1):71–79.
- [85] Kusano, M., Yang, Z., Okazaki, Y., Nakabayashi, R., Fukushima, A., and Saito, K. (2015). Using metabolomic approaches to explore chemical diversity in rice. *Molecular plant*, 8(1):58–67.
- [86] Kuzuhara, Y., Kawamura, K., Yoshitoshi, R., Tamaki, T., Sugai, S., Ikegami, M., Kurokawa, Y., Obitsu, T., Okita, M., Sugino, T., et al. (2015). A preliminary study for predicting body weight and milk properties in lactating holstein cows using a three-dimensional camera system. *Computers and Electronics in Agriculture*, 111:186–193.
- [87] Kwon, K., Park, A., Lee, H., and Mun, D. (2023). Deep learning-based weight estimation using a fast-reconstructed mesh model from the point cloud of a pig. *Computers and Electronics in Agriculture*, 210:107903.
- [88] Larriestra, A., Wattanaphansak, S., Neumann, E., Bradford, J., Morrison, R., and

- Deen, J. (2006). Pig characteristics associated with mortality and light exit weight for the nursery phase. *The Canadian Veterinary Journal*, 47(6):560.
- [89] Le Cozler, Y., Allain, C., Xavier, C., Depuille, L., Caillot, A., Delouard, J., Delattre, L., Luginbuhl, T., and Faverdin, P. (2019). Volume and surface area of Holstein dairy cows calculated from complete 3D shapes acquired using a high-precision scanning system: Interest for body weight estimation. *Computers and Electronics in Agriculture*, 165:104977.
- [90] Le Cozler, Y., Brachet, E., Bourguignon, L., Delattre, L., Luginbuhl, T., and Faverdin, P. (2022). Three-dimensional (3D) imaging technology to monitor growth and development of holstein heifers and estimate body weight, a preliminary study. *Sensors*, 22(12):4635.
- [91] Lee, S., Ahn, H., Seo, J., Chung, Y., Park, D., and Pan, S. (2019). Practical monitoring of undergrown pigs for IoT-based large-scale smart farm. *IEEE Access*, 7:173796–173810.
- [92] Li, X., Xiang, Y., and Li, S. (2023a). Combining convolutional and vision transformer structures for sheep face recognition. *Computers and Electronics in Agriculture*, 205:107651.
- [93] Li, Y., Hu, J., Wen, Y., Evangelidis, G., Salahi, K., Wang, Y., Tulyakov, S., and Ren, J. (2023b). Rethinking vision transformers for MobileNet size and speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16889–16900.
- [94] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8):2674.
- [95] Lin, Q., Ye, G., Wang, J., and Liu, H. (2022). Roboflow: a data-centric workflow management system for developing ai-enhanced robots. In *Conference on Robot Learning*, pages 1789–1794. PMLR.

- [96] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.
- [97] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.
- [98] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022.
- [99] Lizana, X. C., Riegel, R., Gomez, L. D., Herrera, J., Isla, A., McQueen-Mason, S. J., and Calderini, D. F. (2010). Expansins expression is associated with grain size dynamics in wheat (*triticum aestivum* l.). *Journal of experimental botany*, 61(4):1147–1157.
- [100] López-Vergé, S., Gasa, J., Farré, M., Coma, J., Bonet, J., and Solà-Oriol, D. (2018). Potential risk factors related to pig body weight variability from birth to slaughter in commercial conditions. *Translational Animal Science*, 2(4):383–395.
- [101] Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2_Part_1):209–220.
- [102] Martínez-Barajas, E., Delatte, T., Schluepmann, H., de Jong, G. J., Somsen, G. W., Nunes, C., Primavesi, L. F., Coello, P., Mitchell, R. A., and Paul, M. J. (2011). Wheat grain development is characterized by remarkable trehalose 6-phosphate accumulation pregrain filling: tissue distribution and relationship to snf1-related protein kinase1 activity. *Plant physiology*, 156(1):373–381.

- [103] Martinez-Carrasco, R. and Thorne, G. N. (1979). Physiological factors limiting grain size in wheat. *Journal of experimental botany*, 30(4):669–679.
- [104] Martins, B., Mendes, A., Silva, L., Moreira, T., Costa, J., Rotta, P., Chizzotti, M., and Marcondes, M. (2020). Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements. *Livestock Science*, 236:104054.
- [105] Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.-i., Ebana, K., Yano, M., and Saito, K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *The Plant Journal*, 81(1):13–23.
- [106] Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J.-I., Ebana, K., Yano, M., and Saito, K. (2012). Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. *The Plant Journal*, 70(4):624–636.
- [107] McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., Singh, N., DeClerck, G., Agosto-Perez, F., Korniliev, P., et al. (2016). Open access resources for genome-wide association mapping in rice. *Nature communications*, 7(1):1–14.
- [108] Meckbach, C., Tiesmeyer, V., and Traulsen, I. (2021). A promising approach towards precise animal weight monitoring using convolutional neural networks. *Computers and Electronics in Agriculture*, 183:106056.
- [109] Mehta, S., Ghazvininejad, M., Iyer, S., Zettlemoyer, L., and Hajishirzi, H. (2020). Delight: Deep and light-weight transformer. *arXiv preprint arXiv:2008.00623*.

- [110] Mehta, S. and Rastegari, M. (2021). MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.
- [111] Mehta, S. and Rastegari, M. (2022). Separable self-attention for mobile vision transformers. *arXiv preprint arXiv:2206.02680*.
- [112] Meuwissen, T. H., Hayes, B. J., and Goddard, M. (2001a). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, 157(4):1819–1829.
- [113] Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001b). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829.
- [114] Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., et al. (2017). Mixed precision training. *arXiv preprint arXiv:1710.03740*.
- [115] Momen, M., Bhatta, M., Hussain, W., Yu, H., and Morota, G. (2021). Modeling multiple phenotypes in wheat using data-driven genomic exploratory factor analysis and bayesian network learning. *Plant Direct*, 5(1):e00304.
- [116] Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., and Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC genomics*, 22:1–23.
- [117] Morita, S., Yonemaru, J.-I., and Takanashi, J.-i. (2005). Grain growth and endosperm cell size under high night temperatures in rice (*oryza sativa* l.). *Annals of Botany*, 95(4):695–701.
- [118] Morota, G. and Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Frontiers in Genetics*, 5:363.

- [119] Morota, G., Koyama, M., Rosa, G. J., Weigel, K. A., and Gianola, D. (2013). Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. *Genetics Selection Evolution*, 45(1):17.
- [120] Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., and Fernando, S. C. (2018). Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of Animal Science*, 96(4):1540–1550.
- [121] Muñoz-Tamayo, R., Nielsen, B. L., Gagaoua, M., Gondret, F., Krause, E. T., Morgavi, D. P., Olsson, I. A. S., Pastell, M., Taghipoor, M., Tedeschi, L., et al. (2022). Seven steps to enhance open science practices in animal science. *PNAS Nexus*, 1(3):pgac106.
- [122] Nguyen, A. H., Holt, J. P., Knauer, M. T., Abner, V. A., Lobaton, E. J., and Young, S. N. (2023). Towards rapid weight assessment of finishing pigs using a handheld, mobile RGB-D camera. *Biosystems Engineering*, 226:155–168.
- [123] Nir, O., Parmet, Y., Werner, D., Adin, G., and Halachmi, I. (2018). 3D Computer-vision system for automatically estimating heifer height and body mass. *Biosystems Engineering*, 173:4–10.
- [124] Norton, T. and Berckmans, D. (2017). Developing precision livestock farming tools for precision dairy farming. *Animal Frontiers*, 7(1):18–23.
- [125] Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., et al. (2015). Promoting an open research culture. *Science*, 348(6242):1422–1425.
- [126] Obata, T. and Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*, 69(19):3225–3243.

- [127] Oikawa, A., Matsuda, F., Kusano, M., Okazaki, Y., and Saito, K. (2008). Rice metabolomics. *Rice*, 1(1):63–71.
- [128] Olsen, O.-A. (2004). Nuclear endosperm development in cereals and arabidopsis thaliana. *The Plant Cell*, 16(suppl_1):S214–S227.
- [129] Onogi, A. (2023). A bayesian model for genomic prediction using metabolic networks. *bioRxiv*, pages 2023–03.
- [130] Oszvald, M., Primavesi, L. F., Griffiths, C. A., Cohn, J., Basu, S. S., Nuccio, M. L., and Paul, M. J. (2018). Trehalose 6-phosphate regulates photosynthesis and assimilate partitioning in reproductive tissue. *Plant Physiology*, 176(4):2623–2638.
- [131] O’Connell, M., Lynch, P., Bertholot, S., Verlait, F., and Lawlor, P. (2007). Measuring changes in physical size and predicting weight of sows during gestation. *Animal*, 1(9):1335–1343.
- [132] Paudel, S., de Sousa, R. V., Sharma, S. R., and Brown-Brandl, T. (2023). Deep learning models to predict finishing pig weight using point clouds. *Animals*, 14(1):31.
- [133] Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., and Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences*, 101(27):9971–9975.
- [134] Peng, S., Piao, S., Ciais, P., Myneni, R. B., Chen, A., Chevallier, F., Dolman, A. J., Janssens, I. A., Penuelas, J., Zhang, G., et al. (2013). Asymmetric effects of daytime and night-time warming on northern hemisphere vegetation. *Nature*, 501(7465):88–92.
- [135] Pérez, P. and de los Campos, G. (2014). Bglr: a statistical package for whole genome regression and prediction. *Genetics*, 198(2):483–495.

- [136] Pérez-Enciso, M. and Steibel, J. P. (2021). Phenomes: the current frontier in animal breeding. *Genetics Selection Evolution*, 53(1):22.
- [137] Pérez-Rodríguez, P. and de Los Campos, G. (2022). Multitrait bayesian shrinkage and variable selection models with the bglr-r package. *Genetics*, 222(1):iyac112.
- [138] Pezzuolo, A., Guarino, M., Sartori, L., González, L. A., and Marinello, F. (2018a). On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. *Computers and Electronics in Agriculture*, 148:29–36.
- [139] Pezzuolo, A., Guarino, M., Sartori, L., and Marinello, F. (2018b). A feasibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns. *Sensors*, 18(2):673.
- [140] Philipp, N., Weichert, H., Bohra, U., Weschke, W., Schulthess, A. W., and Weber, H. (2018). Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat. *PLoS One*, 13(10):e0205452.
- [141] Pielot, R., Kohl, S., Manz, B., Rutten, T., Weier, D., Tarkowská, D., Rolčik, J., Strnad, M., Volke, F., Weber, H., et al. (2015). Hormone-mediated growth dynamics of the barley pericarp as revealed by magnetic resonance imaging and transcript profiling. *Journal of Experimental Botany*, 66(21):6927–6943.
- [142] Ponnu, J., Wahl, V., and Schmid, M. (2011). Trehalose-6-phosphate: connecting plant metabolism and development. *Frontiers in Plant Science*, 2:70.
- [143] Qiao, Y., Truman, M., and Sukkarieh, S. (2019). Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Computers and Electronics in Agriculture*, 165:104958.

- [144] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [145] Ren, R.-J., Wang, P., Wang, L.-N., Su, J.-P., Sun, L.-J., Sun, Y., Chen, D.-F., and Chen, X.-W. (2020). Os4bglu14, a monolignol β -glucosidase, negatively affects seed longevity by influencing primary metabolism in rice. *Plant Molecular Biology*, 104(4):513–527.
- [146] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- [147] Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.8.12.
- [148] Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., Boggess, M., Cheng, H., Clutter, A., Cockett, N., et al. (2019). Genome to phenome: improving animal health, production, and well-being—a new usda blueprint for animal genome research 2018–2027. *Frontiers in genetics*, 10:327.
- [149] Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics*, 44(2):217–220.
- [150] Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics*, 15(3):469–485.
- [151] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller,

- M. (2011). proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC bioinformatics*, 12(1):1–8.
- [152] Runcie, D. E., Qu, J., Cheng, H., and Crawford, L. (2021). Megalmm: mega-scale linear mixed models for genomic predictions with thousands of traits. *Genome biology*, 22(1):1–25.
- [153] Rutherford, K. M., Donald, R. D., Lawrence, A. B., and Wemelsfelder, F. (2012). Qualitative behavioural assessment of emotionality in pigs. *Applied Animal Behaviour Science*, 139(3-4):218–224.
- [154] Saito, K. (2013). Phytochemical genomics—a new trend. *Current opinion in plant biology*, 16(3):373–380.
- [155] Saito, K. and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annual review of plant biology*, 61:463–489.
- [156] Sakai, H., Cheng, W., Chen, C. P., and Hasegawa, T. (2022). Short-term high night-time temperatures pose an emerging risk to rice grain failure. *Agricultural and Forest Meteorology*, 314:108779.
- [157] Salau, J. and Krieter, J. (2020). Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting. *Animals*, 10(12):2402.
- [158] Saleem, M. H., Potgieter, J., and Arif, K. M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precision Agriculture*, 22(6):2053–2091.
- [159] Sandhu, J., Irvin, L., Chandaran, A. K., Oguro, S., Paul, P., Dhatt, B., Hussain, W., Cunningham, S. S., Quinones, C. O., Lorence, A., et al. (2024). Natural variation in lonely

- guy-like 1 regulates rice grain weight under warmer night conditions. *Plant Physiology*, page kiae313.
- [160] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4510–4520.
- [161] Schrag, T. A., Westhues, M., Schipprack, W., Seifert, F., Thiemann, A., Scholten, S., and Melchinger, A. E. (2018). Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics*, 208(4):1373–1385.
- [162] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128:336–359.
- [163] Sharma, A., Vans, E., Shigemizu, D., Boroevich, K. A., and Tsunoda, T. (2019). Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific reports*, 9(1):11399.
- [164] Silva, F. F., Morota, G., and Rosa, G. J. d. M. (2021). High-throughput phenotyping in the genomic improvement of livestock.
- [165] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [166] Smeeckens, S. (2015). From leaf to kernel: trehalose-6-phosphate signaling moves carbon in the field. *Plant Physiology*, 169(2):912–913.
- [167] Song, X., Bokkers, E., Van der Tol, P., Koerkamp, P. G., and Van Mourik, S. (2018).

- Automated body weight prediction of dairy cows using 3-dimensional vision. *Journal of Dairy Science*, 101(5):4448–4459.
- [168] Sreenivasulu, N., Butardo Jr, V. M., Misra, G., Cuevas, R. P., Anacleto, R., and Kavi Kishor, P. B. (2015). Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. *Journal of Experimental Botany*, 66(7):1737–1748.
- [169] Steibel, J. P. (2023). Phenomics in animal breeding. In *Encyclopedia of Smart Agriculture Technologies*, pages 1–8. Springer.
- [170] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- [171] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826.
- [172] Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR.
- [173] Tasdemir, S., Urkmez, A., and Inal, S. (2011). Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis. *Computers and electronics in agriculture*, 76(2):189–197.
- [174] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- [175] Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

- [176] VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11):4414–4423.
- [177] Vose, R. S., Easterling, D. R., and Gleason, B. (2005). Maximum and minimum temperature trends for the globe: An update through 2004. *Geophysical Research Letters*, 32(23).
- [178] Wada, H., Hatakeyama, Y., Onda, Y., Nonami, H., Nakashima, T., Erra-Balsells, R., Morita, S., Hiraoka, K., Tanaka, F., and Nakano, H. (2019). Multiple strategies for heat adaptation to prevent chalkiness in the rice endosperm. *Journal of experimental botany*, 70(4):1299–1311.
- [179] Wada, K. (2018). labelme: Image polygonal annotation with python. <https://github.com/wkentaro/labelme>.
- [180] Wang, K., Li, Y., Wang, Y., and Yang, X. (2017). On the asymmetry of the urban daily air temperature cycle. *Journal of Geophysical Research: Atmospheres*, 122(11):5625–5635.
- [181] Wang, R., Gao, R., Li, Q., Zhao, C., Ru, L., Ding, L., Yu, L., and Ma, W. (2024). An ultra-lightweight method for individual identification of cow-back pattern images in an open image set. *Expert Systems with Applications*, page 123529.
- [182] Wang, S., Wei, J., Li, R., Qu, H., Chater, J. M., Ma, R., Li, Y., Xie, W., and Jia, Z. (2019). Identification of optimal prediction models using multi-omic data for selecting hybrid rice. *Heredity*, 123(3):395–406.
- [183] Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., and Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 568–578.

- [184] Wang, Y., Mücher, S., Wang, W., Guo, L., and Kooistra, L. (2023). A review of three-dimensional computer vision used in precision livestock farming for cattle growth management. *Computers and Electronics in Agriculture*, 206:107687.
- [185] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- [186] Wase, N., Abshire, N., and Obata, T. (2022). High-throughput profiling of metabolic phenotypes using high-resolution gc-ms. In *High-Throughput Plant Phenotyping*, pages 235–260. Springer.
- [187] Wei, J., Tang, X., Liu, J., and Zhang, Z. (2023). Detection of pig movement and aggression using deep learning approaches. *Animals*, 13(19):3074.
- [188] Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., and Dawe, D. (2010). Rice yields in tropical/subtropical asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proceedings of the National Academy of Sciences*, 107(33):14562–14567.
- [189] Westhues, M., Schrag, T. A., Heuer, C., Thaller, G., Utz, H. F., Schipprack, W., Thiemann, A., Seifert, F., Ehret, A., Schlereth, A., et al. (2017). Omics-based hybrid prediction in maize. *Theoretical and applied genetics*, 130(9):1927–1939.
- [190] Wheeler, T. and Von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341(6145):508–513.
- [191] Xavier, C., Le Cozler, Y., Depuille, L., Caillot, A., Lebreton, A., Allain, C., Delouard, J., Delattre, L., Luginbuhl, T., Faverdin, P., et al. (2022). The use of 3-dimensional

- imaging of Holstein cows to estimate body weight and monitor the composition of body weight change throughout lactation. *Journal of Dairy Science*, 105(5):4508–4519.
- [192] Xia, J., Chen, J., Piao, S., Ciais, P., Luo, Y., and Wan, S. (2014). Terrestrial carbon cycle affected by non-uniform climate warming. *Nature Geoscience*, 7(3):173–180.
- [193] Xiao, J., Liu, G., Wang, K., and Si, Y. (2022). Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM. *Computers and Electronics in Agriculture*, 194:106738.
- [194] Xing, Y., Zhang, Q., et al. (2010). Genetic and molecular bases of rice yield. *Annual review of plant biology*, 61(1):421–442.
- [195] Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., and Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, 171:105300.
- [196] Xu, S., Xu, Y., Gong, L., and Zhang, Q. (2016). Metabolomic prediction of yield in hybrid rice. *The Plant Journal*, 88(2):219–227.
- [197] Xu, Y., Xu, C., and Xu, S. (2017). Prediction and association mapping of agronomic traits in maize using multiple omic data. *Heredity*, 119(3):174–184.
- [198] Xu, Y., Zhao, Y., Wang, X., Ma, Y., Li, P., Yang, Z., Zhang, X., Xu, C., and Xu, S. (2021). Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. *Plant biotechnology journal*, 19(2):261–272.
- [199] Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al. (2010). Common snps explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–569.

- [200] Ying, J.-Z., Shan, J.-X., Gao, J.-P., Zhu, M.-Z., Shi, M., and Lin, H.-X. (2012). Identification of quantitative trait loci for lipid metabolism in rice seeds. *Molecular plant*, 5(4):865–875.
- [201] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27.
- [202] Yu, H., Lee, K., and Morota, G. (2021). Forecasting dynamic body weight of nonrestrained pigs from images using an RGB-D sensor camera. *Translational Animal Science*, 5(1):txab006.
- [203] Zhang, J., Zhuang, Y., Ji, H., and Teng, G. (2021). Pig weight and body size estimation using a multiple output regression convolutional neural network: A fast and fully automatic method. *Sensors*, 21(9):3218.
- [204] Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., et al. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35):9326–9331.
- [205] Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in oryza sativa. *Nature communications*, 2(1):1–10.
- [206] Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., Mock, H.-P., Matros, A., Ebmeyer, E., Schachschneider, R., et al. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proceedings of the National Academy of Sciences*, 112(51):15624–15629.

- [207] Zheng, Z., Li, J., and Qin, L. (2023). YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Computers and Electronics in Agriculture*, 209:107857.
- [208] Zhu, F., Paul, P., Hussain, W., Wallman, K., Dhatt, B. K., Sandhu, J., Irvin, L., Morota, G., Yu, H., and Walia, H. (2021). Seedextractor: an open-source gui for seed image analysis. *Frontiers in plant science*, 11:581546.