


## Article

# Prediction of Hanwoo Cattle Phenotypes from Genotypes Using Machine Learning Methods

Swati Srivastava <sup>†</sup>, Bryan Irvine Lopez <sup>†</sup>, Himansu Kumar , Myoungjin Jang, Han-Ha Chai, Woncheoul Park, Jong-Eun Park <sup>\*</sup> and Dajeong Lim <sup>\*</sup>

Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Wanju 55365, Korea; swati051@gmail.com (S.S.); irvinlopez@korea.kr (B.I.L.); himanshu.genetics@gmail.com (H.K.); minijmj@naver.com (M.J.); hanha@korea.kr (H.-H.C.); wcpark1982@korea.kr (W.P.)

\* Correspondence: jepark0105@korea.kr (J.-E.P.); lim.dj@korea.kr (D.L.)

† These authors contributed equally to this work.

**Simple Summary:** Machine learning has been extensively used in analyzing big data and in conditions where the number of parameters is much bigger than the number of observations. Recently, there have been an increasing number of successful applications of machine learning in genomic prediction as this method makes weaker assumptions, is capable of dealing with the dimensionality problem, and can be more flexible for describing complex relationships. In this study, we evaluated the predictive ability of three machine learning methods, namely, random forest (RF), extreme gradient boosting (XGB), and support vector machine (SVM), when predicting the carcass traits of Hanwoo cattle. These machine learning algorithms were compared with the standard linear method (GBLUP). Our results revealed that XGB method had the best predictive correlation for carcass weight and marbling score. Meanwhile, the best predictive correlation for backfat thickness and eye muscle area was delivered by GBLUP. Moreover, in terms of mean squared error (MSE) of prediction, GBLUP delivered the lowest MSE value for all traits.

**Abstract:** Hanwoo was originally raised for draft purposes, but the increase in local demand for red meat turned that purpose into full-scale meat-type cattle rearing; it is now considered one of the most economically important species and a vital food source for Koreans. The application of genomic selection in Hanwoo breeding programs in recent years was expected to lead to higher genetic progress. However, better statistical methods that can improve the genomic prediction accuracy are required. Hence, this study aimed to compare the predictive performance of three machine learning methods, namely, random forest (RF), extreme gradient boosting method (XGB), and support vector machine (SVM), when predicting the carcass weight (CWT), marbling score (MS), backfat thickness (BFT) and eye muscle area (EMA). Phenotypic and genotypic data (53,866 SNPs) from 7324 commercial Hanwoo cattle that were slaughtered at the age of around 30 months were used. The results showed that the boosting method XGB showed the highest predictive correlation for CWT and MS, followed by GBLUP, SVM, and RF. Meanwhile, the best predictive correlation for BFT and EMA was delivered by GBLUP, followed by SVM, RF, and XGB. Although XGB presented the highest predictive correlations for some traits, we did not find an advantage of XGB or any machine learning methods over GBLUP according to the mean squared error of prediction. Thus, we still recommend the use of GBLUP in the prediction of genomic breeding values for carcass traits in Hanwoo cattle.

**Keywords:** genomic prediction; machine learning; Hanwoo



**Citation:** Srivastava, S.; Lopez, B.I.; Kumar, H.; Jang, M.; Chai, H.-H.; Park, W.; Park, J.-E.; Lim, D. Prediction of Hanwoo Cattle Phenotypes from Genotypes Using Machine Learning Methods. *Animals* **2021**, *11*, 2066. <https://doi.org/10.3390/ani11072066>

Academic Editors: El Hamidi Hay and Tiago Paim

Received: 10 June 2021

Accepted: 9 July 2021

Published: 11 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Korean native cattle (Hanwoo) was originally raised for draft purposes, but the increase in local demand for red meat turned that purpose into full-scale meat-type cattle

rearing; it is now considered one of the foremost economically important species and a vital food source for Koreans [1]. This breed has been subjected to intensive selection for particular meat quality and production attributes over the past few decades; thus, a dramatic improvement has been obtained in terms of carcass weight and rib eye area [2]. The application of genomic selection in Hanwoo breeding schemes in recent years was expected to lead to higher genetic progress. In beef cattle, genomic prediction offers great promise to predict total genetic value of selection candidates, especially for traits that cannot be measured directly, such as carcass traits. The successful application of genomic selection relies on the accuracy of genomic estimated breeding values (GEBVs), which are mostly determined using estimation methods.

Up to date, different genomic prediction methods based on linear models have been developed, such as genomic best linear unbiased prediction (GBLUP) [3], single-step GBLUP [4], the Bayesian alphabet (Bayes A, Bayes B, Bayes C $\pi$ , and BayesR) [5–7], and the ridge regression BLUP (RR-BLUP) [8]. However, these statistical methods typically make strong assumptions about functional forms and the statistical distribution of marker effects. Thus, these methods pose statistical challenges related to high-dimensional genomic data and have difficulty capturing complex relationships between genotypes and phenotypes such as genotype-by-environment-by-trait interactions [9,10]. Recently, there have been an increasing number of successful applications of machine learning in genomic prediction [11]. These machine learning approaches make weaker assumptions, are capable of dealing with the dimensionality problem, and can be more flexible for describing complex relationships [12].

Machine learning methods, such as random forest (RF) [13], boosting [14], and support vector machine (SVM) [15], provide an appealing alternative to conventional statistical methods for genomic prediction of quantitative traits. They may provide an importance measure of predictor variables (SNPs) on a given trait and good predictive performance. RF and boosting are independent of model specification and, hence, may account for non-additive effects. Moreover, SVM is powerful at recognizing subtle patterns in complex datasets [15]. Recently, the extreme gradient boosting (XGB) [16] method was introduced with a similar principle to the gradient boosting method but with increased speed and less overfitting. Several studies using RF, XGB, and SVM have been used for genomic-based prediction in animal and plants [11,17,18].

The objective of this study was to compare the predictive performance of three machine learning methods, namely, RF, XGB, and SVM, with the conventional genomic prediction model (GBLUP) when predicting the carcass weight (CWT), marbling score (MS), backfat thickness (BFT) and eye muscle area (EMA) of Hanwoo cattle. Comparisons in terms of predictive correlation and mean squared error were used as metrics.

## 2. Materials and Methods

### 2.1. Data

The data on four carcass traits were collected from 7234 Hanwoo cattle slaughtered at the age of around 30 months. Carcass weight (CWT), marbling score (MS), backfat thickness (BFT), and eye muscle area (EMA) were the traits under study. The animals in this study were produced through the purebred mating system done using artificial insemination of semen collected from bulls initially selected on the basis of their performance and progeny carcass traits. All of these animals were gathered from different herds in nine provinces across South Korea. The ethics approval for this study was given by the Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration, Korea (2018-293). The descriptive statistics for each trait are presented in Table 1. Phenotypic records were adjusted for fixed effects as a function of a univariate analysis using the PREDICTF90 software package [19] in a pedigree-based model described in our previous study [20]. Briefly, the fixed effects that were used for all traits were herds, year-month of birth, year-month of slaughter, and slaughter place, along with sex and age as covariates.

**Table 1.** Descriptive statistics for carcass traits of Hanwoo cattle.

Trait	Mean	SD	Min	Max
CWT (in kg)	439.33	49.47	159	645
MS (1–9)	5.99	1.84	1	9
BFT (in mm)	14.24	4.78	1	45
EMA (in cm <sup>2</sup> )	96.15	11.96	35	155

SD, standard deviation; CWT, carcass weight; MS, marbling score; BFT, backfat thickness; EMA, eye muscle area.

A total of 7324 animals were genotyped for 53,866 SNPs using the customized Hanwoo 50K SNP Chip (Illumina, Korea) according to the manufacturer's protocol. The genomic DNA was quantified from tissue samples using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). The following threshold levels were applied for quality control using PLINK [21]: SNPs with minor allele frequency lower than 0.01, call rate lower than 0.90, Hardy–Weinberg disequilibrium with a *p*-value lower than 0.0000001, and situated on the sex chromosomes were removed from the genotype data. After quality control, 45,624 SNPs were retained for genomic prediction.

## 2.2. Statistical Methods

In this study, three machine learning algorithms, namely, random forest (RF), extreme gradient boosting (XGB), and support vector machine (SVM), were evaluated. These machine learning algorithms were compared with the standard linear method, GBLUP. The predictive performance of the different methods was assessed using a fivefold cross-validation scheme composed of five subpopulations that were randomly split into more or less equally sized groups. In cross-validation, each subpopulation (~1446) was given a chance to be used as the validation set and the other four subpopulations were used as the training set. The predictive correlation was calculated as a Pearson correlation between predicted and observed phenotypes. Furthermore, the mean squared errors of prediction were calculated.

### 2.2.1. Genomic Best Linear Unbiased Prediction (GBLUP)

The general animal model could be expressed as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{s}_i\alpha_i + \mathbf{Z}\mathbf{g} + \mathbf{e},$$

where  $\mathbf{y}$  is a vector of observed phenotypes,  $\mu$  is the overall mean,  $\mathbf{1}$  is a vector of ones,  $\mathbf{s}_i$  is a vector of genotypes for SNP<sub>*i*</sub> (coded as 0, 1, or 2),  $\alpha_i$  is the size of the effect of the marker (allele substitution effect),  $\mathbf{g}$  is a vector of the genomic breeding values of all individuals  $[\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)]$ , where  $\sigma_g^2$  is the additive genetic variance and  $\mathbf{G}$  is the marker-based genomic relationship matrix [3],  $\mathbf{Z}$  is an incidence matrix linking  $\mathbf{g}$  to  $\mathbf{y}$ , and  $\mathbf{e}$  is the vector of random residual effect  $[\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)]$ . The software MTG2 version 2.21 [22] was used to estimate variance components with restricted maximum likelihood (REML) and to calculate the genomic breeding values (GEBVs).

### 2.2.2. Random Forest (RF)

Random forest is a type of bagging method which is also known as bootstrap aggregating, and it was first proposed by Breiman [13]. It is a compilation of uncorrelated forests of trees whose prediction is more accurate than that of any single or group of trees. It estimates and fits a number of decision trees on various subsamples of the dataset and then uses their average to improve predictive accuracy and control overfitting. This method involves feature selection, generating predictors with the least correlation [13]. Therefore, the initial step was to identify significant features in our data. For this, the feature selection library present in scikit-learn was used. In order to find the best estimator, random search by cross-validation was used on hyperparameters. These parameters were 'n\_estimator' used to find number of trees in the forest, 'max\_features' used to find the maximum num-

ber of features considered for splitting a node, 'max\_depth' used to find the maximum number of levels in each tree, 'min\_samples\_split' used to find the minimum number of data points placed in a node before the node was split, and 'bootstrap' used for sampling data points (with or without replacement). This "fit" and "score" methods with parameters used in this approach were optimized by cross-validation search. The model parameters used in this study for each trait are shown in Table 2. These parameters were selected by the randomsearch function in scikit-learn using Python [23] and were used to build the RandomForestRegressor model.

**Table 2.** Parameters used to build model for RF method for each trait.

Trait	N_Estimator	Criterion	Max_Features	Min_Samples_Leaf	Min_Samples_Split	Max_Depth	Bootstrap
CWT	400	MSE	auto	4	10	70	TRUE
MS	600	MSE	auto	4	2	40	TRUE
BFT	2000	MSE	auto	2	2	90	TRUE
EMA	1400	MSE	auto	4	2	100	TRUE

MSE, mean square error; CWT, carcass weight; MS, marbling score; BFT, backfat thickness; EMA, eye muscle area.

### 2.2.3. Extreme Gradient Boosting Method (XGB)

The extreme gradient boosting (XGB) method [16] is a kind of ensemble machine learning algorithm that converts weak learners into strong learners, either for regression or for classification problems to reduce bias in supervised learning. This method applies the principle of boosting weak learners (CARTs generally) using the gradient descent architecture. It controls overfitting and can reduce prediction errors by utilizing more regularized model formation. The feature selection method was also used in this method to identify significant features through scikit-learn. Scores generated by this method generally gain value, generated by the decrease in prediction error of the objective function to a split node in a tree. Some of the important parameters considered to build the model were 'booster' to determine the type of learner, either its tree or linear function, 'eta' analogous to learning rate, 'min\_child\_weight' to determine the minimum sum of weights, 'max\_depth' to find the maximum number of levels in each tree, 'max\_leaf\_nodes' as the maximum number of terminal nodes, and 'gamma' reflecting the minimum loss function. The scikit-learn XGBRegressor [23] in Python was used for model construction. In this study, booster was selected as 'gbtree' (i.e., tree based model), 'eta' was kept as 0.3, 'min\_child\_weight' was kept as 1 (minimum sum of weights), maximum depth was selected as 6, maximum leaf nodes were selected as 6 (maximum number of terminal nodes), and gamma was kept as 0 (i.e., minimum loss function).

### 2.2.4. Support Vector Machine (SVM)

The support vector machine (SVM) is a supervised regression method that supports linear and nonlinear regression. Generally, SVM is used for classification or regression problems. It works on the basis of enlarging the feature space using various kernels such as linear, polynomial, and sigmoid Gaussian RBF (radial basis function). Linear kernels are mainly used for linear problems, whereas RBF kernels are used for nonlinear problems. A thorough guide and review on this method can be found in Smola and Schölkopf [24]. In this study, we applied the RBF kernel for building the model using epsilon-support vector regression. 'StandardScaler' and RandomForestRegressor models were built using scikit-learn [23], considering all features.

## 3. Results and Discussion

### 3.1. Genetic Parameters

Variance components and heritability estimates for each trait are presented in Table 3. Overall, the estimate of heritability for carcass traits in Hanwoo cattle was medium to high. Heritability estimates for CWT, MS, BFT, and EMA were 0.38, 0.44, 0.36, and 0.35,

respectively. The standard errors of estimated heritability for all traits were 0.02. The estimated heritability for each trait in this work was lower, higher, or in the range of previously reported estimates [25–27]. The observed differences between the estimates in this study and previous works may have been due to the population structure, number of records, fixed effects, and information (pedigree and/or genomic) used.

**Table 3.** Additive genetic variance ( $\sigma^2_a$ ), residual variance ( $\sigma^2_e$ ), phenotypic variance ( $\sigma^2_p$ ), and heritability estimates ( $h^2$ ) for carcass traits of Hanwoo cattle.

Trait	$\sigma^2_a$	$\sigma^2_e$	$\sigma^2_p$	$h^2$
CWT	773.00	1266.48	2039.48	0.38
MS	1.29	1.67	2.96	0.44
BFT	7.77	14.04	21.81	0.36
EMA	43.80	81.89	125.69	0.35

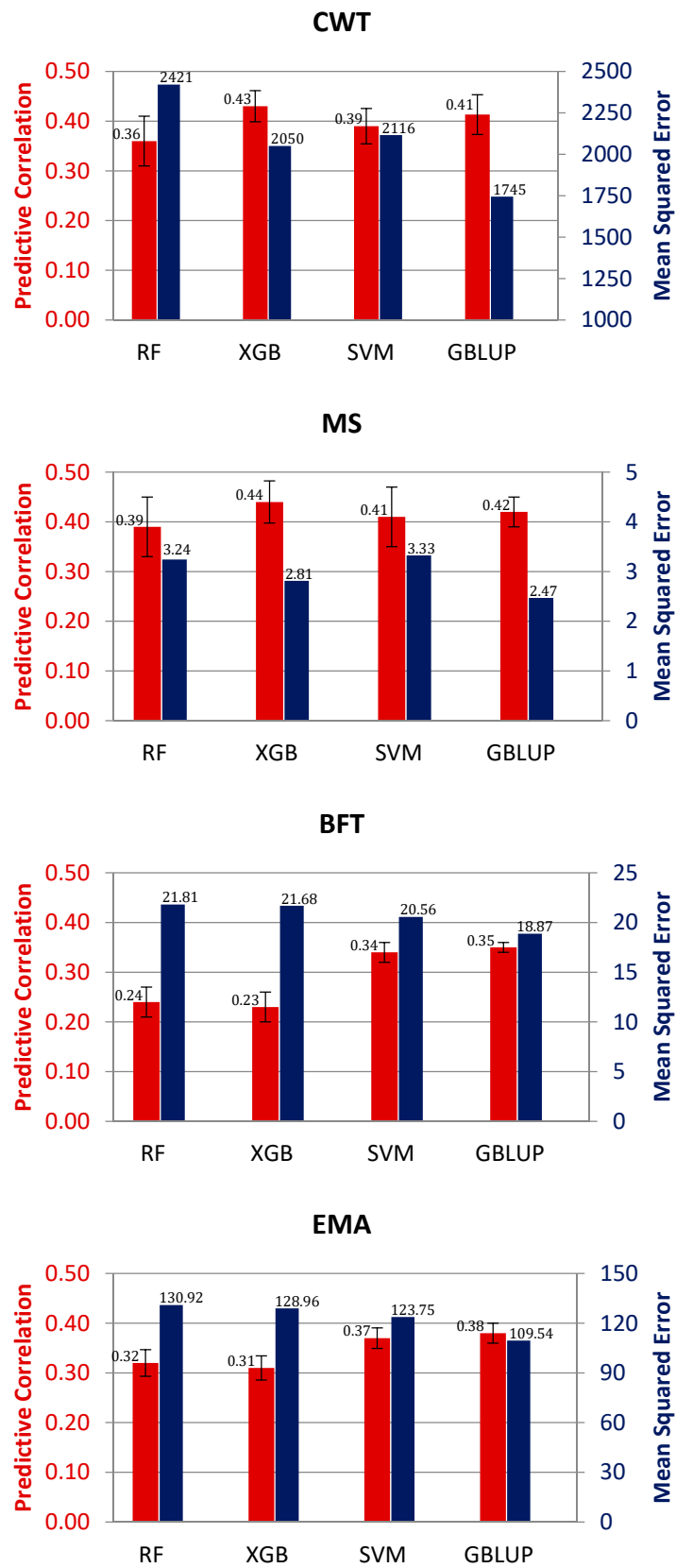
CWT, carcass weight; MS, marbling score; BFT, backfat thickness; EMA, eye muscle area.

### 3.2. Genomic Prediction

Figure 1 displays the predictive correlation and mean squared error (MSE) using GBLUP and three machine learning methods, namely, RF, XGB, and SVM, for four carcass traits of Hanwoo cattle. The results showed that the boosting method XGB showed the highest predictive correlation for CWT and MS, followed by GBLUP, SVM, and RF. Meanwhile, the largest predictive correlation for BFT and EMA was delivered by GBLUP, followed by SVM, RF, and XGB. The average correlations using GBLUP for CWT, MS, BFT, and EMA were 0.41, 0.42, 0.35, and 0.38, while they were 0.43, 0.44, 0.23, and 0.31 when using XGB, respectively. Correspondingly, the mean correlations using SVM were 0.39, 0.34, 0.42, and 0.37, and those using RF were 0.36, 0.39, 0.24, and 0.32, respectively. Predictive correlation is a common and simple way of measuring predictive performance, but MSE is a preferred parameter because it takes into account both prediction bias and variance. In this sense, GBLUP delivered the lowest MSE for all traits among methods. Meanwhile, among the machine learning methods, the lowest MSE for CWT and MS was achieved with XGB, whereas the best performer was SVM for BFT and EMA.

The random forest method has been used in many genomic prediction studies. González-Recio and Forni [11] compared the RF method with Bayes A and Bayesian LASSO using simulated discrete data and disease resistance data in pigs. They reported that RF outperformed those methods, with better classification performance within and across datasets. In this study, the RF method delivered the lowest predictive ability among the methods in general. This is consistent with the results of Abdollahi-Arpanahi et al. [28], who reported that GBLUP and Bayes B had a higher predictive correlation and lower MSE value than RF using a real dataset of Holstein bulls with sire conception rate records, genotyped for 58k SNPs. Moreover, Ogutu et al. [29] reported that RR-BLUP, boosting, and SVM methods had higher predictive correlations than RF in a study using simulation.

The SVM method is a popular machine learning algorithm used in genome-enabled prediction due to its capability to handle potential nonlinearity between features and target traits in both animals and plants [17,30,31]. Previous studies have shown contrasting results regarding the predictive performance of SVM over linear models [17,29,32,33]. In this study, the predictive correlation of the SVM model ranked second in two traits, and the difference in performance with the GBLUP model was small for all traits (Figure 1). Zhao et al. [17] compared the predictive ability of SVM, GBLUP, and BayesR methods using pig datasets. They reported that the prediction accuracy was very similar among methods. Meanwhile, Tusell et al. [33] showed that the SVM models could outperform the conventional GBLUP in predicting average residual feed intake and average daily gain crossbred performances from purebred sire genotypes.



**Figure 1.** Predictive correlation (red color) and mean squared error (blue color) of prediction obtained using different statistical methods for carcass weight (CWT), marbling score (MS), backfat thickness (BFT), and eye muscle area (EMA). RF: random forest, XGB: extreme gradient boosting, SVM: support vector machine, GBLUP: genomic best linear unbiased prediction.

Among machine learning methods, only the boosting method XGB outperformed GBLUP for some traits (CWT and MS) in terms of predictive correlation, as shown in Figure 1. Previous studies showed that the boosting method had a better predictive performance than other machine learning methods such as RF, SVM, and convolutional neural networks [28,29]. This could be due to its efficient ‘weak learner’ algorithm and stepwise assembling method with sequential learning to build the model, unlike parallel learning in the case of RF (bagging method). Another potential reason for such a better predictive performance could be that the boosting method trees are constructed following a greedy search algorithm or optimizing an objective function (e.g., ranking and Poisson regression), whereas RF constructs trees independently, using random samples of data.

Among machine learning methods, there is no universal prediction model. Predictive ability depends on the trait and is affected by many factors. In the machine learning field, the “no free lunch theorem” [34] states that there is no algorithm uniformly better for all species and traits that will work optimally for each problem each time. Thus, the best method may be case-dependent, and an initial evaluation of different methods is recommended to deal with a particular problem.

#### 4. Conclusions

Our results indicated that machine learning method XGB had the best predictive correlation for CWT and MS. Meanwhile, the highest predictive correlation for BFT and EMA was achieved by GBLUP. Although XGB presented higher predictive correlations for some traits, we did not find an advantage of XGB or any other machine learning method over GBLUP in terms of mean squared error of prediction. Therefore, we still recommend the use of the conventional statistical method GBLUP in the prediction of genomic breeding values for carcass traits in Hanwoo cattle.

**Author Contributions:** Conceptualization, S.S., B.I.L., J.-E.P. and D.L.; methodology, S.S. and B.I.L.; formal analysis, S.S. and B.I.L.; data curation, D.L.; writing—original draft preparation, S.S. and B.I.L.; writing—review and editing, H.K., M.J., J.-E.P., W.P., H.-H.C. and D.L.; visualization, D.L. and J.-E.P.; supervision, D.L. and J.-E.P.; project administration, D.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the AGENDA project (PJ01316901 and PJ015658) and the 2021 RDA Research Associate Fellowship Program of the National Institute of Animal Science, Rural Development Administration, Republic of Korea.

**Institutional Review Board Statement:** The ethics approval for this study was given by the Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration, Korea (2018-293).

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We acknowledge the Korean Institute of Animal Products Quality Evaluation (KAPE) for providing the phenotypic data.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Lee, S.-H.; Park, B.-H.; Sharma, A.; Dang, C.-G.; Lee, S.-S.; Choi, T.-J.; Choy, Y.-H.; Kim, H.-C.; Jeon, K.-J.; Kim, S.-D.; et al. Hanwoo cattle: Origin, domestication, breeding strategies and genomic selection. *J. Anim. Sci. Technol.* **2015**, *56*, 2. [[CrossRef](#)]
2. Kim, S.; Alam, M.; Park, M.N. Breeding initiatives for Hanwoo cattle to thrive as a beef industry—A review study. *J. Anim. Breed. Genom.* **2017**, *1*, 102–124. [[CrossRef](#)]
3. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
4. Aguilar, I.; Misztal, I.; Johnson, D.L.; Legarra, A.; Tsuruta, S.; Lawlor, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. *J. Dairy Sci.* **2010**, *93*, 743–752. [[CrossRef](#)]
5. Meuwissen, T.H.; Hayes, B.J.; Goddard, M. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]
6. de los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A.; Manfredi, E.; Weigel, K.; Cotes, J.M. Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. *Genetics* **2009**, *182*, 375–385. [[CrossRef](#)] [[PubMed](#)]

7. Moser, G.; Lee, S.H.; Hayes, B.J.; Goddard, M.E.; Wray, N.R.; Visscher, P.M. Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLoS Genet.* **2015**, *11*, e1004969. [[CrossRef](#)]
8. Endelman, J.B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **2011**, *4*. [[CrossRef](#)]
9. van Eeuwijk, F.A.; Bink, M.C.A.M.; Chenu, K.; Chapman, S.C. Detection and use of QTL for complex traits in multiple environments. *Curr. Opin. Plant Biol.* **2010**, *13*, 193–205. [[CrossRef](#)]
10. de los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* **2013**, *193*, 327–345. [[CrossRef](#)]
11. González-Recio, O.; Forni, S. Genome-wide prediction of discrete traits using bayesian regressions and machine learning. *Genet. Sel. Evol.* **2011**, *43*, 7. [[CrossRef](#)]
12. González-Recio, O.; Rosa, G.J.M.; Gianola, D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* **2014**, *166*, 217–231. [[CrossRef](#)]
13. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
14. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, 1189–1232.
15. Aruna, S.; Dr, S. A Novel SVM based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer. *Int. J. Comput. Appl.* **2011**, *31*, 14–20.
16. Chen, T.; Guestrin, C. XgBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
17. Zhao, W.; Lai, X.; Liu, D.; Zhang, Z.; Ma, P.; Wang, Q.; Zhang, Z.; Pan, Y. Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. *Front. Genet.* **2020**, *11*, 598318. [[CrossRef](#)]
18. Li, B.; Zhang, N.; Wang, Y.-G.; George, A.W.; Reverter, A.; Li, Y. Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Front. Genet.* **2018**, *9*, 237. [[CrossRef](#)] [[PubMed](#)]
19. Misztal, I.; Tsuruta, S.; Lourenco, D.; Aguilar, I.; Legarra, A.; Vitezica, Z. Manual for BLUPF90 family of programs. *Athens Univ. GA* **2014**, 199.
20. Lopez, B.I.; Lee, S.-H.; Park, J.-E.; Shin, D.-H.; Oh, J.-D.; de las Heras-Saldana, S.; van der Werf, J.; Chai, H.-H.; Park, W.; Lim, D. Weighted Genomic Best Linear Unbiased Prediction for Carcass Traits in Hanwoo Cattle. *Genes* **2019**, *10*, 1019. [[CrossRef](#)]
21. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
22. Lee, S.H.; Van der Werf, J.H.J. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* **2016**, *32*, 1420–1422. [[CrossRef](#)] [[PubMed](#)]
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
24. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
25. Choi, T.J.; Alam, M.; Cho, C.I.; Lee, J.G.; Park, B.; Kim, S.; Koo, Y.; Roh, S.H. Genetic parameters for yearling weight, carcass traits, and primal-cut yields of Hanwoo cattle1. *J. Anim. Sci.* **2015**, *93*, 1511–1521. [[CrossRef](#)]
26. Do, C.; Park, B.; Kim, S.; Choi, T.; Yang, B.; Park, S.; Song, H. Genetic Parameter Estimates of Carcass Traits under National Scale Breeding Scheme for Beef Cattle. *Asian-Australas. J. Anim. Sci.* **2016**, *29*, 1083–1094. [[CrossRef](#)] [[PubMed](#)]
27. Lopez, B.I.M.; An, N.; Srikanth, K.; Lee, S.; Oh, J.-D.; Shin, D.-H.; Park, W.; Chai, H.-H.; Park, J.-E.; Lim, D. Genomic Prediction Based on SNP Functional Annotation Using Imputed Whole-Genome Sequence Data in Korean Hanwoo Cattle. *Front. Genet.* **2021**, *11*, 603822. [[CrossRef](#)] [[PubMed](#)]
28. Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Sel. Evol.* **2020**, *52*, 12. [[CrossRef](#)] [[PubMed](#)]
29. Ogutu, J.O.; Piepho, H.-P.; Schulz-Streeck, T. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* **2011**, *5*, S11. [[CrossRef](#)] [[PubMed](#)]
30. Moser, G.; Tier, B.; Crump, R.E.; Khatkar, M.S.; Raadsma, H.W. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet. Sel. Evol.* **2009**, *41*, 56. [[CrossRef](#)]
31. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Application of support vector regression to genome-assisted prediction of quantitative traits. *Theor. Appl. Genet.* **2011**, *123*, 1065. [[CrossRef](#)]
32. Alves, A.A.C.; da Costa, R.M.; Bresolin, T.; Fernandes Júnior, G.A.; Espigolan, R.; Ribeiro, A.M.F.; Carvalheiro, R.; de Albuquerque, L.G. Genome-wide prediction for complex traits under the presence of dominance effects in simulated populations using GBLUP and machine learning methods. *J. Anim. Sci.* **2020**, *98*, skaa179. [[CrossRef](#)] [[PubMed](#)]
33. Tusell, L.; Bergsma, R.; Gilbert, H.; Gianola, D.; Piles, M. Machine Learning Prediction of Crossbred Pig Feed Efficiency and Growth Rate From Single Nucleotide Polymorphisms. *Front. Genet.* **2020**, *11*, 567818. [[CrossRef](#)] [[PubMed](#)]
34. Wolpert, D.H. The lack of a priori distinctions between learning algorithms. *Neural Comput.* **1996**, *8*, 1341–1390. [[CrossRef](#)]