# Comparative Study of Statistical Models for Genomic Prediction

**Sayanti Guha Majumdar, Anil Rai and Dwijesh Chandra Mishra**

*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

## SUMMARY

Genomic prediction has been used for breeding of animals and plants with complex quantitative traits by predicting Genomic Estimated Breeding Values (GEBVs) of target population. The accuracy of genomic prediction depends on various factors including sampling population, genetic architecture of target species, statistical models, etc. There are large numbers of statistical models for genomic prediction available in the literature. These models perform differently due to different genetic architecture of the datasets. In this article, performances of linear least squared regression, BLUP, LASSO, ridge regression, SpAM, HSIC LASSO, SVM, ANN along with our newly developed integrated model framework have been evaluated in wheat dataset containing 599 wheat lines and 1279 SNP markers. In general, the performances of SVM, ridge regression and integrated model framework were found to be superior for genomic prediction. This study will help researcher in selection of appropriate statistical method to predict phenotypic values.

*Keywords*: ANN, Genomic prediction, Integrated model framework, LASSO, Ridge regression, SVM.

## 1. INTRODUCTION

Genomic prediction or genomic selection (GS) is an emerging field of genomic-assisted breeding methodology where whole genome marker data is used to predict genomic estimated breeding value (GEBV). The aim of this method is to increase genetic gain by shortening breeding cycles and increasing accuracy of prediction of GEBV. There are several statistical models available in literature for genomic prediction, viz. least squared regression (LSR), best linear unbiased prediction (BLUP) (Henderson, 1975), least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), ridge regression (Hoerl and Kennard, 1970), Sparse Additive Models (SpAM) (Ravikumar *et al.*, 2009), Hilbert-Schmidt Independence Criterion LASSO (HSIC LASSO) (Gretton *et al.*, 2005 and Yamada *et al.*, 2014), support vector machine (SVM) (Vapnik, 1995), artificial neural network (ANN) (Bain, 1873 and James, 1890). Another statistical model framework has been developed in our previous study by combining one additive model, i.e. SpAM and one non-additive model, i.e. HSIC LASSO. The newly

developed model can be mentioned as integrated model framework (Guha Majumdar *et al.*, 2019). The accuracy of prediction of different models varies on the basis of the underlying statistical methods. All the models differ among themselves due to assumption about the distribution, variance among the genetic markers used etc. In our present study, we have compared the performance of all the above mentioned models for genomic prediction in case of wheat. These models are described briefly below.

### 1.1 Linear Least-Squares Regression

In GS, the main goal is to predict the individual's breeding value by modeling the relationship between the individual's genotype and phenotype. The Linear least-square regression is the simplest model, which can be written as

$$y_i = \mu + \sum_{j=1}^{p} X_{ij}\beta_j + e_i$$

where, $i = 1 \ldots n$ individual, $j = 1 \ldots p$ marker position/segment, $y_i$ is the phenotypic value for

---

*Corresponding author*: Sayanti Guha Majumdar
*E-mail address*: sayanti23gm@gmail.com

individual $i$, $\mu$ is the overall mean, $X_{ij}$ is an element of the incidence matrix corresponding to marker j, individual $i$, $\beta_j$ is a random effect associated with marker $j$, and $e_i$ is a random residual which follows $N(0, \sigma_e^2)$.

The basic problem in using this model is that, it does not work well, if the available number of markers (explanatory variables) is greater than the number of individual available (observations). In order to overcome this problem Meuwissen *et al.* (2001) adopted a stepwise procedure of least squares regression for GS. First, least squares regression analysis was performed on each segment (marker) separately using above model. Then the likelihood of every segment was plotted against the position of the segment which helped in identifying the segments having significant effects. Finally, segments having significant effects were used simultaneously by the model to estimate their individual effects. But this approach also has some drawbacks, like it does not fully take advantage of the whole genome marker information as markers with a significant effect are included in the final model.

## 1.2 Best Linear Unbiased Prediction (BLUP)

The BLUP theory and the mixed model formulation were first described by Henderson (1949), and BLUP was recommended as a method of GS by Meuwissen *et al.* (2001). The random effects model of BLUP (Henderson, 1975)can be written as

$$y = \mu + \sum_{j=1}^{p} Z_j \beta_j + e$$

where, $y$ is the $(n \times 1)$ vector of phenotypic data, $\mu$ is the $(n \times 1)$ overall mean vector, $Z_j$ is the $j$th column of the design matrix, $\beta_j$ is the genetic effect associated with the $j$th marker, and $p$ is the number of markers. $\mu$ is the intercept which is fixed, and $\beta_j$ is the random effects with $E(\beta_j) = 0$, $Var(\beta_j) = \sigma_{\beta_j}^2$, $Var(e) = \sigma^2 I$ and $Cov(\beta, e) = 0$. The intercept $\mu$ can be replaced by $X\alpha$ to include all the fixed effects if other covariates are also available. Then, the model can be written as

$$y = X\alpha + Z\beta + e$$

where $\alpha$ is a $(p_1 \times 1)$ vector of unknown fixed effects, where, the first element is considered as the population mean, and $X$ is the incidence matrix which relates $y$ to $\alpha$. The above equation is usually known as mixed model or mixed effects model. The fixed effect vector $\alpha$ is estimated by BLUE, whereas, BLUP is the predictor of the random effects.

Henderson (1953) proposed that $(\alpha, \beta)$ can be obtained by maximizing the joint likelihood of $(y, \beta)$ given by:

$$L(y, \beta) = f(y|\beta) f(\beta)$$
$$= \frac{1}{(2\pi)^{n/2}|R|^{1/2}} \left[ -\frac{1}{2}(y - X\alpha - Z\beta)' R^{-1}(y - X\alpha - Z\beta) \right]$$
$$\times \frac{1}{(2\pi)^{p/2}|G|^{1/2}} \left[ -\frac{1}{2}\beta' G^{-1}\beta \right]$$

A set of linear equations [Henderson's Mixed Model Equations (MME)] can be obtained by maximizing the likelihood $L(y, \beta)$ with respect to $\alpha$, $\beta$ and equating it to zero:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

where, $R = Var(e)$ and $G = Var(\beta)$. The BLUE of $\alpha$ and the BLUP of $\beta$ can be obtained by solving the MME. The assumption of Henderson's derivation is that $\beta$ and $e$ are normally distributed and maximizes the joint likelihood of $(y, \beta)$ over the unknowns $\alpha$ and $\beta$.

## 1.3 Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO technique (Tibshirani, 1996) is being used for efficient feature selection based on the assumption of linear dependency between input features and output values. In case of LASSO, optimization problem is given as

$$\min_{\beta \in \mathbb{R}^d} \frac{1}{2} \| y - X^T \beta \|_2^2 - \lambda \| \beta \|_1$$

where, $\beta = [\beta_1, ..., \beta_p]^T$ is a regression coefficient vector, $\beta_k$ denotes the regression coefficient of the k-th feature, $\| . \|_1$ and $\| . \|_2$ are the $\ell_1$ and $\ell_2$-norms, and $\lambda > 0$ is the regularization parameter. The $\ell_1$-regularizer in LASSO tends to produce a sparse solution, which means that the regression coefficients of non-significant features become zero.

LASSO is specifically suitable, when the number of features is larger than the number of training samples (Tibshirani, 1996). So, by using LASSO we can overcome the limitations of linear least-square regression method. However, this method performs well for additive effect data only.

## 1.4 Ridge Regression

In case of multi-collinear marker data, the performance of variable selection methods is generally very poor. In order to address this problem, we can use penalized regression model i.e. ridge regression of Hoerl and Kennard 1970. Ridge regression minimizes the penalized sum of squares:

$$|y - X\beta|^2 + \lambda^2 \beta' \beta$$

where, $\lambda$ is the penalty parameter, and the estimate of the regression coefficient is given by: $\widehat{\beta} = (X'X + \lambda I)^{-1} X' y$.

where, $I$ is a $p \times p$ identity matrix. The penalty parameter $\lambda$ can be calculated by several different methods, for example, by plotting $\widehat{\beta}$ as a function of $\lambda$ and choosing the smallest $\lambda$ that results in a stable estimate of $\widehat{\beta}$. Hoerl *et al.* (1975) have proposed another way to choose $\lambda$ using an automated procedure. The estimate of $\lambda$ is given by:

$$\lambda = \frac{rs^2}{(\widehat{\beta})'(\widehat{\beta})}$$

where, $r$ is the number of parameters in the model except the intercept, $s^2$ is the residual mean square obtained by linear least squares estimation, and $\widehat{\beta}$ is the vector of least squares estimates of regression coefficients.

Ridge regression estimator of $\beta$ is biased and this increase in bias is compensated by the decrease in variance. As a result, we get an estimator $\widehat{\beta}_R$ with smallest MSE. Another advantage of ridge regression is that it can be used when available markers are more than the sample size to overcome the "$p > n$" problem.

Meuwissen *et al.* (2001) employed ridge regression in GS. It was assumed that the marker effects were random, and they were drawn from a normal distribution with $Var(\beta_j) = \sigma_\beta^2$, whereas, additive genetic variance among individuals is expressed as

$\sigma_\beta^2 = \sigma_a^2 / n_k$, where, $\sigma_a^2$ represents additive genetic variance among individuals and $n_k$ is the number of marker loci (Habier *et al.* 2007). This method is suitable for data with additive effects, i.e. for linear features only.

## 1.5 Sparse Additive Models (SpAM)

High dimensional feature selection can be performed with sparse additive models (SpAM) (Ravikumar *et al.*, 2009). The SpAM optimization model can be defined as

$$\min_{\beta_1, \dots, \beta_p \in \mathbb{R}^n} \left\| y - \sum_{k=1}^{p} K^{(k)} \beta_k \right\|_2^2 + \lambda \sum_{k=1}^{p} \sqrt{\frac{1}{n} \| K^{(k)} \beta_k \|_2^2}$$

where, $\beta_k = [\beta_{k,1}, \dots, \beta_{k,n}]^T, k = 1, \dots, p$ are regression coefficient vectors, $K_{i,j}^{(k)} = K(x_{k,i}, x_{k,j})$ is Gram matrix and $i, j = 1, \dots, n$, $K(x, x')$ is kernel function, $\beta_{k,j}$ is a coefficient for $[K(x_{k,1}, x_{k,j}), \dots, K(x_{k,n}, x_{k,j})]^T$ and $\lambda > 0$ is a regularization parameter. SpAM is considered as a convex method which can be efficiently optimized by the back-fitting algorithm.

A disadvantage of SpAM is that it can only deal with additive effects. In case of epistatic effects in the data SpAM may fail to select significant markers. Also, SpAM is computationally expensive procedure.

## 1.6 HSIC LASSO

A kernalized non-linear LASSO was proposed by Yamada *et al.* (2014), which isalso called as HSIC (Hilbert-Schmidt Independence Criterion, Gretton *et al.*, 2005) LASSO. The optimization problem can be expressed as

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \| \bar{L} - \sum_{k=1}^{p} \beta_k \bar{K}^{(k)} \|_{Frob}^2 + \lambda \| \beta \|_1$$

s.t. $\beta_1, \dots, \beta_p \geq 0$,

where, $\| \cdot \|_{Frob}$ is the Frobenius norm, $\bar{K}^{(k)} = \Gamma K^{(k)} \Gamma$ and $\bar{L} = \Gamma L \Gamma$ are centered Gram matrices, $K_{i,j}^{(k)} = K(x_{k,i}, x_{k,j})$ and $L_{i, j} = L(y_i, y_j)$ are Gram matrices, $K(x, x')$ and $L(y, y')$ are kernel functions, $\Gamma = I_n - \frac{1}{2} 1_n 1_n^T$ is the centering matrix, $I_n$ is the n-dimensional identity matrix, and $1_n$ is the

n-dimensional vector with all ones. A non-negativity constraint is employed in this model so that meaningful features are selected. This model differs from the original formulation of LASSO as in this case kernel functions $K$ and $L$ is different and non-negativity constraint is imposed. The first term in this equation denotes that we are regressing the output kernel matrix $\bar{L}$ by a linear combination of feature-wise input kernel matrices $\{\bar{K}^{(k)}\}_{k=1}^{p}$.

## 1.7 Support Vector Machine (SVM)

SVM was proposed by Vapnik in 1995. SVM is a supervised machine learning technique, originally used as a classifier. To train the SVM, training dataset is used and the classifier produces maximum margin separation between two classes of observations. The idea can be used for estimating unknown regression function. Maenhout *et al.* (2007) and Long *et al.* (2011) have implemented SVM regression for GS in plant breeding. SVM regression can model the relationship between the marker genotypes and the phenotypes with a linear as well as nonlinear mapping function.

Let us consider a training sample $S = \{(x_i, y_i), x_i \in R^n, y_i \in R, i = 1 \dots n\}$, where $x_i$ is a vector of genotypic values of the $p$ markers for individual $i$, and $y_i$ is the vector of phenotypic value for individual $i$. SVM model describing the relationship between the phenotype and the genotype of an individual has been described below:

$$f(x) = b + wx$$

Where $b$ is a constant which reflects the maximum error while estimating $w$ and $w$ is a vector of unknown weights. The function $f(x)$ can be obtained by minimizing the expression

$$\lambda \sum_{i=1}^{n} L(y_i - f(x_i)) + \frac{1}{2}\|w\|^2$$

Where $L(.)$ denotes the loss function which may be squared loss function, absolute loss function or $\varepsilon$-insensitive loss function measuring the quality of the estimation. $\lambda$ is the regularization parameter which is responsible for the trade-off between the sparsity and the complexity of the model. The norm $\|w\|$ of vector $w$ is associated with model complexity inversely. A support vector $x_i$ satisfies the equation $y_i(wx_i + b) = 1$ by definition.

## 1.8 Artificial Neural networks (ANN)

Neural network (NN) is a nonparametric statistical method which can model relationship between genotypes and phenotype with both linear and complex nonlinear functions. NN mimics the idea of how neurons in the human brain work and interact, and conducts computations. NN was first introduced by Bain (1873) and James (1890). Every unit in NN is analogous to a brain neuron and they connect among themselves with several functions which are analogous to synapses (Hastie *et al.* 2009). The NN is composed of three types of layers, viz. an input layer, a hidden layer and an output layer. This model is known as the feed-forward NN. The NN which is used to estimate a regression function usually consists of only one output layer unit. The hidden layer units are functions of linear combinations of the inputs, whereas, the output layer units are functions of the hidden layer units. The output function of a feed-forward NN can be defined as:

$$f(I_k) = \beta_0 + \sum_{l=1}^{L} \beta_l \sigma(w_l, b_l, I_k), k = 1, 2, \dots, K$$

where $K$ is the number of units in the input layer, $M$ is the number of output layer units, $L$ is the number of hidden layer units, $I_k$ is the $k^{\text{th}}$ input, $\beta_0 \in R^M$ is the intercept, $\beta_l(l = 1, 2, \dots, L)$ are the output layer weights connecting the $l^{\text{th}}$ hidden layer unit to the output layer units, $\sigma$ is the activation function modeling the connection between the hidden layer and the output layer, and $w_l \in R^K$ and $b_l \in R$ are the unknown learning parameters of the hidden layer unit $l(l = 1, 2, \dots, L)$ connecting the $k^{\text{th}}$ neuron in the input layer.

In GS, marker genotypes are represented by $I_k$ where $K$ is the number of individuals in the analysis. We can choose the activation function $\sigma$ as the sigmoid or the Gaussian radial basis function. Gianola *et al.* implemented NNs for GS in 2011.

## 1.9 Integrated Model Framework

The integrated model framework (Guha Majumdar *et al.*, 2019) for estimation of GEBV has been developed by combining SpAM and HSIC LASSO and can be used to capture both linear and non-linear effect of the genetic markers on the phenotypic data. The model can be expressed as

$$y_{Int} = wy_{Sp} + (1-w)y_{HL}$$

where, $y_{Int}$ is the predicted phenotype of the integrated model framework, $w$ is $\frac{\sigma_{HL}^2}{\sigma_{Sp}^2 + \sigma_{HL}^2}$, where $\sigma_{HL}^2$ and $\sigma_{Sp}^2$ are the error variances of models HSIC LASSO and SpAM respectively, $y_{Sp}$ is the predicted phenotype from Sparse Additive Models and $y_{HL}$ is the predicted phenotype from HSIC LASSO. The estimation of $\sigma_{Sp}^2$ and $\sigma_{HL}^2$ can be performed by following refitted cross validation approach of Fan *et al.*, 2012.

## 2. MATERIALS AND METHODS

### 2.1 Data Description

We have a real dataset of wheat for implementing genomic prediction models (Crossa *et al.*, 2010). Genotyping of wheat was done by using 1447 Diversity Array Technology markers generated by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com. au). This dataset includes 599 lines observed for trait grain yield (GY) for four mega environments. For the convenience of our study the GY for first mega environment has been considered. The final number of DArT markers in the dataset after editing was 1279 which has been used in this study.

After implementation of statistical models prediction accuracy (PA) and mean squared error (MSE) have been estimated for all the models. PA can be defined as the correlation between the actual phenotypic $(y_{actual})$ values and the predicted phenotypic $(y_{pred})$ values (Howard *et al.*, 2014). MSE can be expressed as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Where, $\hat{y}$ is the predicted value of phenotype, $y$ is the actual value of phenotype, and $n$ is the number of individuals in the dataset. In order to implement the statistical models, the statistical software R was used. Before implementing the models, the dataset has been split into training and testing data. 80% of the observation has been chosen randomly for training purpose and rest 20% data has been kept as testing data. This procedure of splitting was repeated 500 times to get 500 training and testing datasets which were used to implement all the statistical models.

### 2.2 Implementation in R

#### Least squares regression (LSR)

In order to implement least square regression, the *lm* function from the *stats* package (R Development Core Team 2019) in R was used. In first step, simple linear regression model was fitted for each of the individual markers and most significant 100 markers are selected according to their p-values. Then in a final model those 100 markers were included to simultaneously fit a linear regression model. This two-step method was applied as there are more number of markers than the number of individuals. Finally, phenotypic values were predicted from testing dataset with the selected marker data and estimated regression coefficient of marker effects.

#### BLUP

BLUP was implemented by using *mixed.solve* function from *rrBLUP* package (Endelman, 2011) in R. The model was fitted using training data. Then the phenotypic value was predicted using testing dataset and the predicted coefficients of marker effects from the fitted model.

#### LASSO

The *glmnet* function of the *glmnet* package (Friedman *et al.*, 2010) in R was used with default parameter values to implement LASSO. The prediction was performed with the help of *predict* function of the same package by minimizing cross-validation error.

#### Ridge Regression

Ridge regression can also be implemented through *glmnet* function of the *glmnet* package (Friedman *et al.*, 2010) in R by setting the value of alpha equal to zero. Then the prediction was performed in testing set by using *predict* function.

#### SpAM

The sparse additive model was implemented by using *samQL* function of the *SAM* package (Zhao *et al.*, 2014) in R with default parameter values. The *predict* function of the same package was used to perform the prediction of phenotypic value in testing dataset.

#### HSIC LASSO

In-house R function was developed to implement HSIC LASSO or kernalized LASSO. The *penalized* function of *penalized* package (Goeman *et al.*, 2010)

has been used to fit this kernelized LASSO model. Then *predict* function of the same package is used to predict the phenotypic value of the testing dataset.

## SVM

The *ksvm* function of the *kernlab* package (Karatzoglou *et al.*, 2004) in R with the default parameters was used to perform SVM regression on the training dataset. After fitting the model, the *predict* function was used to obtain the predicted phenotypic values for the testing set.

## Neural network

The NN model was implemented using the *brnn* function of the *brnn* package (Rodriguez and Gianola, 2013) in R. This function uses a two layer NN and maps the input information into some basis function. The number of neurons was set to be three and the number of epochs to train the model was 30. The *predict.brnn* function of the same package was used in the next step to predict the phenotype using testing dataset.

## Integrated Model Framework

The *GSelection* package (Guha Majumdar *et al.*, 2019) in R was developed by us to implement the integrated model framework in GS. To fit the model in training data *feature.selection* function was used. The error variances of SpAM and HSIC LASSO model were estimated with *spam.var.rcv* and *hsic.var.rcv* function of the same package. Then the prediction of the phenotypic value in testing dataset was performed by using *genomic.prediction* function.

## 3. RESULTS AND DISCUSSION

In this study, various statistical models have been implemented in real dataset of wheat for genomic prediction. LSR, BLUP, LASSO, Ridge regression, SpAM, HSIC LASSO, SVM, ANN, Integrated model framework are compared on the basis of their performance in genomic prediction of breeding values. The results have been shown in Table 1.

**Table 1. Statistical comparison of various models for genomic prediction**

| Models | Prediction Accuracy (PA) | Standard Error of PA | MSE |
|---|---|---|---|
| LSR | 0.0476 | 0.0142 | 2.5199 |
| BLUP | 0.1941 | 0.0076 | 2.0420 |
| LASSO | 0.4299 | 0.0070 | 1.8036 |
| Ridge | 0.5253 | 0.0058 | 1.1740 |
| SpAM | 0.4941 | 0.0056 | 1.4436 |
| HSIC LASSO | 0.1490 | 0.0023 | 0.0730 |
| SVM | 0.5784 | 0.0053 | 1.1667 |
| ANN | 0.4822 | 0.0065 | 1.4194 |
| Integrated Model Framework | 0.4950 | 0.0056 | 1.3211 |

It is evident from Table 1 that the newly developed integrated model has been performed better than LSR, BLUP, LASSO, SpAM, HSIC LASSO and ANN in terms of prediction accuracy. Only one parametric statistical model, i.e. ridge regression and one non-parametric statistical model i.e. SVM has better prediction accuracy than the integrated model. Also the mean square error is less in case of SVM and ridge
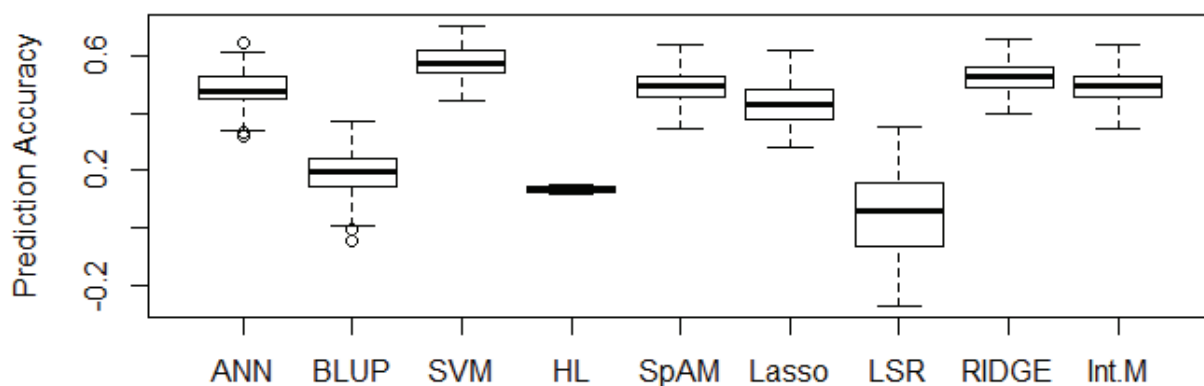


**Fig. 1. The boxplots of prediction accuracy corresponding to different statistical models**

regression than integrated model. Another observation is that the mean square error of HSIC LASSO is very less than the other models. This is due to the reason that HSIC LASSO is nonlinear parametric model where MSE may not be highly desirable criteria for evaluation of model performance. The prediction accuracy of different models are shown with the help of a boxplot in Fig. 1.

It is known from the literature that genetic architecture is responsible for the differences of the accurate predictions of breeding value among the GS methods. Genetic architecture of a population depends on the presence of additive and epistatic genetic effects. The parametric models assume that the markers are independent, i.e. additive in nature. But in practical situation, both additive and epistatic effects are present in the genetic architecture of the population. Because of that reason the parametric models, viz. LSR, BLUP do not perform well in genomic prediction. Although, ridge regression, which is a biased parametric estimator, performs very well in this study. It is also observed that the non-parametric models (viz. SVM, ANN) and newly developed integrated model perform very well in genomic prediction of breeding value. This is due to the reason that these models can capture both linear (additive) and non-linear (epistatic) effect of the markers in the dataset.

## 4. CONCLUSION

The performances of various statistical models in case of genomic prediction have been compared in the present study. The study is conducted in the real dataset of wheat. So, this article will give a clear idea about several statistical models on how they behave in practical situation of genomic prediction. This will help in choosing the appropriate model for our dataset. The superiority of models like ridge regression, SVM, integrated model framework has been depicted in the above study. The accuracy of these models depends on several factors including the trait of interest, extent of additive and epistatic effect present in the dataset, heritability of the trait etc. The performances of these models can be improved further if we consider dominance effect and genotype by environment interaction in the study.

## REFERENCES

Bain, A. (1873). Mind and Body: The Theories of Their Relation. D. Appleton and Company, New York.

Crossa, J., de los Campos, G., Perez, P., Gianola, D., Burgueño, J., *et al.* (2010). Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers, *Genetics*, **186**, 713-724. https://doi.org/10.1534/genetics.110.118521

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP, *Plant Genome*, **4**, 250-255.

Fan, J., Guo, S. and Hao, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression, *J. Roy. Statist. Soc.*, **74(1)**, 37-65.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, **33**, 1-22. URL http://www.jstatsoft.org/v33/i01/

Gianola, D., Okut, H., Weigel, K.A. and Rosa, G.J.M.(2011).Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat, *BMC Genetics*, **12**, 87-100.

Goeman, J.J. (2010). L1 penalized estimation in the Cox proportional hazards model, *Biometrical Journal*, **52(1)**, 70-84.

Gretton, A., Bousquet, O., Smola, A. and Scholkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms, Algorithmic Learning Theory (ALT), pp 63-77.Springer.

Guha Majumdar, S., Rai, A. and Mishra, D.C.(2019).Integrated Framework for Selection of Additive and Non additive Genetic Markers for Genomic Selection, *Journal of Computational Biology*.http://doi.org/10.1089/cmb.2019.0223

Guha Majumdar, S., Rai, A. and Mishra, D.C.(2019).GSelection: Genomic Selection. R package version 0.1.0.https://CRAN.R-project.org/package=GSelection

Habier, D., Fernando, R.L. and Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values, *Genetics*, **177**, 2389-2397.

Hastie, T., R. Tibshirani, and J. Friedman.(2009).The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.

Henderson, C.R. (1949). Estimates of changes in herd environment. *Journal of Dairy Science*, **32**, 706.

Henderson, C.R. (1953). Estimation of Variance and Covariance Components, *Biometrics*, **9(2)**, 226-252.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model, *Biometrics*, **31(2)**, 423-447.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: biased estimation for non orthogonal problems, *Technometrics*, **12**, 55-67.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: applications to non orthogonal problems, *Technometrics*, **12**, 69-82.

Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975). Ridge Regression: Some Simulation, *Communications in Statistics: Theory and Methods*, **4(2)**, 105-123.

Howard, R., Carriquiry, A.L. and Beavis, W.D. (2014).Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures, *G3 (Bethesda)*, **4(6)**, 1027-46.

James, W. (1890).The Principles of Psychology. H. Holt and Company, NewYork.

Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A.(2004). kernlab - An S4 Package for Kernel Methods in R, *Journal of Statistical Software*, **11**, 1-20. URL http://www.jstatsoft.org/v11/i09/.

Long, N., Gianola, D., Rosa, G.J.M. and Weigel, K.A. (2011). Application of support vector regression to genome-assisted prediction of quantitative traits, *Theoretical and Applied Genetics,* **123**, 1065-1074.

Maenhout, S., Baets, B.D., Haesaert, G. and Bockstaele, E.V. (2007). Support vector machine regression for the prediction of maize hybrid performance, *Theoretical and applied genetics,* **115(7)**, 1003-1013. doi: 10.1007/s00122-007-0627-9.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001).Prediction of total genetic value using genome-wide dense marker maps, *Genetics*, **157**, 1819-1829.

Pérez-Rodriguez, P., and Gianola, D. (2013).brnn: brnn (Bayesian regularization for feed-forward neural networks), R package version 0.3. http://CRAN.R-project.org/package=brnn.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL https://www.R-project.org/.

Ravikumar, P., Lafferty, J., Liu, H. and Wasserman, L. (2009). Sparse additive models, *J. Roy. Statist. Soc., Series B (Statistical Methodology)*, **71(5)**, 1009-1030.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso, *Journal of Royal Statistical Society Series B (Methodological)*, **58,** 267-288.

Vapnik, V.(1995).The Nature of Statistical Learning Theory, Ed. 2.Springer, New York.

Yamada, M., Jitkrittum, W., Sigal, L., Xing, E.P. and Sugiyama, M. (2014). High-Dimensional Feature Selection by Feature-Wise Kernelized Lasso, *Neural Computation*, **26**, 185-207.

Zhao, T., Li, X., Liu, H. and Roeder, K. (2014). SAM: Sparse Additive Modelling, R package version 1.0.5. https://CRAN.R-project.org/package=SAM