Indian Journal of Animal Sciences **87** (10): 1226–1231, October 2017/Article

Performance evaluation of support vector machine (SVM)-based predictors in genomic selection

SEYED AMIR KASNAVI[1], MAHDI AMIN AFSHAR[2], MOHAMMAD MAHDI SHARIATI[3], NASSER EMAM JOMEH KASHAN[4] and MAHMOOD HONARVAR[5]

Islamic Azad University, Tehran, Iran

Received: 13 March 2017; Accepted: 21 April 2017

ABSTRACT

The aim was to compare predictive performance of SVM-based predictors constructed using different kernel functions (radial, sigmoid, linear and polynomial) in different genetic architectures of a trait (number of QTL, distribution of QTL effects) and heritability levels. To this end, a genome comprised of five chromosomes, one Morgan each, was simulated on which 10,000 bi-allelic single nucleotide polymorphisms (SNP) were distributed. Cross validation employing a grid search was used to tune the meta-parameters of each kernel function. Pearson's correlation between the true and predicted genomic breeding values ($r_{p,t}$) and mean squared error of predicted genomic breeding values (MSE_p) were used, respectively, as measures of the predictive accuracy and the overall fit. Meta-parameter optimization had a significant effect on predictive performance of SVM-based predictors in such a way that by using improper meta-parameters, the predictive power of models decreased significantly. In all models, the accuracy of prediction increased following increase in heritability and decrease in the number of QTLs. In most of scenarios, radial- and sigmoid-based SVM predictors outperformed polynomial and linear models. The linear-and polynomial-based SVM had lower $r_{p,t}$ and higher MSE_p and, therefore, were not recommended for genomic selection. The prediction accuracy of radial and sigmoid models was approximately the same in most of the studied scenarios; however, considering all pros and cons of radial and sigmoid kernels, radial kernel was recommended as the best kernel function for constructing SVM. All of studied SVM-based predictors were efficient users of time and memory.

Key words: Genetic architecture, Genomic breeding values, QTL effects, SNP, Support vector machine

The genomic selection (GS) introduced by Meuwissen *et al.* (2001) is an advanced form of marker assisted selection (MAS) in which, selection decisions are made on genomic breeding values, predicted from thousands of single nucleotide polymorphism (SNP). In the past 15 years, different statistical models have been developed for accurately predicting genomic breeding values (Howard *et al*. 2014, Hayes and Deatwyler 2015). Recently, nonparametric models from the machine-learning repository such as support vector machines (SVM) have been proposed for genomic prediction (Neves *et al.* 2012, Honarvar and Ghiasi 2013, Howard *et al.* 2014, Ghafouri-Kesbi *et al.* 2016). The SVM is a state-of-the-art classification method introduced by Boser *et al.* (1992) which is widely used in bioinformatics (and other disciplines) owing to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data (Scholkopf *et al.* 2004). To construct SVM, kernel functions are used. A kernel is just a transformation of input data that allows an algorithm like SVMs to process it easily. It allows the user to apply a classifier like SVM to data that have no obvious fixed-dimensional vector space representation (Hastie *et al.* 2009). The prime example of such data in bioinformatics are sequence data, either DNA or protein, and protein structure. The performance of SVM-based predictor in genomic prediction may rely on selected kernel function which used to construct SVM. Popular kernels are linear, radial, polynomial and sigmoid. A comparison between different kernels, which are used to construct SVM regarding their predictive performance is needed to identify the kernel that provides the greatest predictive accuracy. Therefore, in this study the impacts of kernel function selection on the performance of SVM-based predictors in different scenarios of genomic selection was studied.

Present address: [1]Ph.D. Student (amir.kasnavi@srbiau.ac.ir), [2]Assistant Professor (aminafshar@gmail.com), [4]Professor, Department of Animal Science, Science and Research Branch, Islamic Azad University, Tehran, Iran. [3]Assistant Professor (mm.shariati@um.ac.ir), Department of Animal Science, Ferdowsi University of Mashhad, Mashhad, Iran. [5]Assistant Professor (honarvar.mahmood@gmail.com), Department of Animal Science, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran.

MATERIALS AND METHODS

Simulation of data: The *hypred* package (Technow 2013)

was used to simulate a population of animals genotyped for 10,000 SNP. Simulations started with a base population of 100 individuals that were randomly mated for 1,000 discrete generations which allow arriving a mutation-drift balance. A genome was simulated with 5 chromosomes, 100 cM each. The 10,000 bi-allelic SNP markers were distributed along the genome. The coding of each genotype with alleles A_1 and A_2 were, 2 for A_1A_1, 0 for A_2A_2 and 1 for A_1A_2 or A_2A_1.

The population was simulated with an effective population size of $N_e = 100$. At the end of 1,000 generations, the linkage disequilibrium (LD) reached 0.17 which guaranteed that each QTL is in LD with at least one SNP. In this step, we assigned QTL effects to some polymorphic SNPs that were evenly distributed over the genome. Gamma, uniform and normal distributions were assumed for QTL effects (Ghafouri-Kesbi *et al*. 2016). In generation 1001, the population size increased to 1,000 individuals and was labeled as the reference population. Animals in the reference population were genotyped and recorded for the trait. Thereafter, random mating was performed for another generation. The animals in the generation 1002 had

Table 1. Parameters used for simulation program.

Genome size	500 cM
Number of chromosomes	5
Number of marker	10000
Mutation rate per marker	2.5×10^{-3}
Mutation rate per QTL	2.5×10^{-5}
Distribution of additive QTL effects	Normal, Uniform, Gamma
Number of QTL	100, 1000
Effective population size (*Ne*)	100
Heritability	0.3, 0.5, 0.7
Historical population	Generations 1-1000
Reference population	Generation 1001
Validation population	Generation 1002

known genotypes but without phenotypic records and treated as validation population for which genomic breeding values had to be predicted.The parameters used for the simulation of genome are listed in Table 1.

Support vector machines (SVM): The SVM belongs to kernel methods. Kernel methods can be thought of as instance-based learners. Rather than learning some fixed set of parameters corresponding to the features of their inputs, they 'remember' the *i-th* training example. In case of genomic selection, the input are genotypic and phenotypic information of animals in the reference population (χ_i, y_i) and learn for it a corresponding weight w_i. Prediction of unlabeled inputs, i.e. those not in the training set [i.e. phenotypic information of candidate animals ($\hat{Y}$)] is treated by the application of a kernel between the unlabeled input x' and each of the training inputs, x_i. For quantitative responses, a kind of SVM termed Support Vector Regression (SVR) was used. In SVR, with input dataset $G = \{(xi, di)\}_i^n$ (where x_i is the input vector, d_i is the desired real-valued labeling, and n is the number of the input records), x is first mapped into a higher-dimension feature space F via a nonlinear mapping Θ, then linear regression was performed in this space. In other words, SVR approximates a function using the equation (Hastie *et al*. 2009)

$$y = f(x) = w(\Theta)(x) + b$$

The coefficients w and b are estimated by minimizing

$$R(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(d_i, y_i)$$

Where $L_\varepsilon(d, y)$ is the empirical error measured by ε-insensitive loss function

$$L_\varepsilon(d,y) = \begin{cases} |d-y| - \varepsilon, if\ |d-y| \geq 0 \\ 0, otherwise \end{cases}$$

and the term $1/2||w||^2$ is a regularization term. The constant C is specified by the user, and it determines the trade-off between the empirical risk and the regularization term. The ε is also specified by the user, and it is equivalent to the approximation accuracy of the training data. The estimations of w and b are obtained by transforming Eq. (*) into the primal function

$$R\left(W, \in^{(*)}\right) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\in_i + \in_i^*)$$

By introducing Lagrange multipliers, the optimization problem can be transformed into a quadratic programming problem. The solution takes the following form

$$y = f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) + b$$

Where K is the kernel function $K(x, xi) = \Theta(x)^T\Theta(xi)$. By using a kernel function, we can deal with problems of arbitrary dimensionality without having to compute the mapping Θ explicitly. Different kernel functions can be selected to map (or transform) input data to feature space.The potential candidate kernels can be linear, polynomial, radial, and sigmoid, such as:

Linear Kernel (Lin) : $K(x, x') = x^T x'$

Radial Kernel (Rad) : $K(x, x') = exp\left(-\frac{\gamma\|x - x'\|^2}{2\sigma^2}\right)$,

with metaparametery γ

Polynomial Kernel (Pol) : $K(x, x') = \gamma(x^T x' + \alpha)^d$

with meta parameter γ, d, α

Sigmoid Kernel (Sig) : $K(x, x') = tanh(\gamma x^T x' + \alpha)$

with meta parameter γ, α

Except for linear kernel, other kernels have their own meta-parameters that need to be tuned. For radial kernel

there is only one meta-parameter which is gamma (γ). Meta-parameters for polynomial kernel are gamma (γ), degree of polynomial (*d*) and intercept (α). For sigmoid kernel, gamma, and intercept (α) are meta-parameters. Cross validation employing a grid search was used to tune the meta-parameters. With a function including different combinations of meta-parameters, we determined optimum combination of meta-parameters for each kernel. The SVM was run using the *R* package "e1071" (Meyer *et al.* 2013).

Accuracy of genomic breeding values: The predictive accuracy was assessed using the Pearson's correlation coefficient between the predicted genomic breeding values and true (simulated) genomic breeding values ($r_{p,t}$). In addition, mean squared error of genomic breeding values prediction (MSE_p) was employed as a measure of the overall fit achieved with each method. Ten replicates were analyzed, and the mean and standard deviations are presented.

RESULTS AND DISCUSSION

The results of meta-parameters optimization of each kernel together with corresponding cross validation error (*CV-error*), accuracy of prediction ($r_{p,t}$) and mean square error of prediction (MSE_p) achieved by each combination of meta-parameters are presented in Table 2. A wide range of *CV-error* was obtained by fitting different combinations of meta-parameters. In radial kernel, γ= 0.0001 resulted in the best radial-based SVM identified with less *CV-error*. For sigmoid kernel, values 0.0001 and 0.00, respectively, fitted the SVM best. For polynomial kernel (not all the meta-parameters combinations are shown), a combination of meta-parameters as γ=0.01, α=0.00 and *d*=3 resulted in most suitable polynomial-based SVM. As expected, in all the kernel functions, the best combination of meta-parameters which resulted in lowest *CV-error* also provided predictions of genomic breeding values with highest $r_{p,t}$ and lowest MSE_p. Meta-parameters optimization is known to improve classification and prediction accuracy of SVM (Zhu *et al.* 2010, Gaspar *et al.* 2012, Blondel *et al.* 2015). Especially it is important when data set has poor quality which is expected when working with real data. In such situations,

Table 2. Cross validation error (*CV-error*), predictive accuracy ($r_{p,t}$ (SD)) and mean square error of prediction (MSE_p(SD)) of different combination of meta-parameters for each kernel function.

Kernel*	Meta-parameters[b]			*CV-error*	$r_{p,t}$	MSE_p
	γ	α	d			
Rad	0.00001			645.75	0.69 (0.022)	217.61 (27.50)
	0.00005			531.70	0.72 (0.013)	185.00 (11.60)
	0.0001			467.58	0.73 (0.035)	181.20 (16.76)
	0.0005			494.16	0.58 (0.035)	391.60 (35.24)
	0.001			573.08	0.56 (0.035)	407.05 (16.70)
	0.005			594.16	0.08 (0.025)	394.11 (38.35)
	0.01			676.07	0.00 (0.003)	379.00 (25.23)
Sig						
	0.00001	0.00		593.18	0.61 (0.041)	238.56 (32.60)
		0.0001		532.75	0.66 (0.032)	207.66 (26.45)
		0.001		534.42	0.66 (0.040)	199.15 (28.30)
	0.0001	0.00		436.60	0.73 (0.027)	174.12 (19.57)
		0.0001		542.75	0.73 (0.034)	180.62 (24.63)
		0.001		548.20	0.69 (0.041)	191.25 (29.12)
	0.001	0.00		37998.02	0.07 (0.021)	43268.1 (5972.8)
		0.0001		35641.24	0.09 (0.034)	42194.1 (5866.2)
		0.001		32362.03	0.08 (0.022)	51618.3 (5261.2)
	0.01	0.00		53254.90	0.00 (0.061)	45748.6 (6127.8)
		0.0001		49864.52	0.00 (0.043)	44285.5 (5963.2)
		0.001		51284.30	0.00 (0.051)	47491.2 (6085.8)
Pol						
	0.01	0.00	3	525.26	0.63 (0.031)	300.30 (15.54)
		0.001		598.98	0.52 (0.043)	301.62 (29.80)
		0.01		727.93	0.49 (0.031)	332.72 (31.55)
		0.1		744.19	0.56 (0.057)	290.30 (15.88)
	0.01	0.00	4	653.29	0.41 (0.031)	345.44 (22.61)
		0.001		641.12	0.40 (0.035)	329.37 (19.25)
		0.01		675.99	0.43 (0.030)	321.42 (29.13)
		0.1		694.70	0.39 (0.042)	354.56 (24.36)

*Lin, Linear Kernel; Rad, Radial Kernel; Pol, Polynomial Kernel; Sig, Sigmoid Kernel; γ, gamma; α, intercept; d, degree of polynomial; CV-error, cross validation error.

predictive power of SVM is greatly improved by optimization. By optimizing meta-parameters and using optimum values, the generalization power of kernel was maximized, which otherwise might not occur using random or default parameters (Gaspar *et al.* 2012). Improper meta-parameters can decrease the predictive power of SVM in such a way that the predictive accuracy reaches almost zero when the meta-parameters are far from optimum values.

The accuracy of prediction ($r_{p,t}$) and mean squared error of prediction (MSE_p) of SVM-based predictors constructed with different kernel functions studied in different combinations of QTL effect distributions (normal, uniform and gamma), heritabilities (0.3, 0.5 and 0.7) and numbers of QTL (1000, 100) are presented in Tables 3–5. Following the increase in heritability, the $r_{p,t}$ increased and MSE_p decreased with significantly higher values for linear- and polynomial-based SVM. In linear kernel, for example, $r_{p,t}$ increased by 60% and MSE_p decreased by 400% following increase in heritability from 0.3 to 0.7,while in radial kernel, these values were, respectively, 10% increase in $r_{p,t}$ and 30% decrease in MSE_p. Later result revealed higher sensitivity of linear and polynomial kernels to heritability of trait. With a simulation study, Ghafouri-Kesbi *et al.* (2016) reported an increase in accuracy of genomic breeding values from 0.42 to 0.69 by increasing heritability from 0.3 to 0.5. When genetic architecture is based on presence of epistasis and/or low heritability, predictions are not very accurate for almost all genomic selection methods (Neves *et al.* 2012, Howard *et al.* 2014, Combs and Bernardo 2015). According to Deatwyler *et al.* (2013), formula to approximate accuracy has a direct relationship with heritability and an inverse relationship with number of independent chromosome segments or QTLs in the population. By decreasing the number of QTLs from 1000 to 100, the prediction accuracy did not changed significantly ($P>0.05$), but MSE_p decreased significantly ($P<0.05$) with approximately the same values for all SVM-based predictors. By decreasing the number of QTLs, the total genetic variance is divided between a smaller number of QTLs and, therefore, the efficiency of methods to estimate

Table 3. Predictive accuracy ($r_{p,t}$ (SD)) and mean square error of prediction (MSE_p(SD)) of support vector machine (SVM)-based predictor with different kernel function in the normal distribution of QTL effects

Heritability	Kernel*	No. of QTL	$r_{p,t}$	MSE_p
	Lin	1000	0.41 (0.031)^{a}	1443.36 (135.90)^{a}
	Rad	1000	0.65 (0.017)^{b}	224.44 (17.03)^{b}
	Pol	1000	0.51 (0.021)^{a}	314.59 (25.52)^{c}
	Sig	1000	0.62 (0.020)^{b}	271.05 (32.45)^{bc}
h^2=0.3				
	Lin	100	0.48 (0.038)^{a}	124.11 (7.98)^{a}
	Rad	100	0.66 (0.033)^{b}	22.87 (2.48)^{b}
	Pol	100	0.51 (0.032)^{a}	30.94 (4.78)^{b}
	Sig	100	0.59 (0.036)^{b}	28.80 (6.07)^{b}
	Lin	1000	0.55 (0.037)^{a}	605.29 (41.50)^{a}
	Rad	1000	0.74 (0.024)^{b}	175.85 (19.45)^{b}
	Pol	1000	0.62 (0.024)^{a}	278.28 (17.16)^{c}
	Sig	1000	0.74 (0.027)^{b}	185.48 (14.40)^{b}
h^2=0.5				
	Lin	100	0.60 (0.015)^{a}	61.00 (13.71)^{a}
	Rad	100	0.74 (0.017)^{b}	19.40 (2.73)^{b}
	Pol	100	0.65 (0.022)^{ab}	26.10 (2.56)^{b}
	Sig	100	0.73 (0.027)^{b}	20.04 (3.85)^{b}
	Lin	1000	0.71 (0.010)^{a}	270.81 (19.47)^{a}
	Rad	1000	0.79 (0.001)^{b}	157.57 (11.51)^{b}
	Pol	1000	0.68 (0.037)^{a}	242.97 (18.79)^{a}
	Sig	1000	0.80 (0.017)^{b}	127.48 (9.49)^{b}
h^2=0.7				
	Lin	100	0.68 (0.044)^{a}	27.05 (3.17)^{a}
	Rad	100	0.80 (0.017)^{b}	15.36 (1.87)^{b}
	Pol	100	0.67 (0.014)^{a}	29.05 (3.04)^{a}
	Sig	100	0.81 (0.026)^{b}	14.34 (2.96)^{b}

*Lin, Linear Kernel; Rad, Radial Kernel; Pol, Polynomial Kernel; Sig, Sigmoid Kernel. Mean values that do not have a common superscript are significantly different ($P<0.05$).

Table 4. Predictive accuracy ($r_{p,t}$ (SD)) and mean square error of prediction (MSE_p(SD)) of support vector machine (SVM)-based predictor with different kernel function in the uniform distribution of QTL effects

Heritability	Kernel*	No. of QTL	$r_{p,t}$	MSE_p
	Lin	1000	0.44 (0.063)^{a}	420.38 (92.29)^{a}
	Rad	1000	0.66 (0.014)^{b}	81.01 (8.09)^{b}
	Pol	1000	0.51 (0.055)^{a}	114.82 (7.75)^{c}
	Sig	1000	0.66 (0.026)^{b}	89.16 (7.27)^{b}
h^2=0.3				
	Lin	100	0.41 (0.036)^{a}	42.18 (5.85)^{a}
	Rad	100	0.64 (0.045)^{b}	7.62 (1.61)^{b}
	Pol	100	0.51 (0.025)^{a}	10.15 (1.05)^{b}
	Sig	100	0.63 (0.069)^{b}	8.46 (1.67)^{b}
	Lin	1000	0.61 (0.059)^{a}	138.73 (68.87)^{ac}
	Rad	1000	0.76 (0.016)^{b}	58.27 (4.66)^{b}
	Pol	1000	0.60 (0.058)^{a}	101.91 (14.83)^{c}
	Sig	1000	0.74 (0.014)^{b}	57.26 (3.31)^{b}
h^2=0.5				
	Lin	100	0.60 (0.024)^{a}	15.75 (3.23)^{a}
	Rad	100	0.74 (0.026)^{b}	6.48 (0.37)^{b}
	Pol	100	0.62 (0.017)^{a}	8.61 (0.84)^{c}
	Sig	100	0.74 (0.012)^{b}	5.59 (0.39)^{b}
	Lin	1000	0.71 (0.020)^{a}	105.9 (10.28)^{a}
	Rad	1000	0.79 (0.025)^{b}	54.24 (4.78)^{b}
	Pol	1000	0.71 (0.021)^{b}	84.57 (9.84)^{c}
	Sig	1000	0.80 (0.013)^{b}	50.13 (3.02)^{b}
h^2=0.7				
	Lin	100	0.72 (0.018)^{a}	8.68 (1.21)^{a}
	Rad	100	0.79 (0.014)^{b}	5.72 (0.43)^{b}
	Pol	100	0.70 (0.018)^{a}	8.65 (0.90)^{c}
	Sig	100	0.81 (0.013)^{b}	4.60 (0.40)^{b}

*Lin, Linear Kernel; Rad, Radial Kernel; Pol, Polynomial Kernel; Sig, Sigmoid Kernel. Mean values that do not have a common superscript are significantly different ($P<0.05$).

Table 5. Predictive accuracy ($r_{p,t}$ (SD)) and mean square error of prediction (MSE_p(SD)) of support vector machine (SVM)-based predictor with different kernel function in the gamma distribution of QTL effects

Heritability	Kernel*	No. of QTL	$r_{p,t}$	MSE_p
	Lin	1000	0.44 (0.037)^a	2053.45 (590.03)^a
	Rad	1000	0.65 (0.029)^b	223.03 (114.27)^b
	Pol	1000	0.49 (0.023)^a	392.15 (64.27)^b
	Sig	1000	0.64 (0.017)^b	256.11 (142.17)^b
h^2=0.3				
	Lin	100	0.43 (0.026)^a	140.42 (26.63)^a
	Rad	100	0.73 (0.024)^b	28.26 (15.99)^b
	Pol	100	0.53 (0.033)^a	24.06 (11.90)^b
	Sig	100	0.61 (0.025)^b	32.17 (18.94)^b
	Lin	1000	0.59 (0.028)^a	657.80 (192.52)^a
	Rad	1000	0.74 (0.032)^b	194.01 (70.84)^b
	Pol	1000	0.63 (0.024)^{ab}	278.79 (147.01)^b
	Sig	1000	0.75 (0.039)^a	172.30 (112.83)^b
h^2=0.5				
	Lin	100	0.56 (0.036)^a	109.83 (74.84)^a
	Rad	100	0.70 (0.024)^b	26.08 (13.67)^b
	Pol	100	0.59 (0.069)^a	23.09 (15.20)^b
	Sig	100	0.71 (0.049)^b	38.50 (16.13)^b
	Lin	1000	0.73 (0.010)^a	304.02 (125.58)^a
	Rad	1000	0.79 (0.027)^b	161.57 (78.22)^b
	Pol	1000	0.68 (0.026)^a	301.31 (145.26)^b
	Sig	1000	0.80 (0.027)^b	134.15 (62.34)^b
h^2=0.7				
	Lin	100	0.73 (0.030)^a	25.80 (11.88)^a
	Rad	100	0.81 (0.027)^a	20.84 (11.34)^{ab}
	Pol	100	0.70 (0.051)^{ab}	28.69 (13.67)^a
	Sig	100	0.80 (0.030)^a	11.57 (5.27)^b

Lin, Linear Kernel; Rad, Radial Kernel; Pol, Polynomial Kernel; Sig, Sigmoid Kernel. Mean values that do not have a common superscript are significantly different (P<0.05).

Table 6. CPU time (excluding tuning time) and memory required to complete the same analyses

	Lin	Rad	Pol	Sig
Time (sec)	375	385	320	330
Memory	0.62 GIG	0.61 GIG	0.60 GIG	0.64 GIG

*Lin, Linear Kernel; Rad, Radial Kernel; Pol, Polynomial Kernel; Sig, Sigmoid Kernel.

such large QTL effects increases leading to increased accuracy (Ghafouri-Kesbi *et al.* 2016).

Regarding accuracy of prediction ($r_{p,t}$), radial and sigmoid-based SVM outperformed polynomial and linear models (P<0.05), while their predictive performance was approximately the same in most of the studied scenarios. The linear kernel-based SVM had the lowest $r_{p,t}$ and highest MSE_p. The radial kernel function only has one parameter, therefore, constructing the radial SVM predictor would be easier compared to polynomial model which has multiple parameters. Although both radial and sigmoid models can realize the nonlinear mapping in high-dimensional space, there is less limitation in using radial kernel function, whereas, sigmoid may have invalidation values in some parameters (Zhu *et al.* 2010). Therefore, radial kernel, would be recommended for constructing SVM-based predictor. Blondel *et al.* (2015) compared different machine learning methods for ranking individuals according to their genomic breeding value and reported that SVM achieved good accuracy when used a radial kernel. In addition, mean square error of prediction (MSE_p) of radial and sigmoid-based SVM was significantly smaller compared to polynomial and linear models. As MSE_p takes into account both accuracy and bias of prediction, one could conclude that predictions of radial and sigmoid models had less bias. The MSE_p measures the deviations of predicted values from true values and is an important indicator of the quality of predictor. The model is better when MSE is lower and perfect fit is achieved when MSE = 0 (Blondel *et al.* 2015). The high MSE_p of linear model show that it predicts genomic breeding values of candidate animals with higher error. The difference of radial and sigmoid models, were non-significant (P>0.05) but with higher levels of heritability, the sigmoid-based SVM provided predictions with less MSE_p.

A narrow range (320 to 385 seconds) for the computing time requirements as well as required memory (0.60 to 0.64 GIG) for the studied models was observed (Table 6). Therefore, computing time and required memory cannot limit applying any of the SVM-based predictors. Computing time is important, particularly for cross-validation and implementation in practice which requires frequent re-estimation of breeding values. In all the papers we examined including Neves *et al*. (2012) and Ghafouri-Kesbi *et al.* (2016), SVM has been one of the most efficient methods regarding computing time and memory requirement.

In conclusion, our findings showed that meta-parameter

optimization had a significant effect on predictive performance of SVM-based predictors in such a way that by using improper meta-parameters, the predictive power of models decreased significantly. Radial and sigmoid-based SVM models had better predictive performance compared with linear and polynomial models. Difference between radial and sigmoid-based SVM models was negligible. However, considering all aspects, the use of radial-based-SVM in genomic selection is recommended.

REFERENCES

Blondel M, Onogi A, Iwata H and Ueda N. 2015. A ranking approach to genomic selection.*PLoS ONE* **10**(6): e0128570.

Boser B, Guyon I and Vapnik V. 1992. An training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp 263–68. 27–29 July 1992. Pittsburgh, USA.

Combs E and Bernardo R. 2015. Accuracy of genome wide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* **6**: 1.

Daetwyler H D, Calus M P L, Pong-Wong R, de los Campos G and Hickey J M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* **193**: 347–65.

Ghafouri-Kesbi F, Rahimi-Mianji G, Honarvar M and Nejati-Javaremi A. 2016. Predictive ability of random forests, boosting, support vector machines and genomic best linear unbiased prediction in different scenarios of genomic evaluation. *Animal Production Science* **57**: 229–36.

Hastie T J, Tibshirani R and Friedman J. 2009. *The Elements of Statistical Learning.* 745 p, Springer, New York, USA.

Hayes B J and Daetwyler H D. 2015. Genomic selection.Course note. February 2015. Armidale, Australia. Available at: http://jvanderw.une.edu.au/GenomicPredictionCoursNotes Armidale2015.pdf.

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. *Genetics* **4**: 1027–46.

Honarvar M and Ghiasi H. 2013. A comparison of genomic predictions using support vector machines (SVMs) and GBLUP methods. *Agrochimica Research* **57**: 3–21.

Meuwissen T H E, Hayes B J and Goddard M E. 2001. Prediction of total genetic value using genome wide densemarker maps.*Genetics* **157**: 1819–29.

Meyer D, Dimitriadou E, Hornik K, Weingessel A and Leisch K. 2013. Misc functions of the department of statistics (e1071), TU Wien. Available at: http://cran.r-project.org/web/packages/e1071/index.html.

Neves H H R, Carvalheiro R and Queiroz S A. 2012. A comparison of statistical methods for genomic selection in a mice population.*BMC Genetics* **13**: 100.

Scholkopf B, Tsuda K and Vert J P. 2004. Kernel methods in computational biology. MIT Press series on *Computational Molecular Biology*. 425 pp, MIT Press. Cambridge, Massachuse.

Technow F. 2013. hypred: Simulation of genomic data in applied genetics. Available at: http://cran.r-project.org/web/packages/hypred/index.html.

Zhu Y, Tan Y, Hua Y, Wang M, Zhang G and Zhang. 2010. Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. *Journal of Digital Imaging* **23**: 51–65.