RANKING INDIVIDUALS BASED ON PREDICTED PERFORMANCE USING MEAN NORMALIZED DISCOUNTED CUMULATIVE GAIN VALUE

Chesang S. ^{1*}***, Muasya T. K.*** ^{2} ***and Ngeno K*** ^{1}
[1]Department of Animal Science Chuka University PO Box 109-60400, Chuka, Kenya
[2]Animal Breeding and Genomics Group, Department of Animal Science Egerton University
PO Box 536- 20115, Egerton, Kenya
** chesangsumukwo@gmail.com; csumukwo@chuka.ac.ke;*

Citation:

Chesang S., Muasya T. K. and Ngeno K. (2023). Ranking individuals based on predicted performance using mean normalized discounted cumulative gain value. In: Isutsa, D. K. (Ed.). *Proceedings of the Chuka University 9th Annual International Research Conference held in Chuka University, Chuka, Kenya from 24th to 25th November*, 2022. 25-29 pp.

ABSTRACT

Accuracy of genomic prediction always relies on an appropriate choice of a statistical model to capture the relationship between the genetic architecture of a trait and the underlying marker calls in a panel of high-density marker data. However, the ranking problem has become an important subject in machine learning (ML) models, due to its widespread applications in many decision-making processes because the measures of rank quality are usually based on sorting, which is not directly optimizable. To counter this, mean normalized discounted cumulative gain value (MNDCGV), a standard quality measure in information retrieval with capabilities of ranking individuals according to breeding values has been proposed. Few studies have emphasized on the ranking of individuals based on predicted phenotypic values using MNDCGVs but none have been reported in animals. The focus of this study, therefore, was, to evaluate the prediction performance of DeepGS, RR-BLUP and Ensemble GS models using MNDCGV. The MNDCGV results showed the accuracy of GEBVs estimated using DeepGS was approximately equal to 0.75~0.78, RR-BLUP 0.66~0.76 and Ensemble 0.76~0.79 as a result of top-ranked alpha increasing from 10% to 70%. The Ensemble and DeepGS model outperformed the conventional RR-BLUP model by a significant margin ($P<0.05$), therefore they can be used as a supplement to RR- BLUP. Thus, Ensemble and DeepGS models can be given a top priority as GS model and as an alternative to conventional GS models in predicting the performance and ranking of individuals with high breeding values to be used for selection purposes in indigenous chicken breeding programs. Ensemble model performed very well in ranking individuals with better performance compared to DeepGS and RR-BLUP, with improvement values of 0.01 and 0.11, respectively. Thus, Ensemble model can be given a top priority as GS model for performance improvement.

Key words: Cross-validation, Genomic selection, Mean normalized discounted cumulative gain value

INTRODUCTION

Selection criteria based on phenotypes have led to substantial genetic improvement as a result of its ability to utilizes phenotypic or pedigree information in prediction of breeding values. However, with technological advances, there has been a growing interest in developing new statistical models and algorithms capable of predicting untested phenotypes in methods commonly used in genomic selection (Gonzalez-Camacho *et al*., 2018). This led to the development of genome-enabled prediction models in animal breeding (VanRaden *et al*., 2008) such as the Stepwise Regression, RR-BLUP and Bayesian Estimation. The conventional genomic prediction models predict phenotypes using all available single nucleotide polymorphism (SNP) marker data collected from a population using statistical models to predict the marker-trait associations (Meuwissen *et al*., 2001).

The accuracy of genomic prediction depends on an appropriate choice of statistical model. The model chosen should be able to capture the relationship between the genetic architecture of a trait and the underlying marker calls in a panel of high-density marker data (Mcdowel, 2016). Therefore, models such as DeepGS with ability of incorporating interactions between marker features have the capacity to achieve higher accuracy by capturing non-additive effects and noisy data.

Ma *et al.* (2018) proposed an Ensemble model which intergrades DeepGS and RR-BLUP model where parameters are optimized using the particle swarm optimization (PSO) algorithm. This model was developed by inspiration from the social behaviour of bird flocking or fish schooling (Kennedy and Eberhart, 1995). Among many different statistical models developed, no much variation in accuracy of predicted has frequently been observed and reported (Roorkiwal *et al*., 2016).

Application of machine learning (ML) in genome-enabled predictions as a means of improving accuracy has been accelerated in recent years (Bellot *et al.*, 2018). The increased use of ML has been due to a growing interest in using semi- and non-parametric models such as, deep convolutional neural network, artificial neural network and ensemble model for genome-enabled prediction of quantitative traits to account for non-additive gene effects and higher non- linearities as well as genotype-environment interactions (Ehret *et al*., 2015). According to Blondel *et al*. (2015) predictive accuracy of most models are typically assessed using the Pearson's correlation coefficient (PCC) between observed trait values and the predicted trait values despite poor correlations with ranking accuracy.

In order to select the most favourable individuals to be used in genetic improvement programs, it is important to correctly rank individuals from the most favourable to least favourable based on accurately predicting breeding values (Blondel *et al*., 2015). The focus of this study therefore, was to evaluate the prediction performance of GS models using the mean normalized discounted cumulative gain value (MNDCGV) as described by (Blondel *et al*., 2015) which have been described as rewarding more strongly models which assign a high rank to individuals with high breeding value.

MATERIALS AND METHODS

Phenotypes and genotypes source

The phenotypes and genotypes data used for this study were obtained from a study done by Yuan *et al*. (2018) from Chinese indigenous chicken breeds. The study used 394 birds where hens and cocks of the different breeds were phenotyped for body weight (BW). Phenotyping for BW was measured at hatch and every week until 12 weeks of age. For genotypes, blood samples were collected for DNA extraction for all birds used in the study using phenol-chloroform method and diluted to 50 ng/ml. Genotyping was performed using Illumina 60K Chicken SNP BeadChip (Groenen *et al*., 2011). Quality control was done using customized scripts in R software version 3.3.0 (R Core Team, 2017).

Missing Markers and Coding

Imputation using A.mat function on the RR-BLUP package installed in R software (R Core team) were done to impute missing genotypes. This function takes care of markers with 50% missing genotypes. Genotypes were coded into {0 1 2} based on R code script.

Data Analysis

Cross-validation was used to evaluate the prediction performance of GS models as proposed by Crossa *et al*. (2016) and Gianola and Schon (2016). Five-fold cross-validation was used for this study, in which individuals in the dataset were randomly partitioned into five groups of approximately equal size. Using genotypic and phenotypic data, the GS models were trained and validated from four groups with 90% individuals for the training set and 10% for the validation set. The trained GS model was used to predict phenotypic trait of individuals from the remaining group using only genotypic data. This process was repeated five times until each group was used once for testing. The predicted phenotypic traits values were combined for performance evaluation. The entire five-fold cross-validation experiment was repeated ten times with different seeds used to shuffle the order of individuals in the original dataset. Therefore, for each given level of alpha, this procedure produced ten different mean normalized discounted cumulative gain values (MNCGVs), and the average was used as the final result.

The predicted performance for determining individuals with high phenotypic values to be used for selection for each GS model was assessed by measuring the MNDCGVs as described by Blondel *et al.* (2015). Given *n* individuals, the predicted and observed phenotypic values form a $n \times 2$ matrix of score pairs (X, Y). The MNDCGVs for selecting the top-ranked k^{th} individuals were calculated in an iterative manner as follows:

$$MNV(K, X, Y) = \frac{1}{K}\left((K-1)MNV(K-1, X, Y)\right) + \left(\frac{\sum_{i=1}^{k} y(i, X)d(i)}{\sum_{i=1}^{k} y(i, Y)d(i)}\right)$$

where, $d(i) = 1/(\log_2 i + 1)$ is a monotonically decreasing discount function at position i; $y(i,Y)$ is the i^{th} value of observed phenotypic values Y sorted in descending order, here $y(1,Y) \geq y(2,Y) \geq \ldots \geq y(n,Y)$; $y(i,X)$ is the corresponding value of Y in the score pairs (X, Y) for the i^{th} value of predicted scores X sorted in descending order; MNV is the mean normalized value of selecting the top individual. Thus, MNDCGV has a range of 0 to 1 when all the observed phenotypic values are larger than zero; a higher MNV (k, X, Y) indicate a better performance of the model to select the top-ranked k individuals with high phenotypic values.

Statistical analysis

The significance level of the difference between paired samples was examined using the student's t-test as implemented in R software. The mean separation was done to ascertain if the means for the top ranked individuals predicted by the three models were significantly different using least significant difference (LSD) at alpha=0.05.

RESULTS

The MNDCGV was used to evaluate the performance of DeepGS, RR-BLUP and Ensemble (integrated DeepGS and RR-BLUP) for selecting individuals with high phenotypic values for body weight. With top-ranked alpha increasing from 10% to 70%, the results showed that the MNDCGV for DeepGS was approximately equal to 0.75~0.78, RR-BLUP 0.66~0.76 and Ensemble 0.76~0.79 as shown in Figure 3. The top alpha for this work did not reach 100% because the number of individuals used was 394 and when subjected to fivefold cross-validation, 79 individuals remained for performance evaluation.

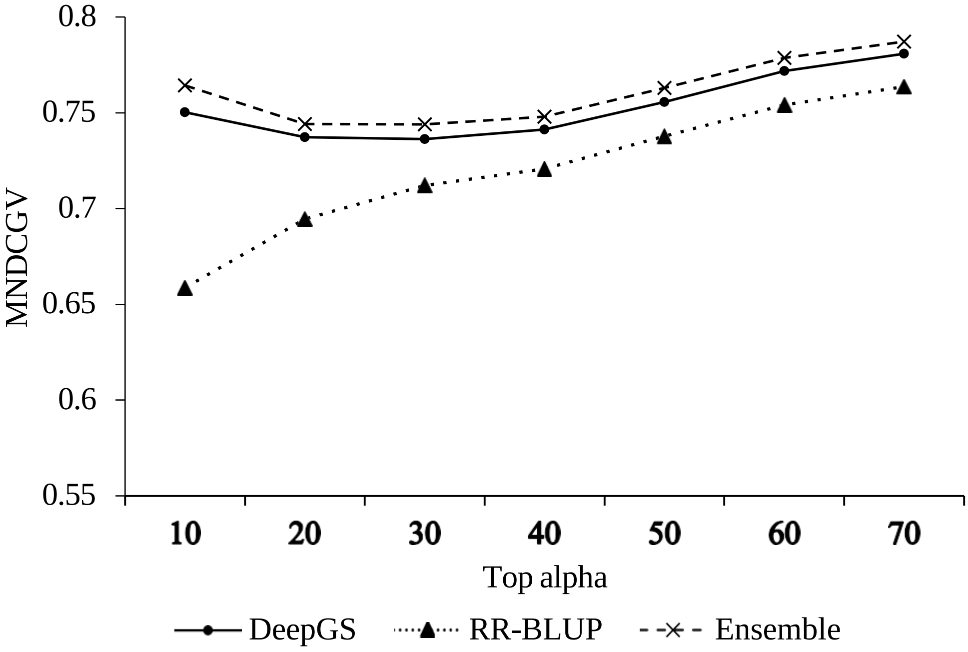

Figure 3: Mean normalized discounted cumulative gain values (MNDCGV) curves for DeepGS, RR-BLUP, and Ensemble model with top-ranked alpha increasing from 10% to 70%

The mean separation was done to using LSD at alpha=0.05 to ascertain if the means for the top ranked individuals predicted by the three models were significantly different. The results showed that MNDCGVs for DeepGS and Ensemble were significantly higher than those of RR-BLUP as shown in Table 3.

Table 3: Mean separation using least significant difference (LSD*) for GS models*

Model	Mean	N
Ensemble	0.76129^{a}	7
DeepGS	0.75332^{a}	7
RR-BLUP	0.72018^{b}	7

Means with the same letter are not significantly different.

This study also aimed at investigating the performance improvement (absolute increment) for the three GS models.

Integrated (Ensemble) model substantially improved the prediction performance over DeepGS and RR-BLUP. The absolute MNDCGV improvement (absolute increment) at the top-ranked level of alpha = 10% of Ensemble (0.76) over RR-BLUP (0.66) was 0.11 and for Ensemble (0.76) over DeepGS (0.75) was 0.01 shown in Table 4 and with P-values corresponding to 0.01 and 0.000107 respectively. Figure 4 shows the trend of the absolute increment of in MNDCGV of DeepGS and the ensemble GS models over RR-BLUP. When compared with RR-BLUP the median results showed that Ensemble model improved the MNDCGVVs by 0.03 and DeepGS improved by 0.02 corresponding to 3% and 2% respectively.

Table 4: Prediction performance based on MNDCGV for Ensemble, DeepGS and RR-BLUP with top-ranked alpha increasing from 10% to 70%

		MNV		**Ensemble *vs* DeepGS**	**DeepGS *vs* RR-BLUP**	**Ensemble *vs* RR-BLUP**
Top α	**DeepGS**	**RR-BLUP**	**Ensemble**	**MNV improvement**	**MNV improvement**	**MNV improvement**
10	0.75	0.66	0.76	0.01	0.09	0.11
20	0.74	0.69	0.74	0.01	0.04	0.05

30	0.74	0.71	0.74	0.01	0.02	0.03
40	0.74	0.72	0.75	0.01	0.02	0.03
50	0.76	0.73	0.76	0.01	0.02	0.03
60	0.77	0.75	0.78	0.01	0.02	0.02
70	0.78	0.76	0.79	0.01	0.02	0.02

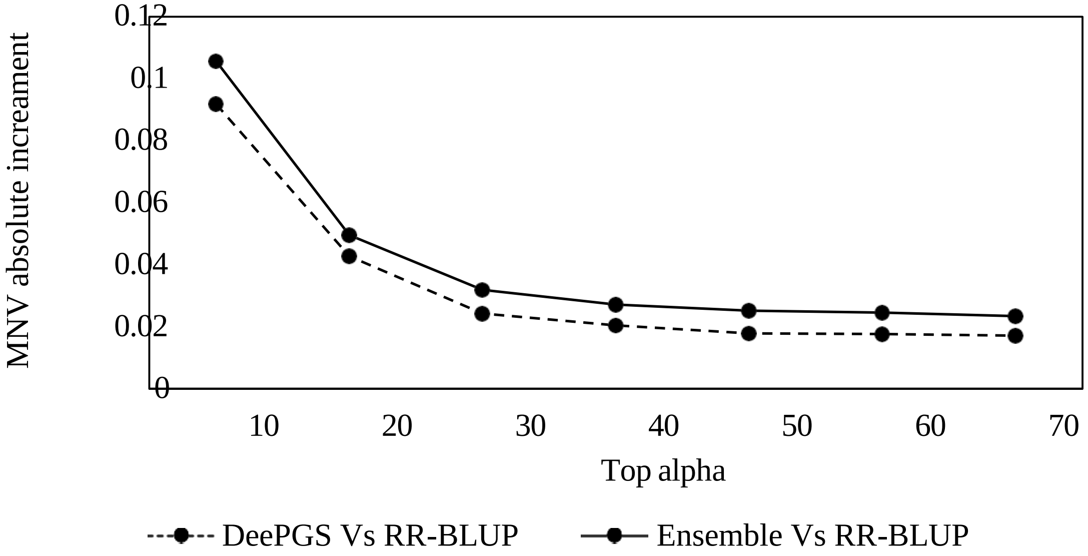

Figure 4: The absolute increase in MNDCGV of DeepGS and the ensemble GS models over RR-BLUP evaluated using fivefold cross-validation with five replicates

DISCUSSION

The ranking problem has become an important topic in machine learning, partly due to its widespread applications in many decision-making processes because the measures of rank quality are usually based on sorting, which is not directly optimizable. To counter this, mean normalized discounted cumulative gain value (MNDCGV), a standard quality measure in information retrieval with capabilities of ranking individuals according to breeding values (performance) has been proposed (Blondel *et al.*, 2015). It has the ability of rewarding models which assign a high rank to individuals with high breeding values. It is also important to select a small number of individuals because selected individuals contribute their genetic materials to the next generation. If too many candidates are selected, selection intensity becomes low and it is not possible to obtain a good improvement of the target trait in the next generation (Blondel *et al.*, 2015). The focus of this study, therefore, was, to evaluate the prediction performance of GS models using the MNDCGV.

The MNDCGV results showed that the DeepGS model was approximately equal to 0.75~0.78, RR-BLUP 0.65859~0.763621 Ensemble 0.764274~0.7871 with top-ranked alpha increasing from 10% to 70%. The Ensemble model and DeepGS model outperformed the conventional RR-BLUP model by a significant margin ($P<0.05$). The mean separation for MNDCGV results for the models showed that there was no significant difference between Ensemble and DeepGS model but RR-BLUP was significant different from the other two models. Therefore, Ensemble and DeepGS can be used as a supplement to RR-BLUP in predicting the performance of individuals with high phenotypes to be used for selection purposes in indigenous chicken breeding programs. A similar trend of results was reported in the phenotypic prediction of a wheat study done by Ma *et al.* (2018). The reason behind this difference in performance ranking is that both Ensemble, DeepGS and RR-BLUP models capture different aspects of the relationship between phenotypes and genotypes, this is attributed to the fact that the models used different algorithms to build regression-based models (Ma *et al.*, 2018). Also, the integrated model (Ensemble) and DeepGS had an advantage over conventional models (RR-BLUP) because of their ability to handle categorical variables and missing values without prior imputation and estimate variable importance and interactions (Blondel *et al*., 2015).

The MNDCGV improvement at the top-ranked level of alpha = 10% results showed that integrated model (Ensemble) performed very well in ranking individuals with better performance compared to DeepGS and RR-BLUP, corresponding to 0.01 for DeepGS and 0.11 for RR-BLUP. This confirms a trend observed by Ma *et al*. (2018) and Blondel *et al*. (2015) in their study on wheat phenotypic prediction and a ranking approach to genomic selection respectively. This is attributed to the fact that ensemble model uses a particle swarm optimization (PSO) algorithm, which has the capability of parallel searching on very large spaces of candidate solutions, without making assumptions about the problem being optimized (Kennedy and Eberhart, 1995). This is also supported by Ma *et al*. (2018) argument on the combination of predictions of DeepGS and RR-BLUP to contribute to better performance.

Performance improvement was also evaluated by comparing the Ensemble and DeepGS model with respect to RR-BLUP. The median results showed that Ensemble model improved the MNDCGV over RR-BLUP by 0.03 and DeepGS improved by 0.02. This further affirms the fact that Ensemble model is superior and robust in performance improvement compared to DeepGS and RR-BLUP. Therefore, Ensemble model can be given a top priority as the GS model and as an alternative to conventional (ridged regression) GS models. The performance of ridge regression model with respect to MNDCGV was not good for this study, these egos what Ma *et al*. (2018) reported but contradicts what Heslot *et al*. (2012) suggested on ridged regression model being the best models for ranking individuals.

CONCLUSION

The performance evaluation results for this study showed that the Ensemble and DeepGS model performed better than RR-BLUP. Therefore, for phenotypic ranking, this work advocates the application of Ensemble and DeepGS model to be used as a supplement to RR-BLUP in ranking of top performing individuals with high phenotypic values to be used for selection purposes. Also, for performance improvement this study suggests use of ensemble model because it has shown to be robust, powerful and effective in comparison to other GS models. Application of MNDCGVs in ranking top performing animals opens up a new avenue for selecting superior animals to be used for genetic improvement programs.

REFERENCES

Bellot, P., de los Campos, G., & Perez-Enciso, M. (2018). Can Deep Learning Improve Genomic Prediction of Complex Human Traits? *Genetics*, 210(3*)*, 809-819.

Blondel, M., Onogi, A., Iwata, H., & Ueda, N. (2015). A ranking approach to genomic selection. *PloS one*, *10*(6), 1-23 Crossa, J., Diego, J., Jorge, F., Paulino, P-R., Juan, B., Carolina, S-P., Prashant, V., Carolina, S., Cesar, P., Deniz, A.,

Clay, S., Matthew, R., Maria, T., Thomas, P., Carlos, G., Roberto, J. P., Peter, W., Sukhwinder S. (2016). Genomic prediction of gene bank wheat landraces. *G3 (Bethesda)*, 6(7), 1819-1834.

Ehret, A. Hochstuhl D, Gianola D, Thaller G. (2015). Application of neural networks with back-propagation to genome- enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle. *Genetics, Selection, Evolution*. 47(1), 22-30.

Gianola, D. and Schon, C.C. (2016). Cross-validation without doing cross-validation in genome-enabled prediction. *G3 (Bethesda)*, 6(10), 3107-3128.

Gianola, D., R. L. Fernando, and A. Stella, (2006) Genomic-assisted prediction of genetic value with semi-parametric procedures. *Genetics* 173, 1761-1776.

Gonzalez-Camacho, J.M., Ornella, L., Parez-Rodriguez, P., Gianola, D., Dreisigacker, S., Crossa, J. (2018). Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. *Plant Genome*, 11(2), 170104-170119.

Groenen, M. A., Megens, H. J., Zare, Y., Warren, W. C., Hillier, L. W., Crooijmans, R. P., & Cheng, H. H. (2011). The development and characterization of a 60K SNP chip for chicken. *BMC Genomics*, *12*(1), 274-283.

Heslot, N., Yang, H. P., Sorrells, M. E., & Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. Crop science, 52(1), 146-160.

Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. In "*Proceedings of IEEE International Conference on Neural Networks*" (Vol. 4, No. 1, pp. 1942-1948).

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., & Ma, C. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Plant*, 248(5), 1307-1318.

McDowell, R. M. (2016). *Genomic selection with deep neural networks*. MSc Thesis, Iowa State University, Ames, Iowa.

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(4):1819-1829.

Okut, H., Wu, X.L., Rosa, G.J., Bauck, S., Woodward B.W., Schnabel, R.D., Taylor, J.F., and Gianola, D., (2013). Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. *Genetics Selection Evolution*, 45:34

Roorkiwal, M., Rathore, A., Das, R. R., Singh, M. K., Jain, A., Srinivasan, S., & Hickey, J. M. (2016). Genome-enabled prediction models for yield related traits in chickpea. *Frontiers in Plant Science*, 7, 1666.

Team, R. C. (2017). R: A language and environment for statistical computing.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414- 4423.

Yuan, Y., Peng, D., Gu, X., Gong, Y., Sheng, Z., & Hu, X. (2018). Polygenic Basis and Variable Genetic Architectures Contribute to the Complex Nature of Body Weight—A Genome-Wide Study in Four Chinese Indigenous Chicken Breeds. *Frontiers in genetics*, *9*, 229.