International Journal of Recent Research and Review, Vol. XII, Issue 3, September 2019

ISSN 2277 – 8322

Ranking Individuals Based on Predicted Performance Using Mean Normalized Discounted Cumulative Gain Value

Sumukwo Chesang[1], Thomas Kainga Muasya[2], Kiplangat Ngeno[3]

[1,2,3]*Animal Breeding and Genomics Group, Department of Animal Science Egerton University P.O.BOX 536- 20115, Egerton, Kenya*

Abstract **- There has been a growing interest in developing new statistical models and algorithms for predicting untested phenotypes in schemes commonly used in genomic selection (GS). 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 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 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 0.66~0.76and Ensemble 0.76~0.79 as a result of top-ranked alpha increasing from 1% 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 model can be given a top priority as GS model and as an alternative to conventional GS models in predicting the performance of individuals with high breeding values to be used for selection purposes in indigenous chicken breeding programs. For performance improvement 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 over Ensemble model respectively.Thus, Ensemble model can be given a top priority as GS model for performance improvement.**

Keywords **- Mean normalized discounted cumulative gain value, Cross-validation, Genomic selection**

I. INTRODUCTION

Traditional breeding programs have led to substantial genetic improvement this is attributed to its abilityto 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 schemes commonly used in genomic selection [9].This led tothe development of genome-enabled prediction models in animal breeding [7]; [3] such as the Stepwise Regression, Ridge Regression–Best Linear Unbiased Prediction (RR-BLUP) and Bayesian Estimation. The conventional genomic prediction models attempt to predict phenotypes by utilizing all available single nucleotide polymorphism (SNP) marker data collected from a population, using one of many possible statistical models to predict the marker-trait associations in a data-driven way [5]. The accuracy of genomic prediction 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 [8]. 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.Reference [14] also proposed an Ensemble model which intergrades DeepGS and RR-BLUP model where parameters are optimized using the particle swarm optimization (PSO) algorithm, which was developed by inspiring from the social behaviour of bird flocking or fish schooling [15]. Among many different statistical models developed, no much variation in accuracy prediction has frequently been observed and reported [12].

Application of chine learning (ML) in genome-enabled predictions as a means of improving accuracy has been accelerated in recent years [17]. 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 [18].According to Blondel *et al*. 2015 predictive accuracy of most models are typically assessed using the Pearson correlation coefficient (PCC) between observed trait values and the predicted trait values despite PCC correlate poorly with ranking accuracy.

In order to select the most favourable individuals to be used in genetic improvement programs, it is importantto correctly rank individuals from the most favourable to least favourable based on phenotypic values rather than to accurately predicting breeding values [1]. 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 [1]. This rewards more strongly models which assign a high rank to individuals with high breeding value. Since itfocuses on the top individuals in the ranking, unlike Pearson correlation which treats all individuals uniformly.

II. MATERIALS AND METHODS

Data source

The data used for this study were obtained from Chinese indigenous chicken breeds[2]. The study used 394 birds from four indigenous breeds consisting of two typical low body weight breeds (Chahua chicken and Silkie chicken) and two intermediate and high body-weight breeds (Beard chicken and Langshan chicken). The birds were randomly selected from the original conserved population and bred by performing artificial insemination of hens with sperm pools. The obtained fertilized eggs were incubated and the chicks were of an approximately equal number of cocks and hens. The hens and cocks of the different breeds were phenotyped for body weight.

Phenotyping

Live body weight (BW) was measured at hatch and every week until 15 weeks of age.

Sample collection, DNA extraction and genotyping

Blood samples were collected from 394 birds at 15 weeks of age. The DNA extraction was done using the phenol-chloroform method and diluted to 50 ng/ml. Genotyping was performed using Illumina 60K Chicken SNP BeadChip [6]. Quality control was conducted on all the birds (after quality control of their phenotypic records) across four breeds by customized scripts in R software version 3.3.0 [4]. Single-nucleotide polymorphisms (SNPs) filtering were done using the following criteria: individual samples were excluded with call rates < 0.9 and minor allele frequency (MAF) < 0.05. After imposing the quality control checks, a total of 46211 SNPs remained.

Genotype pre-processing and coding

The 46211 markers had some missing genotypes. Therefore, missing markers were imputed using A.mat function of the RR-BLUP package installed in R software[4]. Markers with 50% missing genotypes were not imputed, leaving 26698 markers for data analysis. Genotypes were coded into {0 1 2} based on R code script.

III. DATA ANALYSIS

Cross-validation was used to evaluate the prediction performance of GS models as proposed by [16] and [11]. 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 wasassessed by measuring the MNDCGVsas described by [1]. Given *n* individuals, the predicted and observed phenotypic values form an $n \times 2$ matrix of score pairs *(X, Y)*. The MNDCGVs for selecting the top-ranked k^{th}individuals were calculated in an iterative manner as follows:

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

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 GS model to select the top-ranked *k* individuals with high phenotypic values.

IV. 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.

V. 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 1% to 70%, the results showed that the MNDCGVfor DeepGS was approximately equal to 0.75~0.78, RRBLUP 0.66~0.76 and Ensemble 0.76~0.79 as shown in Fig. 1 and Table I**.** The top alpha for this work did not reach 100% because the number of individuals used was 394 when subjected to fivefold cross-validation 79 individuals remained for performance evaluation.

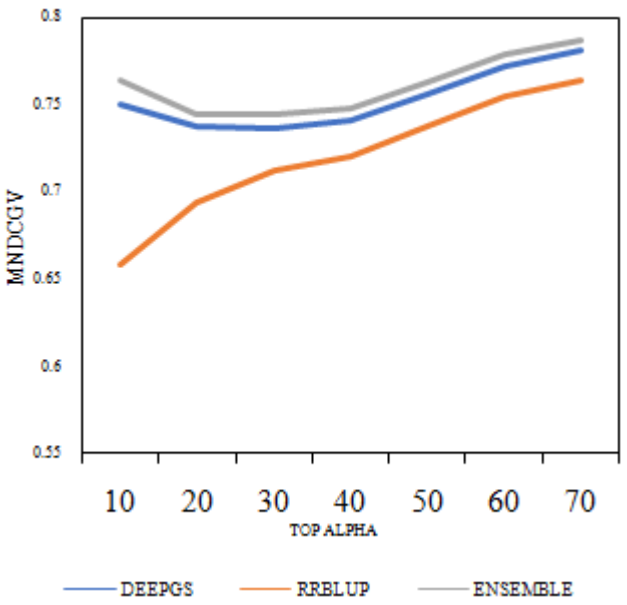

Fig. 1 Mean normalized discounted cumulative gain valuescurves for DeepGS, RR-BLUP, and Ensemble model with top-ranked alpha increasing from 1% to 7%

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, the results showed that MNDCGVs for DeepGS and Ensemble were significantly higher than those of RR-BLUP as shown in Table I.

Table I

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.

We were also interested in 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 at the top-ranked level of alpha = 1% of Ensemble (0.76) over RR-BLUP (0.66) was 0.11 with a P-value of 0.01 and for Ensemble (0.76) over DeepGS (0.75) was 0.01 with a P-value of 0.000107 shown in Table II and Fig. 2**.** 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 II

Prediction performance based on MNDCGV for Ensemble, DeepGS and RR-BLUP with top-ranked α increasing from 1% 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

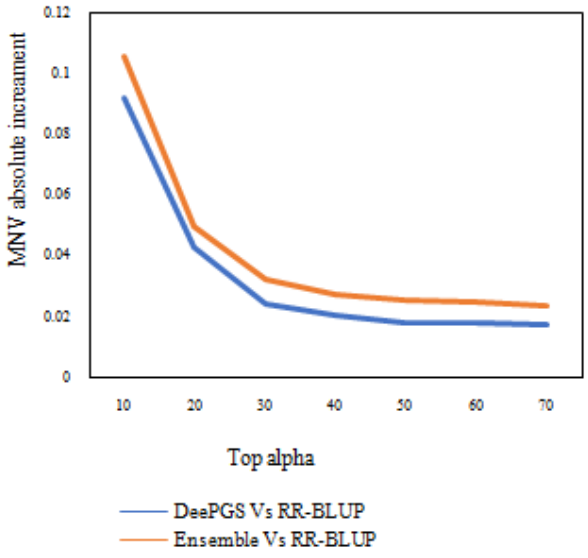

Fig. 2. The absolute increases in MNDCGV of DeepGS and the ensemble GS models over RR-BLUP evaluated using fivefold cross-validation with five replicates

VI. 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[1].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 [1]. 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 was approximately equal to 0.75~0.78, RR-BLUP 0.65859~0.763621 Ensemble 0.764274~0.7871 with Stop-ranked alpha increasing from 1% to 70%. The Ensemble model outperformed the DeepGS model and 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 breeding values 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 [14]. 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 [14]. 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 [1].

The MNDCGV improvement at the top-ranked level of alpha = 1% 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 [14] and [1] 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 [15]. This is also supported by [14] argument on the

combination of predictions ofDeepGS and RR-BLUP may 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 reference [14] reported but contradicts what reference [13] suggested on ridged regression model being the best models for ranking individuals.

VII. CONCLUSION

The mean normalized discounted cumulative gain value is one of the promising models for ranking individuals based on estimated breeding values (phenotypes). It plays an important role in selecting individual candidates with high phenotypic values to be used in breeding programs based on performance evaluation. The performance evaluation for this studyshowed that the Ensemble and DeepGS model performed better than RR-BLUP. Therefore, for phenotypic ranking, this work recommends the application of Ensemble and DeepGS model to be used as a supplement to RR-BLUP in predicting the performance of individuals with high breeding values to be used for selection purposes. Also, for performance improvement ensemble model can be given priority because it has shown to be robust, powerful and effective in comparison to other GS models. The information generated from this study also opens up a new avenue for ranking animals based on MNDCGVs since they are scanty in the animals breeding field.

Ensemble and DeepGS can be used as a supplement to RR-BLUP in predicting the performance of individuals with high breeding values to be used for selection purposes in indigenous chicken breeding programs.

VIII. ACKNOWLEDGEMENT

We would like to thank the centre of excellence in sustainable agriculture and agribusiness management (CESAAM) for giving us grants.

Conflict of Interest

The authors declare no competing interests

IX. REFERENCES

[1] Blondel, M., Onogi, A., Iwata, H., & Ueda, N. (2015). A ranking approach to genomic selection. *PloS one*, *10*(6), e0128570.

[2] 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.

[3] 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

[4] R Core Team (2017). R: A Language and Environment for Statistical Computing. Vienna: R foundation for statistical computing.

[5] 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

[6] 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.

[7] VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414-4423.

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

[9] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Eberhart, Russell, and James Kennedy. "Particle swarm optimization." In Proceedings of the IEEE international conference on neural networks, vol. 4, pp. 1942-1948. 1995.

[16] 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.

[17] 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.

[18] 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.

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

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