



*The Authentic Knowledge: African Journal of
Pure and Applied Science Research
AJPASR - ISSN 2618-0626-p, 2618-0634-e
Volume 1, Issue No.1, 2019
P.O. Box 555 (00202) Nairobi, Kenya
editor@serek.or.ke*

**SOCIETY OF
EDUCATIONAL
RESEARCH
AND
EVALUATION
IN KENYA**

Predictions of Indigenous Chicken Phenotypes from Genotypes: Comparison Between Machine Learning and conventional Linear Models

Sumukwo Chesang, Thomas Kainga Muasya & Kiplangat Ngeno
Animal Breeding and Genomics Group, Department of Animal Science
Egerton University, P.O.BOX 536- 20115, Egerton, Kenya*

Correspondence: Email chesangsumukwo@gmail.com*; muasyaky@yahoo.com; aarapngeno@gmail.com

Abstract

Genomic selection is a new breeding strategy which is rapidly becoming the method of choice of selection. It is useful in predicting the phenotypes of quantitative traits based on genome-wide markers of genotypes using conventional predictive models such as ridge regression best linear unbiased prediction model (BLUP). However, these conventional predictive models are faced with a statistical challenge related to the high dimensionality of marker data, inter and intra-allelic interactions and typically make strong assumptions. Machine learning models can be used as an alternative in the prediction of phenotypes due to their ability to address this challenges. Therefore, the aim of this study was to compare the predictive ability of machine learning using deep convolutional neural network (DeepGS), conventional neural network (Artificial neural network), conventional statistical predictive model ridge regression best linear unbiased (RR-BLUP) model and combination of DeepGS and RR-BLUP (Ensemble model) in predicting body weight (BW) of indigenous chicken based on genome-wide markers. The pearson correlation coefficient (PCC) results from this study for the four models were 0.891, 0.889, 0.892 and 0.812 for DeepGS, RR-BLUP, Ensemble and ANN. This showed that DeepGS did not yield significant difference ($p > 0.05$) from the other models, therefore, it can be used in complement to the commonly used conventional models. For individuals with higher phenotypic values, the PCC results showed a drastic decrease in the performance of DeepGS, rrBLUP, Ensemble and ANN from 0.891, 0.889, 0.892, 0.845 to 0.315, 0.466, 0.342, 0.518 respectively. Therefore, more effort should be put on individuals with higher phenotypic values.

Keywords: Genomic prediction, Machine learning, DeepGS, Artificial neural network

Introduction

Over the past decades, phenotypes have been used to select superior animals for breeding in genetic improvement programmes. However, introduction of genomic selection (GS) technology as a breeding strategy tool has led revolution in genetic improvement programs. Genomic selection is a new breeding strategy which is rapidly becoming the method of choice of selection for animal breeders (Hayes *et al.*, 2009). This is attributed to its ability to accurately identify genetically superior animals at a young age based on genomic estimated breeding values (GEBVs). Genomic selection is useful in predicting the phenotypes of quantitative traits based on genome-wide markers of genotypes using conventional predictive models. The main benefit of GS is its ability to accurately identify genetically superior animals at a young age. The increased accuracy and persistence of prediction based on conventional statistical models at a young age make GS more efficient than progeny testing (calus *et al.*, 2013). The latter is hindered by the long generation interval and low reproductive rate thus lowering genetic progress.

The success of genomic-wide selection (GWS) is related to the choice of the prediction method. Several statistical models have been adopted for predicting GEBV, such as the Best Linear Unbiased Prediction (BLUP) (Henderson, 1975), Stepwise Regression, Ridge Regression–BLUP (RR-BLUP) and Bayesian Estimation (Meuwissen *et al.*, 2001). However, these conventional statistical models do not consider non-linearity among explanatory variable. The association between predictors and target variables which may involve non-additive effects is also ignored, leading to biased results and low predictive abilities (Wenlong *et al.*, 2017). The models make strong assumptions by performing linear regression analysis only thus limiting their predictive performance since they do not capture nonlinear relationships within genotypes and between genotypes and phenotypes (Ma *et al.*, 2017).

Machine learning algorithms provide a new avenue to model noisy data and have been suggested as an attractive, non-linear alternative to traditional statistical models according to literature (LeCun, *et al.*, 2015; Singh *et al.*, 2016; Islam, 2017). They have shown to be superior to conventional statistical models. This is attributed to their ability to model complex and non-linear relationships among acoustic parameters without having to fulfil the strict assumptions required by the conventional parametric model (Favaro *et al.*, 2014). They also have ability to account for non-additive effect in data and can automatically “learn” complex relationships from the training dataset, without pre-defined rules (Ma *et al.*, 2017). The focus of this study was, therefore, to compare the predictive ability of machine learning using deep convolutional neural network (DeepGS), conventional neural network (Artificial neural network) and conventional statistical predictive model ridge regression best linear unbiased (rrBLUP) model in predicting body weight (BW) of indigenous chicken based on genome-wide markers.

Materials and methods

The data used for this study were obtained from a wide-genome study done by Yuan *et al.* (2018) in four Chinese indigenous chicken breeds. 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. Live body weight (BW) was measured at hatch and every week until 15 weeks of age.

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 BeadChips (Groenen *et al.*, 2011). Quality control was conducted on all birds (after quality control of their phenotypic records) across four breeds by customized scripts in R software version 3.3.0 (R Core Team, 2007). 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.

The 46211 markers had some missing genotypes therefore, missing markers were imputed using A.mat function of the rrBLUP package installed in R software. Markers with 50% missing genotypes were not imputed, leaving 26698 markers for data analysis. Genotypes were coded as {0 1 2} based on R code script by Eva KF Chan (<http://www.evachan.org/rscripts.html>)

Data analysis

Conventional Neural Network

The conventional neural network model adopted for this research were a feed-forward artificial neural network (ANN). The data were split into training, validation and test data with 80% used for training, 10% being used for validation and remaining 10% used for testing. Sampling were random in order to avoid any selection bias in the dataset. The Prediction (testing) was done with the remaining data sample that were not used during the training and validation phase. Data preprocessing and analysis were done using the R software, while the construction and training of neural networks were performed using MATLAB. A three-layered, fully-connected, feed-forward artificial neural network were built using a MATLAB software function “feedforwardnet”, with a 18-32-1 architecture (18 nodes in the first hidden layer, 32 nodes in the second hidden layer and one node as an output layer) (Ma *et al.*, 2018). Error optimization was done using back-propagation algorithm (Rumelhart, *et al.*, 1986). Nodes in one layer were fully connected to all nodes in the next layer to form a multilayer ANN architecture.

Conventional statistical model (Ridge Regression-Based Linear Unbiased Prediction (RR-BLUP))

Ridge Regression-Based Linear Unbiased Prediction (RR-BLUP) were adopted as a conventional statistical model. This model was built using the standard linear regression formula, given the genotype matrix as $Z (n \times p)$ and the corresponding phenotype vector as $Y (n \times 1)$;

$$Y = \mu + Z_g + \ell$$

where, n = individuals, p = markers, μ = mean of phenotype vector Y , $g (p \times 1)$ = a vector of marker effects and $\ell (n \times 1)$ = vector of random residual effects. The model was implemented using the function of mixed.solve in R software.

Deep convolutional neural network model (DeepGS)

The DeepGS model was constructed using the deep learning technique as described in a project done by Ma *et al.* (2019) available at GitHub (<https://github.com/cma2015/DeepGS>). Where An 8-32-1 architecture with one input layer, one convolutional layer with 8 neurons, one sampling layer, 3 dropout layers, 2 fully-connected layers and one output layer were constructed. The input layer receives the genotypic markers of a given individual in the $1 \times p$ matrix, where p is the number of genotypic markers. The first convolutional layer filters the input matrix with 8 kernels (weights) which are each 1×18 in size with a stride size of 1×1 , followed by a 1×4 max-pooling layer with a stride size of 1×4 . The output of the max-pooling layer is passed to a dropout layer with a rate of 0.2. The first fully-connected layer with 32 neurons is used after the dropout layer to join together the convolutional characters with a dropout rate of 0.1. Rectified linear unit (ReLU) activation function is applied in the convolutional and first fully connected layers. First fully-connected layer output is then fed to the second fully-connected layer with one neural and a dropout rate of 0.05. Using a linear regression model, the output of the second fully-connected layer is finally connected to the output layer which presents the predicted phenotypic value of the analyzed individual.

The Deep GS model was trained on the training set and validated on the ten-fold cross-validation. Parameters were optimized with the back-propagation algorithm, by setting the number of epochs to 6,000, the learning rate to 0.01, the weight decay to 0.00001 and the momentum to 0.5. The loss function minimized were taken as the mean absolute error (Mae) index (Ma *et al.*, 2019):

$$Mae = \frac{1}{m} \sum_{k=1}^m |predict_k - obs_k|$$

where, m is the number of individuals in the training dataset, and $predict_k$ and obs_k represent the predicted and observed phenotypic values of the k^{th} individual, respectively. Deep GS were implemented using the central processing unit (CPU) based on Deep Learning framework MXNet.

Ensemble model

An ensemble GS model (E) was constructed using the ensemble learning approach by linearly combining the predictions of DeepGS (D) and RR-BLUP (R), using the formula described by Ma *et al.* (2018).

$$Predict_E = (W_D \times predict_D + W_R \times predict_R) / (W_D + W_R)$$

where W_D = weight of DeepGS, W_R = weight of RR-BLUP, $predict_D$ = predictions of DeepGS (D), $predict_R$ = prediction of RR-BLUP

For each fold of cross-validation, parameters (W_D and W_R) were optimized on the corresponding validation dataset using the particle swarm optimization (PSO) algorithm.

Statistical analysis

For each BW trait of a single breed, boxplots were generated by R version 3.3.0 (R CoreTeam, 2017) to screen for outliers. The Pearson's correlation coefficient (PCC) and corresponding significance level ($p < 0.05$) was calculated using R programming language for each model. The significance level of the difference between paired samples were examined using the student's t-test using R software.

Results

The results on Figure 1 show the regression-based predictive ability for body weight based on the four models (DeepGS, rrBLUP, Ensemble and ANN). The experimental results showed that DeepGS, RR-BLUP and Ensemble model yielded a high correlation of 0.891, 0.889 and 0.892 respectively based on Pearson's correlation coefficient (PCC) between predicted phenotypes and observed growth rate at week 5. Artificial neural network yielded relatively lower PCC of 0.813 in comparison with the other models. Generally, all the methods performed similarly, although DeepGS and Ensemble model was slightly better.

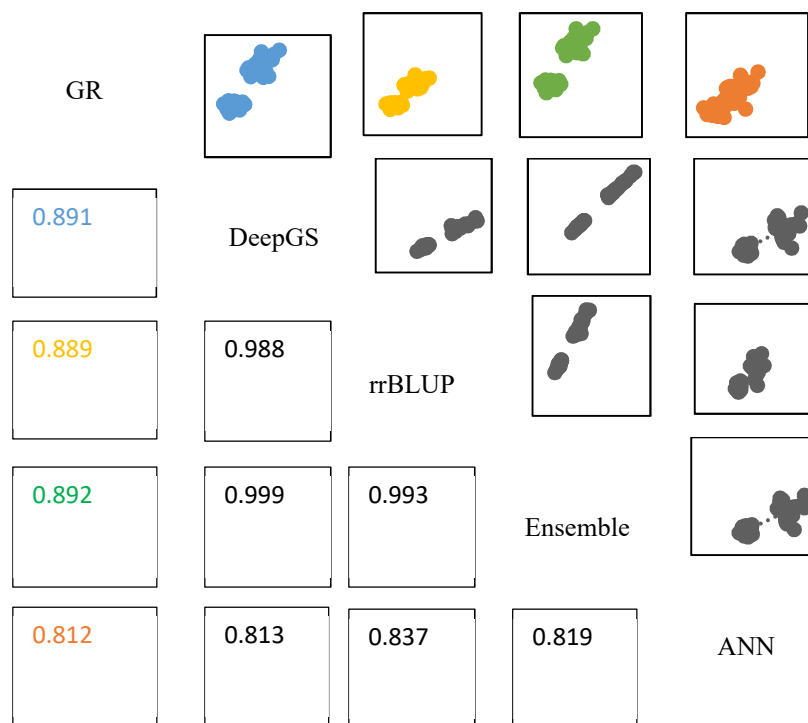


Fig.1 Pearson's correlation coefficient (PCC) analysis between observed and predicted phenotypic values for the four GS models. The correlation results from different GS models are shown in different colours.

The absolute errors between observed and predicted phenotypic values for the four models are presented in Figure 2. The results showed that DeepGS and Ensemble models recorded lower p-values of $1.192e-4$ and $4.65e-4$, respectively, on paired sample t-test compared with 0.0746 for rrBLUP and 0.0506 for ANN.

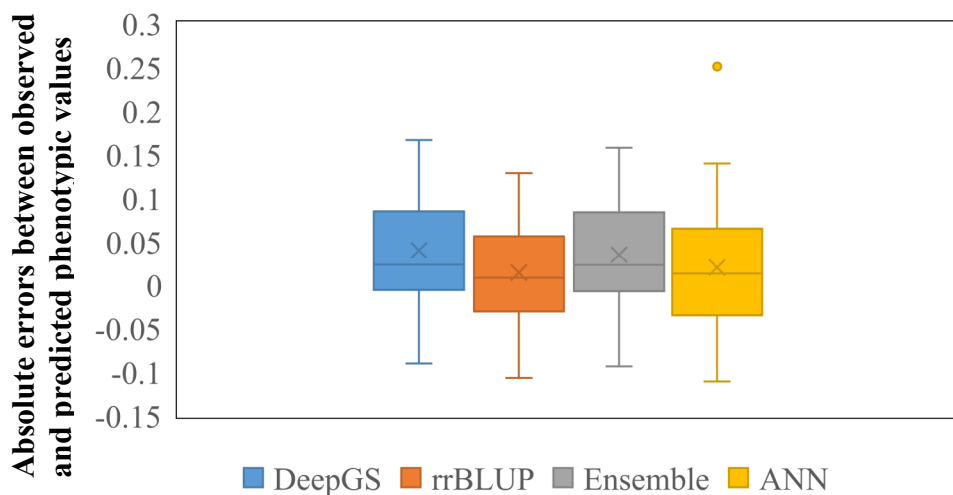


Fig.2 Box plot of the absolute errors between the observed and predicted phenotypic values.

Mean separation was also done to ascertain if the means for body weight obtained from the four models were significantly different from the actual BW mean using the least significant difference (LSD) at $\alpha = 0.05$, the results showed that the means for the four models were not significantly different as presented in Table 1.

Table. 1; mean separation using the least significant difference (LSD) of the GS models

Model	Mean	N
Actual (BW)	0.34416 ^a	39
rrBLUP	0.33032 ^a	39
ANN	0.32424 ^a	39
Ensemble	0.30929 ^a	39
DeepGS	0.30419 ^a	39

Means with different letters within a column are significantly different. The researcher was also interested to know how the top-ranked 200 individuals with the highest BW performed as per prediction of phenotypes is a concern. The PCC value results show a drastic decrease in correlation between observed and predicted phenotypic values corresponding to 0.315, 0.466, 0.342 and 0.518 for DeepGS, rrBLUP, Ensemble and ANN respectively Table 2.

Table 2; PCC between observed and predicted BW based on DeepGS and rrBLUP model

	Observed BW	DeepGS	RR.BLUP	Ensemble	ANN
Observed BW	1	0.314851	0.466294	0.342109	0.518307
DeepGS	0.314851	1	0.934642	0.99835	0.021874
RR.BLUP	0.466294	0.934642	1	0.953517	0.251608
Ensemble	0.342109	0.99835	0.953518	1	0.059165
ANN	0.518307	0.021874	0.251608	0.059165	1

Discussion

The aim of this study was to compare the predictive ability of machine learning (ML) using deep convolutional neural network (DeepGS), conventional neural network (Artificial neural network) and conventional statistical predictive model ridge regression best linear unbiased (rrBLUP) model in predicting body weight (BW) of indigenous chicken based on genome-wide markers. To achieve this, data was split into training, validation and testing datasets, genotype markers were used as input features. The PCC results showed that DeepGS yielded almost the same phenotypic predictions compared to the other three linear models (rrBLUP and ANN). DeepGS did not outperform rrBLUP, Ensemble nor ANN model by a significant margin ($p > 0.05$). This is true because the mean separation results for predicted BW from the four models showed that there was no significant difference between the four models. Similar results were reported by Bellot *et al.* (2018) in the complex human genome, Ma *et al.* (2018) study also reported that DeepGS can be used as a complement to the RR-BLUP model.

The reasons why DeepGS model performed in a similar way or sometimes worse than rrBLUP is that allele frequencies of genotypes are always unbalanced (Bellot *et al.*, 2018) and to a large extent the coding system adopted, which always takes a simple {0 1} or {0 1 2} format. The nature of data, the trait type being predicted and the network architecture also play a crucial role in determining the output of machine learning model, this is because most of the data are noisy which will intern affect the outcome (Bellot *et al.*, 2018). For an instant the PCC on ANN in this study was 0.812 which were way far much higher and contrary to 0.378 reported by Ma *et al.* (2018), this may be as a result of different algorithm adopted, input data set and data splitting criteria. Okut *et al.* (2013) reported correlation training set ranging from 0.776 to 0.858 almost similar to what this study reported after using different training algorithms, different activation functions and different numbers of neurons in hidden layers. Therefore, there is a need for a study to explore how related are chosen markers to functional regions related to the trait being studied since the use of unknown or non-specific markers will lead to the introduction of noise into the dataset (Besic *et al.*, 2017).

The PCC results showed that combining DeepGS and rrBLUP (ensemble model) resulted in better performance compared to RR-BLUP corresponding to 0.892 and 0.889 respectively. This is because validation dataset of ensemble 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). Therefore, the ensemble model can play the same role as DeepGS and rrBLUP model. This study also estimated the absolute error between observed and predicted phenotypic value where the results showed that DeepGS and Ensemble recorded lower p-values compared to rrBLUP and ANN models. This is in consensus with what Ma *et al.* (2018) reported. Lower error indicates a good agreement between observed and predicted phenotypic value. Therefore, the DeepGS and Ensemble model can be used in complementarily for phenotypic predictions.

We were also interested to know how the top-ranked 200 individuals performed. The results indicated that neither DeepGS nor RR-BLUP performed particularly well on body weight prediction since the PCC results were 0.31 and 0.47 respectively. This was similar to results of the study done by Ma *et al.* (2018) that reported almost similar decreasing trend on the prediction of top-ranked 400 individuals with the longest grain length. Therefore, it's important to concentrate on individuals with high breeding values because they play an important role in selection just like selection in a traditional breeding program where an individual with highest phenotypic values is given priority as replacement stock.

Conclusion

Even though the literature demonstrate that machine learning is a powerful tool in predicting phenotypes compared to conventional models, this study finds out that the margin of difference between the tested models was negligible and there was no statistical difference between the four models. The small differences observed might have arisen due to the nature of data used, different training algorithm, network architecture and a number of traits being analyzed. Therefore, DeepGS, Ensemble, RR-BLUP and ANN models can be used interchangeably in making phenotypic predictions.

Acknowledgement

We would like to thank the center of excellence in sustainable agriculture and agribusiness management (CESAAM) for giving us grants. Also appreciate Yangyang Yuan for making data available to be used in this study and Dr. Eva KF Chan for providing with scripts used for coding the data.

conflict of interest

The authors declare no competing interests

Reference

- 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.
- Besic, L., Muhovic, I., Asic, A., Catic, A., Gurbeta, L., & Badnjevic, A. Application of neural networks to the prediction of a phenotypic trait of pacific lampreys based on single nucleotide polymorphism (SNP) genetic markers. *Biomedical Research and Clinical Practice*, 2(5), 1-7.
- Calus, M.P.L., Berry, D.P., Banos, G., de Haas, Y., Veerkamp, R.F., (2013). Genomic selection: the option for new robustness traits. *Advances in Animal Biosciences*. 4:3, 618–625
- Favaro, L., Briefer, E. F., & McElligot, A. G. (2014). Artificial Neural Network approach for revealing individuality, group membership and age information in goat kid contact calls. *Acta Acustica United with Acustica*, 100(4), 782-789.
- 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.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433-443.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2): 423-447.
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization, doi: 10.1109/icnn. 1995.488968. 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.
- Md. Mohaiminul Islam (2017). *Deep Learning Models for Predicting Phenotypic Traits and Diseases from Omics Data*. MSc Thesis, The University of Manitoba, Canada.
- 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
- R Core Team (2017). R: A Language and Environment for Statistical Computing. Vienna: R foundation for statistical computing.
- Singh, R., Lanchantin, J., Robins, G., (2016). Deep Chrome: deep-learning for predicting gene expression from histone modifications. *Bioinformatics*, 32(17), 639-648.

- Wenlong, M., Zhixu, Q., Jie, S., Qian, C., Chuang, M., (2017). Deep GS: Predicting phenotypes from genotypes using Deep Learning.
- 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.