Would large reference populations unveil the potential of deep neural networks for improved genome-enabled prediction of complex traits? The case for body weight in broilers.

CURRENT STATUS: UNDER REVIEW

Tiago Luciano Passafaro  tpassafaro@hotmail.com
University of Wisconsin
Corresponding Author

Fernando B. Lopes
Cobb-Vantress Inc

João R. R. Dórea
University of Wisconsin

Mark Craven
University of Wisconsin

Vivian Breen
Cobb-Vantress Inc

Rachel J. Hawken
Cobb-Vantress Inc

Guilherme J. M. Rosa
University of Wisconsin

DOI:
10.21203/rs.2.22198/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
body weight, broilers, deep neural networks, genome-enabled prediction, and multilayer perceptron
Abstract

Background: Deep neural networks (DNN) are a particular case of artificial neural networks (ANN) composed by multiple hidden layers, and have recently gained attention in genome-enabled prediction of complex traits. Yet, few studies in genome-enabled prediction have assessed the performance of DNN compared to traditional regression models. Strikingly, no clear superiority of DNN has been reported so far, and results seem highly dependent on the species and traits of application. Nevertheless, the relatively small datasets used in previous studies, most with fewer than 5,000 observations may have precluded the full potential of DNN. Therefore, the objective of this study was to investigate the impact of the size of the reference population on the performance of DNN compared to Bayesian regression models for genome-enable prediction of body weight in broilers. Results: Predictive performance of DNN improved as sample size increased, reaching a plateau at about 0.32 of prediction correlation when 60% of the entire training set size was used. Interestingly, DNN showed superior prediction correlation with smaller sample sizes and poorer prediction correlation with larger samples sizes compared to Bayesian Ridge Regression (BRR) and Bayes Cπ without including the tuning data in the training data. Conversely, Bayesian models fitted with the training and tuning sets showed the best performance in terms of prediction correlation, but such advantage vanished for larger sample sizes. DNN presented the lowest mean square error of prediction regardless the amount of data used to train the predictive approaches, as well as with Bayesian models including or not the tuning set into the training set. The predictive bias was lower for DNN compared to Bayesian models regardless the amount of data used with estimates closed to the unit with larger sample sizes. Conclusions: DNN had worse prediction correlation compared to BRR and Bayes Cπ, but improved mean square error of prediction and bias relative to both Bayesian models for genome-enabled prediction of body weight
in broilers. Such findings, highlights advantages and disadvantages between predictive approaches depending on the criterion used for comparison. Nonetheless, further analysis is necessary to detect scenarios where DNN can clearly outperform Bayesian benchmark models.

Introduction

The identification and selection of individuals with superior genetic merit is critical for the improvement of complex traits in animals and plants. Genomic selection was originally proposed by Meuwissen et al. (2001) [1], and been used as a tool to accelerate the genetic improvement of complex traits by earlier and accurate selection of genetically superior individuals compared to traditional pedigree analysis [2, 3]. Advances in genotyping technologies allowed the production of high-density genetic chips in a cost-effective manner, making genomic selection a reality for animal [4, 5, 6] and plant [7, 8] breeding programs.

Genomic selection relies on the information of a large number of genetic markers, posing a statistical challenge for genome-enabled prediction studies in which the number of markers is often much larger than the number of observations. Methods such as G-BLUP [9], Bayes A and Bayes B [1], Bayes C [10], Bayesian Lasso [11], Single-step analysis [12], among others have been proposed to cope with this challenge and also to improve the performance of genome-enabled prediction. In addition, machine learning (ML) techniques have also been implemented in genome-enabled prediction in attempt to improve predictive performance due to their ability to accommodate nonlinear relationships between predictors and response variables. ML methods such as the Reproducing Kernel Hilbert Space [13, 14], Random Forest [15], and Artificial Neural Networks (ANN) [16, 17] have been used in genome-enabled prediction, showing slightly better or similar results compared to linear regression approaches. Recently, a particular case of ANN with
multiple hidden layers, namely, Deep Neural Networks (DNN) has emerged as one of the most powerful machines for pattern recognition, being successfully applied in different fields including computer vision, speech recognition, and machine translation [18].

Deep neural networks are gaining prominence also in genome-enabled prediction and they have been already employed in different studies [19, 20, 21]. However, results reported by these studies have shown no clear superiority of DNN compared to traditional linear regression approaches, with results seem highly dependent on species and traits of application. Nevertheless, the relatively small datasets used in previous studies, most with fewer than 5,000 observations may have precluded the full potential of DNN. For the most successful applications of DNN, the dataset sample sizes had at least 70,000 observations (e.g., MNIST, ImageNet, and VoxCeleb). Thus, large sample sizes could be crucial to unveil the potential of DNN in the genome-enabled field. Bellot et al. (2018) [22] employed DNN for genome-enabled prediction of complex traits in humans using a large dataset composed of 102,221 observations, finding similar performance of DNN and Bayesian regression models. Hence the question remains if DNN cannot indeed outperform Bayesian regression models commonly used in genome-enable prediction of complex traits, or if its performance depends on the species and trait being considered, or if there is also a dependence on size of the reference population used for training the models. Here we try to tackle this latter enquiry, by assessing the relative performance of DNN with varying sizes of training sets. Specifically, we employ a sub-sampling scheme from a large reference population of broiler chickens, and compare the results from DNN and Bayesian regression models on genome-enabled prediction of body weight of broilers.

Results

**Genetic parameter estimates**

Estimates of variance components for body weight were 4,436.6 (SE = 281.07), 1,026.0
(SE = 71.12), and 13477.0 (SE = 163.01) g² for additive genetic, maternal permanent environmental, and residual effects, respectively. These estimates resulted in a phenotypic variance of 18,939.6 (SE = 146.59) g². Estimate of direct heritability for body weight was 0.23 (SE = 0.013), and the proportion of the phenotypic variance due to maternal permanent environmental effect was 0.05 (SE = 0.003).

**Deep neural networks random search**

A random search considering 200 different DNN’s architectures was performed for each sub-sampling of the training set. Deep neural networks were selected based on their prediction correlation on the tuning set. Different architectures of the DNN were selected for each sub-sampling of the training set (Table 1). Overall, DNN with more than one hidden layer showed a greater predictive performance considering up to 50% of the training set size, while simple ANN architectures with one hidden layer and approximately 300-800 units were selected afterwards. All ANN had a L2 norm (ridge regularization) larger than zero, and the dropout rate was smaller than 1, except for the models using 1, 3, and 100% of the entire training set size. The prediction correlation of all ANN is summarized in Figure 1A. Regardless of the ANN architecture, the prediction correlation had an increased trend with larger sample sizes. Interestingly, the distance between the worst to the median prediction correlation of all ANN was greater than the distance between the best to the median prediction correlation of all ANN for each sub-sample of the training set. The MSEP for each ANN are summarized in Figure 1B. Overall, the MSEP had a decreased trend with larger sub-samples sizes. Similarly to the prediction correlation, the distance between the worst to the median MSEP of all ANN was greater than the distance between the best and the median MSEP of all ANN for each sub-sample
of the training set.

**Models’ predictive performance**

As expected, the prediction correlation increased with larger training sample sizes, with a fast increment using up to 50% of the available data, reaching a plateau of approximately 0.32 afterwards, for each genome-enabled prediction approach (Figure 2A). Deep neural networks had the greatest prediction correlation using 1% (0.090) and 3% (0.137) of the training set size, while Bayesian Ridge Regression (BRR) and Bayes Cp fit without the tuning set showed similar or better prediction correlation compared to DNN when more than 5% of the entire training set size was considered. The relative gain of prediction correlation for DNN compared to BRR (Bayes Cp) was 11% (13%) and 7% (7%) when 1% and 3% of the entire training set size was used, respectively. The superiority of the DNN vanished in subsets using more than 5% of the training set size, and the DNN relative gain for prediction correlation was worse compared to both Bayesian methods, varying from -13% to -1%. After fitting the Bayesian regression models with the additional data from the tuning set in each sub-sampling of the training set, the prediction correlation of Bayesian Ridge Regression (BRR-WT) and Bayes Cp (Bayes Cp-WT) were greater than the DNN, regardless of the amount of data used. Moreover, the relative gain of DNN compared to BRR-WT (Bayes Cp-WT) decreased remarkably to -116% (-117%) and -56% (-56%) using 1% and 3% of the training set size, respectively, but such difference in the relative gain was attenuated with larger sample sizes. Overall, the MSEP decreased along with the sample size of the training set for all predictive approaches (Figure 2B). Deep neural networks showed the lowest mean square error of prediction (MSEP) for each subset of the training set, ranging from 26,264.8 to 30,589.3. The relative gain of MSEP was better for DNN compared to BRR (Bayes Cp), ranging from -2% (-2%) to -8% (-8%) when 20% (20%)
and 3% (3%) of the entire training set size was used, respectively. Interestingly, the MSEP of BRR-WT and Bayes $\pi$-WT were greater than DNN for each sub-sampling of the training set, except when 20% of the training data was used.

Deep neural networks showed the smallest predictive bias compared to all Bayesian regression models (Figure 3). Interestingly, the predictive bias of DNN was smaller than one for all partitions of the training set, except when using 30% of the data in the training set. Conversely, Bayesian regression models had a predictive bias greater than one for almost all training set sub-samples, starting after the sub-sampling of 10% and 5% of the training set for models fit with or without the tuning set, respectively. Spearman rank correlations between the predicted body weight from the different genome-enabled prediction approaches through the different sub-sampling of the training set were on average 0.990 (range: 0.981-0.998), 0.917 (range: 0.783-0.968), 0.820 (range: 0.330-0.974), 0.812 (range: 0.333-0.960), 0.908 (range: 0.776-0.957), 0.811 (range: 0.324-0.960), 0.813 (range: 0.328-0.966), 0.772 (range: 0.329-0.926), 0.765 (range: 0.333-0.916), and 0.990 (range: 0.978-0.998) for BRR, Bayes $\pi$, BRR, DNN, BRR, BRR-WT, BRR, Bayes $\pi$-WT, Bayes $\pi$, DNN, Bayes $\pi$, BRR-WT, Bayes $\pi$, Bayes $\pi$-WT, DNN, BRR-WT, DNN, Bayes $\pi$-WT, BRR-WT, Bayes $\pi$-WT, respectively. The agreement on the top 10-ranked broilers selected across the genome-enabled prediction approaches through the different sub-sampling of the training set were on average 91% (range: 86-96 %), 73% (range: 56-83 %), 67% (range: 28-88 %), 65% (range: 28-83 %), 71% (range: 55-81 %), 66% (range: 29-83 %), 66% (range: 29-85 %), 58% (range: 28-74 %), 57% (range: 28-71 %), and 91% (range: 86-96 %) for BRR, Bayes $\pi$, BRR, DNN, BRR, BRR-WT, BRR, Bayes $\pi$-WT, Bayes $\pi$, DNN, Bayes $\pi$, BRR-WT, Bayes $\pi$, Bayes $\pi$-WT, DNN, BRR-WT, DNN, Bayes $\pi$-WT, BRR-WT, Bayes $\pi$-WT, respectively.

Discussion
The heritability estimated for body weight in broiler chickens from a pure line population was of moderate magnitude, accounting for 23% of the phenotypic variance. This result indicates that the response to selection should be effective in a short to medium term. The ratio of maternal permanent environmental variance over the phenotypic variance was low and contributed to 5% of the body weight variation. Although the variance fraction accounted for by the maternal permanent environment effect was relatively low, the inclusion of this effect in the model is essential to avoid an inflation of the variance of the additive genetic effect. Body weight estimates of heritability and the fraction for maternal permanent environmental variance were consistent with other studies using the same trait in broilers from single pure lines [23, 24].

For the DNN implementation, a random search was used for hyperparameter optimization, leading to the selection of different models for each subset of the training set. This result indicates that the choice of the best DNN architecture was strongly affected by the amount of data available during training. Therefore, the random search did not provide a robust DNN structure to predict body weight throughout the training set partitions. Recently, simulated annealing and genetic algorithms have been considered for hyperparameter optimization in machine learning applications [25, 26]. Such approaches may provide a more robust DNN architecture, and as consequence may show a better predictive performance compared to random search. However, Bellot et al. (2018) [22] evaluated the performance of DNN on the genome-enable prediction of complex traits in humans using a genetic algorithm for hyperparameter optimization, and also reported that DNN had similar results with Bayesian regression models.

Hyperparameter optimization is a very difficult task, which involves the exploration of various DNN architectures to find an optimal parameter set within a specific search space. Such component of the learning process is crucial for the success of DNN and is very
demanding on computational resources and time. Parallel computing as employed in our study can be used to alleviate time issues, where each DNN architecture is trained and evaluated independently on different computers. However, parallel computing requires expensive computational resources, which in most situations is not available for many researchers. Despite such challenges, hyperparameter optimization is critical to obtain DNN architectures which could deliver greater predictive performance. For instance, in our study, the difference of predictive performance between the best and worst DNN in each sub-sampling of the training set was considerably large. Therefore, implementing DNN with no hyperparameter optimization may inadvertently define a DNN architecture that delivers a poor predictive performance. Moreover, the hyperparameter optimization cost is relatively minor compared to the cost to collect, store, and analyze genomic data. Therefore, hyperparameter optimization should be considered for genome-enabled prediction applications in animal and plant breeding programs.

The best models selected for each partition of the training set have some type of regularization (i.e. $L2 > 0$ and dropout rate $< 1$) to improve model generalization. The large number of inputs typically observed in genome-enabled prediction, and the high correlation between markers due to linkage disequilibrium may negatively affect the performance of DNN. Regularization approaches such as dropout can prevent complex co-adaptations between units [27], reducing the observed association among inputs from adjacent layers. Therefore, this result suggests that DNN with regularization techniques are recommended to improve predictive performance on new observations for genome-enabled prediction. Similar result was reported by McDowell (2016) [19], who found better predictive performance for DNN with some kind of regularization compared to DNN without regularization for genome-enabled prediction of complex traits in different plant species.

The selection of DNN hyperparameters considering the predictive performance on a tuning
set may not reflect the best predictive performance in the testing set. For instance, for each sub-sampling of the training set at least one DNN with different architecture had a greater predictive performance on the testing set compared to those DNN selected based on the lowest MSEP observed in the tuning set. Therefore, selecting DNN architecture by measuring the predictive performance on a tuning set may not deliver optimized predictive performance on new records. Nevertheless, DNN optimization based on the predictive performance on a testing set provides results that are optimistically biased since some information from the testing set is considered a priori. Thus, in our study the correct strategy was to select the DNN architecture based on the predictive performance in the tuning set.

Deep neural networks are gaining prominence in genome-enabled prediction because of several advantages including flexibility to accommodate complex relationships between output variables and predictors, their high predictive performance, and no parametric assumptions regarding variable distributions [28]. Although DNN has emerged with an enormous potential to transform genome-enable prediction, recent studies showed no evident superiority of DNN relative to traditional genome-enable prediction models. For instance, Rachmatia et al. (2017) [20] used deep belief networks to predict complex traits in maize and found that DNN outperformed linear regression models in only 2 out of 8 traits. McDowell (2016) [19] compared DNN with 5 linear regression methods (i.e. ordinary least squares, lasso, ridge regression, elastic net, and Bayesian ridge regression) on 6 traits from 3 different species (i.e. Arabidopsis, maize, and wheat). In this study DNN outperformed traditional regression methods in about 50% of the time. In another study, Montesinos-Lopez et al. (2018) [21] compared a multi-task DNN with Bayesian multi-trait and multi-environment model using complex traits in maize and wheat under different environments. The authors reported a greater predictive performance of DNN when
genotype x environmental interactions were not included in the analysis and a lower performance when such terms were considered in the analysis. According to these studies, the performance of DNN is strongly affected by many factors including the genetic architecture of a trait, the presence of non-additive effects, hyperparameter optimization, and the DNN architecture considered for genome-enabled prediction (e.g. multilayer perceptron or convolutional neural networks). These findings are consistent with our study, in which Bayesian regression models showed similar or greater prediction correlation than DNN, but worst MSEP.

The lowest MSEP of DNN reflects the predictive bias estimates in each sub-sampling of the training set. Deep neural networks showed greater inflation on the prediction of body weights compared to all Bayesian models using up to 20% of the data, and less biased estimates afterwards, indicating an advantage for DNN over Bayesian models. The Spearman’s correlation and the agreement on the top 10-ranked broilers suggested a re-ranking of animals depending upon to the model used. Such difference in the ranking of broilers is more pronounced between Bayesian regression models fitted with the tuning set in comparison to the other genome-enabled prediction approaches, whereas DNN presented a slightly lower re-ranking of broilers relative to BRR and Bayes Cπ

Interestingly enough, the predictive performance of DNN was better than the BRR and Bayes Cπ when considering small sample sizes. This result is most likely because of the benefit of using in the training process a tuning set exclusive for DNN. However, after re-fitting the Bayesian regression models including also the tuning set data, such an advantage was accounted for and the superiority of DNN vanished. Strategies such as a k-fold cross-validation within the training set could be considered to select DNN architectures. However, in our study, implementing such an approach was extremely difficult due to the computational cost of performing a k-fold cross-validation in such a big
data together with the sub-sampling process in the training set for each genome-enabled prediction approach.

Although DNN often show a greater predictive performance when trained with large sample size, for genome-enable prediction it seems that adding more data per se is not a guarantee to outperform benchmark models. The relative simple nature of the marker inputs (i.e. three genotypes coded as 0, 1 or 2) and the complex essence of quantitative traits may pose a challenge for DNN applied to genome-enabled prediction compared to other successful applications, such as in computer vision [22]. As pointed out by these authors, inputs used in computer vision are more complex and less structured than those available for genome-enabled prediction. Furthermore, the attribute (expected value of trait or genetic risk) used in genome-enabled prediction is often not directly observed, rather it is a function of genetic and environmental factors [22]. Therefore, the characteristics of the response variable and inputs may explain in part the similar predictive performance of DNN and Bayesian methods using large amount of data.

Furthermore, body weight inheritance is suggestive to be mainly accounted for by genetic additive effects, with a lower contribution of non-additive genetic effects. Abdollahi-Arpanahi et al. (2016) [29] concluded that the dominance effects had a minor contribution in the phenotypic variation of body weight relative to additive effects. Additive inheritance is often well fitted by traditional linear models used for genome-enabled prediction. On the other hand, ANN is better suited to capture nonlinear relationships by using multiple layers and nonlinear activation functions. For instance, Dórea et al. (2018) [30] reported greater predictive performance of ANN compared to Partial Least Squares on the prediction of dry matter intake in lactating dairy cows, concluding that such a superiority is possibly explained by the ability of ANN to accommodate nonlinear relationships. Therefore, the additive genetic nature of body weight may be another potential
explanation for the similar predictive performance between DNN and Bayesian models.

It is important to point out some disadvantages of DNN when applied to genome-enable prediction compared to traditional linear regression models. The first drawback has been previously discussed, and reflects the importance of hyperparameter optimization in DNN performance. The second disadvantage is the lack of biological interpretability of the results obtained with DNN. For instance, extracting information from multiple hidden layers is very difficult, turning the algorithm into a “black box” regarding biological interpretation. A practical example of this lack of interpretability is that the effect of each marker cannot be estimated separately, while SNP effects are easily obtained in traditional linear models used for genome-enabled prediction. Another issue of DNN is that such a predictive approach is more susceptible to overfitting than linear models. In our study, we used early stopping, dropout, and a L2 norm to tackle overfitting and the results indeed suggested that such approaches helped to improve generalization. Despite all of these limitations, DNN had a better performance in terms of MSEP but worst prediction correlation compared to the Bayesian regression models. Therefore, DNN should be more explored in genome-enable prediction to find scenarios in which DNN is clearly superior. Common DNN strategies used in the field of computer science including multi-task DNN (i.e. similar to multi-trait analysis), novel algorithms for parameter optimization, and different types of network structures (e.g. convolution and multi-input networks) can be easily adapted and implemented for further analysis in genome-enabled prediction.

Conclusions

Results have shown that the prediction correlation of DNN was comparable to Bayesian regression models with larger training set sizes, while DNN had the lowest MSEP. The inclusion of more data in the training set per se is not a guarantee for DNN to outperform
traditional linear regression models in genome-enabled prediction applications. Overall, the use of DNN for genome-enable prediction is promising but further research investigating novel algorithms for hyperparameter optimization, multi-trait analysis, and other DNN structures are fundamental to evaluate scenarios where DNN can clearly outperform benchmark models.

Methods

(See Methods section in the Supplementary Files)

Abbreviations

**DNN:** Deep Neural Networks  
**ANN:** Artificial Neural Networks  
**BRR:** Bayesian Ridge Regression  
**BRR-WT:** Bayesian Ridge Regression fitted including the tuning set in the training set  
**Bayes Cπ-WT:** Bayes Cπ fitted including the tuning set in the training set  
**MSEP:** Mean Square Error of Prediction  
**SNP:** Single Nucleotide Polymorphism  
**MLPs:** Multilayer Perceptron  
**MSE:** Mean Square Error  
**RG:** Relative Gain

Declarations

*Ethics approval and consent to participate*

Ethics approval and consent to participate, as well as Animal Care and Use Committee approval was not obtained for this study because statistical analysis were performed on a historical data which does not involve human related data. Furthermore, no animal was handled directly.
**Consent for publication**

Not applicable

**Availability of data and materials**

The data that support the findings of this study are available from Cobb upon reasonable request with signed confidentiality agreement contract by contacting Rachel J. Hawken (rachel.hawken@cobb-vantress.com).

**Competing interests**

The author(s) declare(s) that they have no competing interests.

**Funding**

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Brazil provided the financial support for the first author.

**Acknowledgements**

The authors would like to acknowledge Cobb Vantress Inc. for providing the data used for statistical analysis. Furthermore, the authors are grateful to Christina Koch for all of her assistance in using the Center for High Throughput Computing (CHTC) at the University of Wisconsin-Madison.

**Authors’ contributions**

The study was conceived by TLP, FBL, JRRD, and GJMR; the supporting data was provided by VB and RJH, who also helped to understand questions related to the data collection process. Illustrations were drafted by TLP and reviewed by all authors. Data analysis was performed by TLP and FBL. Insightful suggestions and discussions regarding the development of the deep neural networks was provided by MC. TLP wrote the paper and all authors have read and approved the final version of the manuscript.

References
1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819.

2. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 2013;193:327.

3. Meuwissen T, Hayes B, Goddard M. Accelerating improvement of livestock with genomic selection. Annu Rev Anim Biosci. 2013;1:221.

4. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci. 2016;113:E3995.

5. Knol EF, Nielsen B, Knap PW. Genomic selection in commercial pig breeding. Anim Front. 2016;6:15.

6. Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O’Sullivan NP, et al. Implementation of genomic selection in the poultry industry. Anim Front. 2016;6:23.

7. He S, Schulthess AW, Mirdita V, Yusheng Z, Korzun V, Bothe R, et al. Genomic selection in a commercial winter wheat population. Theor Appl Genet. 2016;129:641.

8. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. Genomic selection in plant breeding: Methods, models, and perspectives. Trends Plant Sci. 2017;22:961.

9. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414.

10. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.

11. Park T, Casella G. The Bayesian Lasso. J Am Stat Assoc. 2008;103:681.

12. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including
phenotypic, full pedigree, and genomic information. J Dairy Sci. 2009;92:4648.

13. Gianola D, Van Kaam JBCHM. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics. 2008;178:2289.

14. de los Campos G, Gianola D, Rosa GJM, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res (Camb). 2010;92:295.

15. Sarkar RK, Rao AR, Meher PK, Nepolean T, Mohapatra T. Evaluation of random forest regression for prediction of breeding value from genomewide SNPs. J Genet. 2015;94:187.

16. Gianola D, Okut H, Weigel KA, Rosa GJ. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet. 2011;12:87.

17. Ehret A, Hochstuhl D, Gianola D, Thaller G. Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle. Genet Sel Evol. 2015;47:1.

18. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436.

19. McDowell R. Genomic selection with deep neural networks. Ames, IA; 2016.

20. Rachmatia, H., W. A. Kusuma, and L. S. Hasibuan. 2017. Prediction of maize phenotype based on whole-genome single nucleotide polymorphisms using deep belief networks. J. Phys. Conf. Ser. 835:1.

21. Montesinos-López, A., J. Crossa, D. Gianola, C. M. Hernández-Suárez, and J. Martín-Vallejo. 2018. Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. G3. 8:3829.

22. Bellot P, Campos GDL, Pérez-enciso M. Can deep learning improve genomic prediction of complex human traits? Genetics. 2018;210:809.
23. Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM. Effect of different genomic relationship matrices on accuracy and scale. J Anim Sci. 2011;89:2673.

24. Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z, et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front Genet. 2014;5:1.

25. Young SR, Rose DC, Karnowski TP, Lim S-H, Patton RM. Optimizing deep learning hyper-parameters through an evolutionary algorithm. In: Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments - MLHPC ’15. New York, New York, USA: ACM Press; 2015. p. 1-5.

26. Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, et al. Evolving deep neural networks. In: Artificial Intelligence in the Age of Neural Networks and Brain Computing. Elsevier Inc.; 2017. p. 293-312.

27. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929.

28. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol. 2016;12:878.

29. Abdollahi-Arpanahi R, Morota G, Valente BD, Kranis A, Rosa GJM, Gianola D. Differential contribution of genomic regions to marked genetic variation and prediction of quantitative traits in broiler chickens. Genet Sel Evol. 2016;48:1.

30. Dórea JRR, Rosa GJM, Weld KA, Armentano LE. Mining data from milk infrared spectroscopy to improve feed intake predictions in lactating dairy cows. J Dairy Sci. 2018;101:5878.

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J
Tables

Table 1. The best deep neural network architecture selected based on prediction correlation on the tuning set for each sub-sampling of the training set.
| Size (%) | Number of layers | Number of units per layer | L2\(^2\) | Dropout rate\(^3\) | Accuracy |
|--------|-----------------|--------------------------|---------|-----------------|----------|
| 1      | 4               | 5000\(^{(1)}\).1\(^{(2)}\).600\(^{(3)}\).800\(^{(4)}\) | 0.0600  | 1.0             | 0.090    |
| 3      | 4               | 5000\(^{(1)}\).300\(^{(2)}\).200\(^{(3)}\) | 0.0675  | 1.0             | 0.137    |
| 5      | 3               | 400\(^{(1)}\).200\(^{(2)}\).900\(^{(3)}\) | 0.0100  | 0.5             | 0.145    |
| 7      | 2               | 500\(^{(1)}\).2000\(^{(2)}\) | 0.0450  | 0.8             | 0.166    |
| 10     | 2               | 800\(^{(1)}\).100\(^{(2)}\) | 0.0025  | 0.6             | 0.200    |
| 15     | 2               | 800\(^{(1)}\).900\(^{(2)}\) | 0.0050  | 0.5             | 0.236    |
| 20     | 4               | 600\(^{(1)}\).100\(^{(2)}\).500\(^{(3)}\) | 0.0325  | 0.5             | 0.226    |
| 30     | 1               | 1000\(^{(1)}\) | 0.0100  | 0.7             | 0.274    |
| 40     | 1               | 2000\(^{(1)}\) | 0.0800  | 0.6             | 0.285    |
| 50     | 3               | 600\(^{(1)}\).4000\(^{(2)}\) | 0.0975  | 0.5             | 0.285    |
| 60     | 1               | 300\(^{(1)}\) | 0.0800  | 0.8             | 0.304    |
| 70     | 1               | 400\(^{(1)}\) | 0.0800  | 0.5             | 0.309    |
| 80     | 1               | 800\(^{(1)}\) | 0.0925  | 0.7             | 0.308    |
| 90     | 1               | 400\(^{(1)}\) | 0.0800  | 0.5             | 0.307    |
| 100    | 1               | 500\(^{(1)}\) | 0.0600  | 1               | 0.322    |

\(^{1}\)The number in parenthesis represents the corresponding hidden layer.

\(^{2}\)L2 = ridge regularization.

\(^{3}\)Dropout rate was applied in all layers, except for the output layer.

\(^{4}\)MSEP = mean square error of prediction.

Table 2. Hyperparameters considered in the random search of deep neural networks (DNN)\(^1\).
| Hyperparameter     | Space                                      |
|--------------------|--------------------------------------------|
| Number of units    | 1, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000 |
| Hidden layers      | 1, 2, 3, 4                                 |
| Dropout rate\(^2\) | 0.5, 0.6, 0.7, 0.8, 0.9, 1                 |
| L2\(^3\)          | 0.0000, 0.0025, 0.0050, 0.0075, 0.0100, 0.0125, 0.0150, 0.0175, 0.0200, 0.0225, 0.0250, 0.0350, 0.0375, 0.0400, 0.0425, 0.0450, 0.0475, 0.0500, 0.0525, 0.0550, 0.0575, 0.0600, 0.0700, 0.0725, 0.0750, 0.0775, 0.0800, 0.0825, 0.0850, 0.0875, 0.0900, 0.0925, 0.0950, 0.3000, 0.0325, 0.0350, 0.0375, 0.0400, 0.0425, 0.0450, 0.0475, 0.0500, 0.0525, 0.0550, 0.0575, 0.0600, 0.0625, 0.0650, 0.0675, 0.0700, 0.0725, 0.0750, 0.0775, 0.0800, 0.0825, 0.0850, 0.0875, 0.0900, 0.0925, 0.0950, 0.1000 |

\(^1\)The hyperparameters were randomly select and combined to find the optimal DNN architecture.

\(^2\)The dropout rate was applied in all layers, except for the output layer.

\(^3\)L2 = ridge regularization.

Figures
Predictive performance for each of the 200 deep neural networks generated using the random search, in A) prediction correlation and B) mean square error of prediction (MSEP). The continuous black line represents the median of the 200 deep neural networks for each sub-sample of the training set.

Figure 1
Predictive performance for Bayes $C_\pi$, Bayesian Ridge Regression (BRR), Deep Neural Networks (DNN), Bayes $C_\pi$ fit with the tuning set (Bayes $C_\pi$-WT), and Bayesian Ridge Regression fit with the tuning set (BRR-WT) for each sub-sampling of the entire training set size, in A) prediction correlation and B) mean square error of prediction (MSEP).
Figure 3

Predictive bias for Bayes $C_π$, Bayesian Ridge Regression (BRR), Deep Neural Networks (DNN), Bayes $C_π$ fit with the tuning set (Bayes $C_π$-WT), and Bayesian Ridge Regression fit with the tuning set (BRR-WT) for each sub-sampling of the entire training set size.
Representation of a Multilayer Perceptron (MLPs) architecture. In A) The structure of the deep neural network (DNN) and the training process including forward and backward propagation are depicted. In the forward propagation information flows from the input to the output layers by outputting the calculations of the activation function to the next layer. In the backward propagation, the output is assessed and a loss function $L(W)$ [i.e. mean square error] is used to minimize the overall error function, and consequently update the network weights using stochastic gradient descent. In B) The underlying calculations for each unit in order to provide the output to the next layer. In this process, weight vectors $[W(.)]$ and inputs are linearly combined and transformed based on an activation function, i.e., rectified linear which outputs the maximum between zero and the linear combination of weights and inputs. This figure is based and adapted from the diagram proposed by Angermueller et al. (2016).
Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.
Methods.pdf