Combining environmental and genomic information in artificial neural networks for phenotypic prediction

J. Vandenplas[1], M.P.L. Calus[1], M. Schrauf[1], R. Bonifazi[1], J. ten Napel[1], T. Pook[1], M.C.A.M. Bink[2] and M. Curcic[3,4]

[1] Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

[2] Hendrix Genetics, P.O. Box 114, 5830 AC, Boxmeer, The Netherlands

[3] Department of Ocean Sciences, Rosenstiel School of Marine, Atmospheric and Earth Science, University of Miami, Miami, FL, USA

[4] Frost Institute for Data Science and Computing, University of Miami, Coral Gables, FL, USA

Corresponding author: jeremie.vandenplas@wur.nl

Abstract

Artificial neural networks (ANNs) have received increasing attention for phenotypic and genomic prediction because they can accommodate non-linear relationships and high-dimensional inputs. Their performance has nevertheless been inconsistent when using genomic data, and many reported studies have relied on phenotypes pre-corrected for fixed and non-genetic random effects. The objective of this study was to evaluate the feasibility and predictive ability of ANNs that use environmental and genomic information jointly for phenotypic prediction, thereby avoiding phenotype pre-correction while allowing non-linear relationships among effects to be captured. Three levels of complexity were investigated. The first baseline model (ANN 1) was an ANN including a single input layer combining environmental and genomic data followed by a dropout layer with a 30 % rate, and then a fully connected (FC) layer with a single node and a linear activation function. The second model (ANN 2) consisted in a modular three-network architecture that separated genomic and environmental inputs, each with one FC layer and one node, before combining their predictions to predict phenotypes with a third ANN, also with one FC layer. The third model (ANN 3) was built on the structure of ANN 2 and included additional locally connected layers and non-linear activation functions. The predictive abilities of the ANNs were evaluated using data from the Dutch routine genomic evaluations for dry matter intake (DMI) and enteric methane (CH4) emissions, partitioned into training, validation and test subsets. As benchmark, phenotypes were also predicted with Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP) applied on the training sets only. Modelling environmental and genomic data separately (ANNs 2 and 3) improved predictive ability relative to the single combined-input ANN 1. For DMI, ANN 3 achieved a Pearson correlation of 0.85 with the observed phenotypes for the test subset but did not exceed SNPBLUP (0.87). For CH4 and the test subset, ANN 3 attained the highest Pearson correlation (0.30), exceeding SNPBLUP (0.22). These results indicate that ANN architectures with networks dedicated to a source of information may improve phenotypic prediction, particularly for CH4, although the computational cost of our implementation with the most complex model was substantial compared to SNPBLUP.

Key words: artificial neural networks, phenotypic prediction, environmental effects, DMI, CH4

Introduction

Artificial neural networks (ANNs) have been proposed as an alternative to linear models for phenotypic and genomic prediction because they offer flexibility for modelling non-linear relationships and can accommodate large numbers of predictor variables (Pérez-Enciso

and Zingaretti, 2019; Pook et al., 2020). However, when using genomic data, their predictive performance has not always been favourable, and many published studies rely on phenotypes that were pre-corrected for environmental effects prior to model fitting (Montesinos-López et al., 2018; Waldmann, 2018; Pook et al., 2020).

This study investigated whether ANNs can be used to integrate environmental and genomic data directly for phenotypic prediction. The principal motivation was to remove the need for phenotype pre-correction while allowing potentially non-linear relationships among sources of information to be represented within the prediction model. The comparison was performed using two routine genomic evaluation datasets for the Dutch Holstein population, one for dry matter intake (DMI) and one for enteric methane (CH4) emissions. The Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP) was used as the benchmark method.

Materials and Methods

Data sets

The predictive abilities of the ANNs were evaluated using datasets extracted from the Dutch Holstein genomic evaluations for DMI (E-40-Breeding value dry matter intake, 2022) and CH4 (E-48-Breeding value methane, 2026). After extraction, the DMI dataset included 1 415 first-parity genotyped cows with 25 153 DMI records. The records included in this study were weekly averages and had a mean of 26.59 kg dry matter per day (kg/d) and a standard deviation of 7.29 kg/d. Genotypes included 50 008 SNPs. Environmental effects included in ANNs and SNPBLUP, and therefore contributing to the phenotypic prediction, were experiment, herd, month, year, parity by age at calving, days-in-milk (DIM) fitted with a second-order Legendre polynomial, and permanent environment. Training, validation, and test sets were partitioned by herd and year of measurement, and included 1149, 122 and 144 animals and 20 388, 2253 and 2512 phenotypes, respectively.

The CH4 dataset included 7 284 first-parity genotyped cows with 166 224 sniffer CH4 records. These records were standardized and centered CH4 phenotypes that were calculated from average weekly concentrations (expressed in parts per million), with the mean and standard deviation calculated across the entire dataset. The phenotypes included in this study had a mean of 0.01 and a standard deviation of 1.01. Genotypes included 50 373 SNPs. Environmental effects included in ANNs and SNPBLUP were herd, year, week, parity by DIM fitted with a second-order Legendre polynomial, breed composition, permanent environment, and permanent environment within parity. Training, validation and test sets were defined by year of measurement, with 6453 cows with 45 995 records before 2022 assigned to training, 2675 cows with 105 442 records from 2023 to validation and 4186 cows with 14 787 records from 2024 to testing.

ANN architectures

Three ANN architectures were evaluated for phenotypic prediction. The architectures were designed to assess whether predictive ability could be improved by processing environmental and genomic data separately before combining them. The first baseline ANN (ANN 1) represented a direct approach in which both sources of information were combined at the input stage, whereas the second (ANN 2) and third ANN (ANN 3) used separate subnetworks for environmental and genomic data followed by a third network that combined their outputs. For all ANN architectures, model parameters were optimized using the Adam optimization algorithm.

The ANN 1 consisted of a single network in which the environmental and genomic inputs

were concatenated. The combined input was followed by a dropout layer with a dropout rate of 30%, and subsequently by a fully connected (FC) layer containing a single node with a linear activation function.

The ANN 2 used a modular architecture comprising three subnetworks: a genomic ANN, an environmental ANN, and a final combining ANN. The genomic ANN processed the genomic data using a dropout layer with a dropout rate of 30%, followed by a single-node FC layer with a linear activation function. The environmental ANN consisted of a single-node FC layer with a linear activation function. The outputs of the genomic and environmental ANNs were subsequently used as inputs to the combining ANN, which consisted of a single-node FC layer with a linear activation function to generate the final phenotypic prediction.

The ANN 3 retained the modular structure of ANN 2 but introduced additional layers and non-linear activation functions. The genomic ANN processed the genotypic data through a locally connected layer with three inputs per local receptive field, 32 units, and a ReLU activation function. This layer was followed by a dropout layer with a dropout rate of 30% and a single-node FC layer with a linear activation function. The environmental ANN consisted of a two-node FC layer with a ReLU activation function, followed by a single-node FC layer with a linear activation function. Finally, the combining ANN processed the outputs of the genomic and environmental subnetworks through a two-node FC layer with a Leaky ReLU activation function, followed by a single-node FC layer with a linear activation function to obtain the final phenotypic predictions.

The progression from ANN 1 to ANN 3 therefore allowed the effect of two design principles to be evaluated: first, the separation of environmental and genomic information into dedicated subnetworks and, second, the introduction of additional model complexity and non-linear relationships within and between these sources of information.

Software and benchmark comparison

The ANNs were trained on the training sets using mini-batches of 32 records. Their predictive ability was assessed at each epoch using mean squared errors (MSE) between predicted and observed phenotypes in the validation set. The model with the lowest MSE across up to 50 epochs was used for predictive ability comparison among ANNs using the three subsets. As benchmark, phenotypes of the three subsets were also predicted with SNPBLUP trained on the training sets only.

The SNPBLUP analyses were performed with MiXBLUP (Vandenplas et al., 2022), using the variance component estimates of the Dutch Holstein genomic evaluations. The ANN analyses were performed using the Fortran program *mixnn* developed by Wageningen University and Research and built on the *neural-fortran* library (Curcic, 2019). Environmental data were provided to *mixnn* as sparse incidence matrices, as constructed for mixed models, and SNP genotypes were provided in PLINK 1 binary (BED) format (Chang et al., 2015). All analyses were performed using the same computational environment, with 10 CPU cores available for each analysis.

Results & Discussion

Predictive ability of ANNs

The predictive abilities of the ANNs differed substantially according to the architecture and trait (Tables 1 and 2). Overall, ANN 1, which directly combined environmental and genomic information in a single input layer, showed the poorest predictive ability among the ANN approaches. This pattern was particularly clear for DMI, for which ANN 1 had a considerably higher MSE than ANN 2 and ANN 3 across all subsets. For example, on the test set, the MSE was 49.19 kg^2/d^2 for ANN 1, compared with

14.80 kg^2/d^2 and 13.36 kg^2/d^2 for ANN 2 and ANN 3, respectively (Table 1). Similarly, the Pearson correlation on the test set was only 0.46 for ANN 1, compared with 0.83 and 0.85 for ANN 2 and ANN 3, respectively (Table 2). Therefore, these results suggest that separating environmental and genomic information into dedicated subnetworks can improve predictive performance.

Furthermore, additional flexibility allowed by the subnetworks could be beneficial for phenotypic prediction. For DMI, ANN 3 showed an improvement over ANN 2 on the validation and test sets, with correlations of 0.85 compared with 0.82 and 0.83, respectively (Table 2). The corresponding test set MSE values were 13.36 kg^2/d^2 for ANN 3 and 14.80 kg^2/d^2 for ANN 2 (Table 1). A similar advantage was observed for CH4. On the test set, ANN 3 showed the best performance among the ANN architectures, with an MSE of 0.77 $units^2$ and a correlation of 0.30. In comparison, ANN 2 had an MSE value of 0.93 $units^2$ and a correlation of 0.17, respectively. Thus, although modelling environmental and genomic data separately in ANN 2 improved substantially the predictive ability over combining both sources in a single network, the additional flexibility included in ANN 3 provided an additional advantage for phenotypic prediction.

Comparison with SNPBLUP

For DMI, SNPBLUP provided the best overall predictive performance. On the test set, SNPBLUP achieved a correlation of 0.87 and a MSE of 12.58 kg^2/d^2, compared with a correlation of 0.85 and MSE value of 13.36 kg^2/d^2 for the best ANN (ANN 3; Tables 1 and 2). Thus, although the modular ANN architectures substantially improved upon ANN 1, neither ANN approach exceeded

Table 1. Mean square errors between observed and predicted phenotypes for dry matter intake (DMI) and enteric methane emissions (CH4).

Trait	Subset	SNPBLUP[1]	ANN[2] 1	ANN 2	ANN 3
DMI	Training	4.29	44.24	5.57	6.78
	Validation	11.69	42.90	15.72	13.30
	Test	12.58	49.19	14.80	13.36
CH4	Training	0.32	1.25	1.00	0.56
	Validation	1.03	1.51	0.89	0.91
	Test	1.22	1.47	0.93	0.77

[1] SNPBLUP = Single Nucleotide Polymorphism Best Linear Unbiased Prediction.
[2] ANN = Artificial Neural Network

Table 2. Pearson correlations between observed and predicted phenotypes for dry matter intake (DMI) and enteric methane emissions (CH4).

Trait	Subset	SNPBLUP[1]	ANN[2] 1	ANN 2	ANN 3
DMI	Training	0.96	0.65	0.95	0.94
	Validation	0.87	0.47	0.82	0.85
	Test	0.87	0.46	0.83	0.85
CH4	Training	0.86	0.50	0.60	0.75
	Validation	0.30	0.20	0.30	0.31
	Test	0.22	0.16	0.17	0.30

[1] SNPBLUP = Single Nucleotide Polymorphism Best Linear Unbiased Prediction.
[2] ANN = Artificial Neural Network.

SNPBLUP for DMI. The difference between ANN 3 and SNPBLUP was nevertheless relatively small on the validation and test sets.

For CH4, the results differed from those obtained for DMI. ANN 3 achieved the highest predictive ability on the test set, with a correlation of 0.30, compared with 0.22 for SNPBLUP (Table 2). ANN 3 also had the lowest test set MSE (0.77 units^2), compared with 1.22 units^2 for SNPBLUP (Table 1). On the validation set, however, ANN 3 and SNPBLUP showed similar predictive ability, with correlations of 0.31 and 0.30, respectively (Table 2), and comparable MSE values of 0.91 units^2 and 1.03 units^2 (Table 1). These results indicated that the flexible ANN architecture was particularly advantageous for predicting CH4 phenotypes.

The contrasting results for DMI and CH4 suggested that the benefit of increased model flexibility depends on the trait and prediction scenario. For DMI, the SNPBLUP approach remained competitive and slightly outperformed the ANNs on independent data. In contrast, ANN 3 provided improved prediction for CH4, suggesting that the relationships between the available environmental and genomic information and CH4 phenotypes may benefit from the additional flexibility provided by non-linear and locally connected layers. However, differences in the definition and size of the training, validation and test sets between the two traits, as well as differences in the trait definitions themselves (e.g., CH4 was centered and scaled), should also be considered when comparing results across traits.

Computational performance

The improvement in predictive ability obtained with the more flexible ANNs was accompanied by a substantial increase in computational cost for the training phase. For DMI, wall clock times were 111 seconds (s) for SNPBLUP, 1 455 s for ANN 1, 1 393 s for ANN 2 and 55 708 s for ANN 3. For CH4, the corresponding wall clock times were 741 s, 12 488 s, 12 114 s and 220 859 s. These results underline the trade-off between predictive ability and feasibility. Although ANN 3 provided the strongest overall ANN performance, its computational burden for the training phase was greater than that of the simpler ANN alternatives and SNPBLUP.

Perspective

The results of this study showed the potential of ANNs to predict phenotypes directly from environmental and genomic information, without requiring phenotypes to be pre-corrected before training the ANN. In particular, the results suggest that a modular architecture, in which environmental and genomic information are processed separately before being combined, may provide a useful framework for integrating heterogeneous sources of information.

The extent to which additional network complexity improves prediction appears to depend on the trait and the data structure. While the most complex architecture provided only limited improvements for DMI and did not outperform SNPBLUP, it achieved the highest predictive ability for CH4 on the test set. Further research should therefore investigate which characteristics of traits and datasets determine when more flexible ANN architectures provide an advantage over conventional linear approaches.

An important limitation of the present implementation was the substantial computational cost associated with ANN 3. Future work should therefore focus on improving computational efficiency and systematically evaluating alternative architectures and hyperparameter settings. In addition, the modular framework could be extended to incorporate other sources of information relevant to phenotypic prediction. Such developments could help identify situations in which the additional flexibility of ANNs provides sufficient gains in predictive performance to justify their computational requirements.

Conclusions

ANNs can be applied to phenotypic prediction from environmental and genomic information without relying on pre-corrected phenotypes. The most favourable results were obtained when separate ANNs were assigned to different sources of information, and their predictions were combined in a final network. This modular strategy improved predictive ability relative to a single combined-input network and showed the greatest potential for CH4 prediction. The principal limitation of the most flexible architecture in this study was its substantial computational cost for the training phase.

Acknowledgments

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project LWV20054) and the Breed4Food partners CRV (Arnhem, the Netherlands), Hendrix Genetics (Boxmeer, the Netherlands), and Topigs Norsvin (Den Bosch, the Netherlands). The use of the HPC cluster has been made possible by the RegioDeal Foodvalley (through Shared Research Facilities Wageningen UR).

References

Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7. https://doi.org/10.1186/s13742-015-0047-8.

Curcic, M. 2019. A parallel Fortran framework for neural networks and deep learning. arXiv:1902.06714.

E-40-Breeding value dry matter intake. 2022. Accessed September 14, 2026. https://www.cooperatie-crv.nl/wp-content/uploads/2023/06/E_40-Voeropname-Oktober-2022-Engels.pdf

E-48-Breeding value methane. 2026. Accessed September 14, 2026. https://www.cooperatie-crv.nl/wp-content/uploads/2026/04/E-chapter-methane-april-2026.pdf.

Montesinos-López, O.A., J. Martín-Vallejo, J. Crossa, D. Gianola, C.M. Hernández-Suárez, A. Montesinos-López, P. Juliana, and R. Singh. 2018. A Benchmarking Between Deep Learning, Support Vector Machine and Bayesian Threshold Best Linear Unbiased Prediction for Predicting Ordinal Traits in Plant Breeding. *G3: Genes Genomes Genet.* g3.200998.2018. https://doi.org/10.1534/g3.118.200998.

Pérez-Enciso, M., and L.M. Zingaretti. 2019. A Guide on Deep Learning for Complex Trait Genomic Prediction. *Genes* 10:553. https://doi.org/10.3390/genes10070553.

Pook, T., J. Freudenthal, A. Korte, and H. Simianer. 2020. Using Local Convolutional Neural Networks for Genomic Prediction. *Front. Genet.* 11:561497. https://doi.org/10.3389/fgene.2020.561497.

Vandenplas, J., R. f. Veerkamp, M. p. l. Calus, M. h. Lidauer, I. Strandén, M. Taskinen, M. Schrauf, and J. ten Napel. 2022. MiXBLUP 3.0 - software for large genomic evaluations in animal breeding programs. Pages 1498–1501 in Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers. https://edepot.wur.nl/586651

Waldmann, P. 2018. Approximate Bayesian neural networks in genomic prediction. *Genet. Sel. Evol.* 50:70. https://doi.org/10.1186/s12711-018-0439-1.