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

J. Vandenplas, M.P.L. Calus, M. Schrauf, R. Bonifazi, J. ten Napel, T. Pook, M.C.A.M Bink, M. Curcic

WAGENINGEN UNIVERSITY & RESEARCH
BREED4FOOD
UNIVERSITY OF MIAMI

Artificial neural networks (ANNs)

- Promising for phenotypic and genomic predictions
- Ability to
  - learn non-linear relationships
  - handle high-dimensional data

- However
  - Inconsistent predictive performances with genomic data
  - Usually trained on pre-corrected phenotypes

Aim

To evaluate the feasibility and predictive ability of ANNs

using both environmental and genomic information

for phenotypic prediction

➔ To eliminate the need for pre-correction, and leveraging non-linear relationships among effects

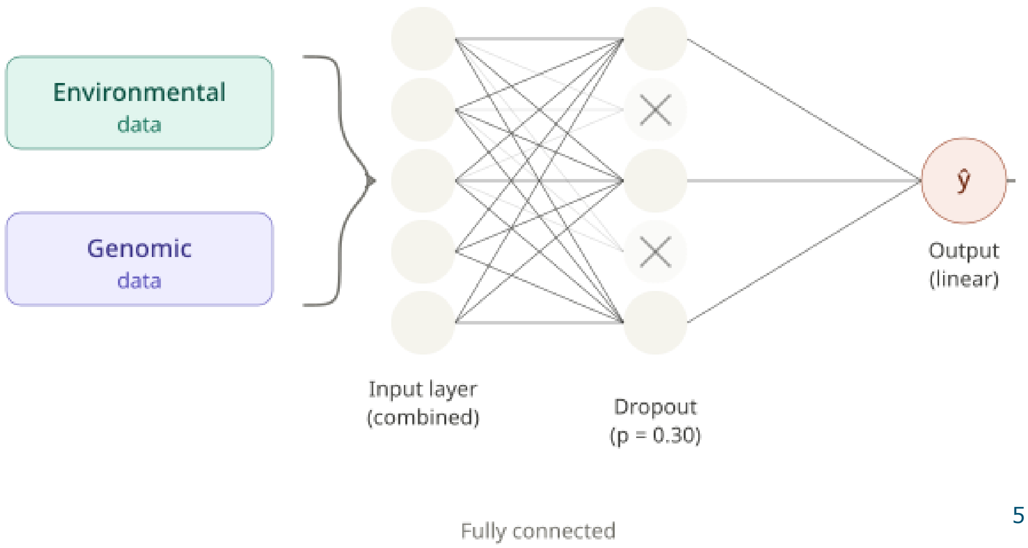
WAGENINGEN UNIVERSITY & RESEARCH

Tools

- MiXBLUP
  - Program for solving large-scale mixed model equations
- mixnn
  - Fortran program that trains a set of ANNs
    - Based on the library neural-fortran https://github.com/modern-fortran/neural-fortran
  - Input
    - Sparse incidence matrices (“fixed” effects)
    - SNP genotype matrices
  - Output: phenotypes

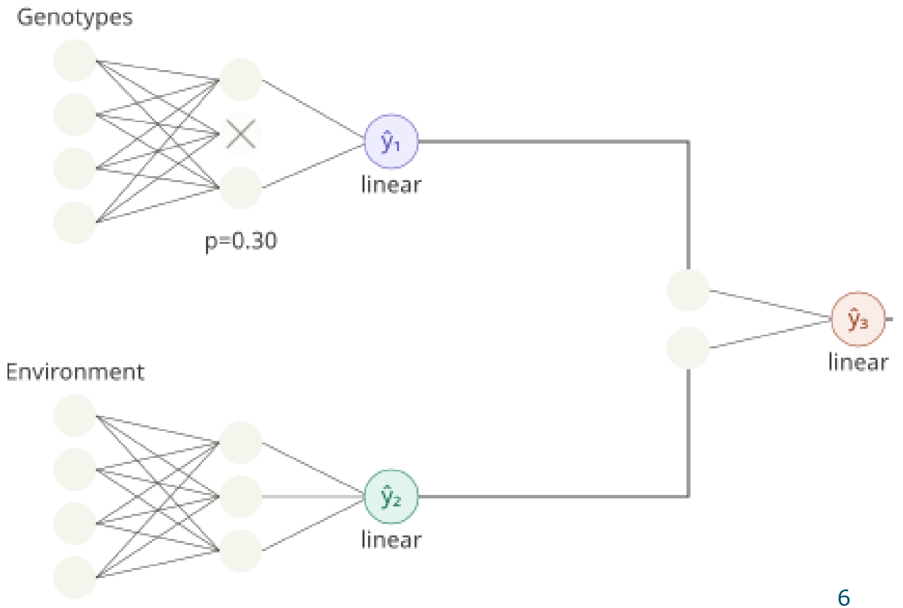
Models – ANN 1

- ANN
  - 1 input layer combining environmental and genomic data
  - Fully connected (output=1, linear)
  - Dropout (0.30)
  - Optimizer: Adam



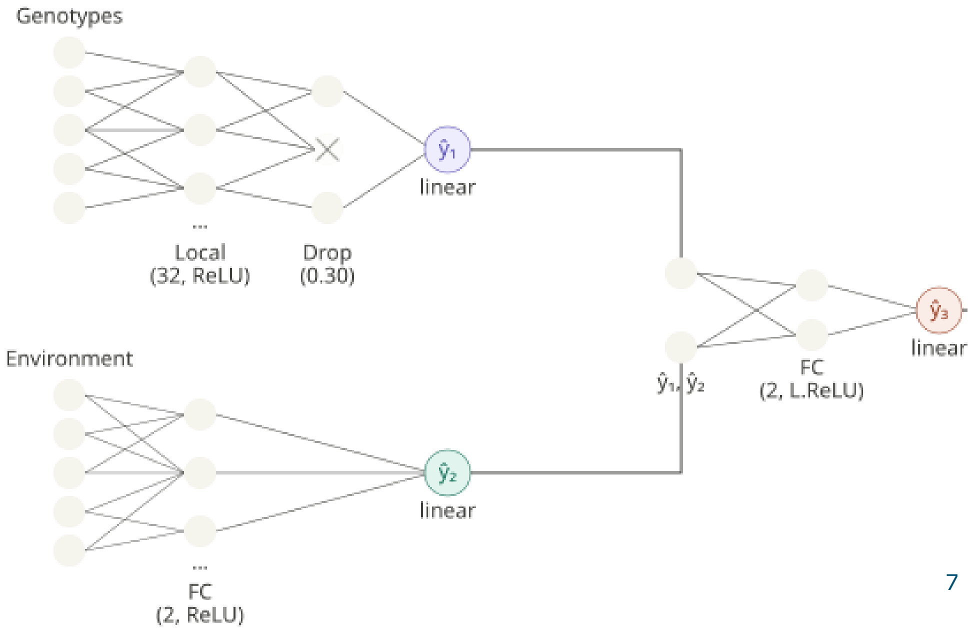
Models – ANN 2

- Set of 3 ANNs
  - ANN 1 (genotypes)
    - Dropout (0.30)
    - FC (output=1, linear)
  - ANN 2 (environment)
    - FC (output=1, linear)
  - ANN 3
    - FC (1, linear)



Models – ANN 3

- Set of 3 ANNs
- ANN 1 (genotypes)
  - Locally connected (3, 32, ReLU)
  - Dropout (0.30)
  - FC (1, linear)
- ANN 2 (environment)
  - FC (2, ReLU)
  - FC (1, linear)
- ANN 3
  - FC (2, Leaky ReLU)
  - FC (1, linear)



WAGENINGEN UNIVERSITY & RESEARCH

Data

- Data
  - Dutch DMI routine genetic evaluation
  - Dutch CH_4 routine genetic evaluation

- 3 subsets
  - Training: training & validation datasets
  - Inference: test dataset (not used in the training phase)

- Predictive abilities presented for the best performing model

Data – DMI evaluation

- 1415 first-parity genotyped cows
- 25,153 DMI records
  - Effects: experiment, herd, month, year, parity*age at calving, DIM (2^{nd} Legendre), permanent environment

- Training, validation and test sets based on herds

Dataset	Training	Validation	Test
Animals	1149	122	144
Phenotypes	20,388	2253	2512

Predictive ability – DMI evaluation

Pearson correlations between true and predicted phenotypes

Dataset	SNPBLUP	ANN 1	ANN 2	ANN 3
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
Time (s)	111	1455	1393	55,708

➔Modelling environmental and genomic information separately improves the predictive ability

➔NN do not outperform SNPBLUP

Data – CH_4 evaluation

- 7284 first-parity genotyped cows
- 166,224 sniffer CH_4 records
  - Effects: herd, year, week, parity*DIM(2nd Legendre), breed composition, perm. env., perm. env. within parity

- Training, validation and test sets based on the year of measurement

Dataset	Training	Validation	Test
Year of measurement	<2022	2023	2024
Animals	6453	2675	4186
Phenotypes	45,995	105,442	14,787

Predictive ability – CH4 evaluation

Pearson correlations between true and predicted phenotypes

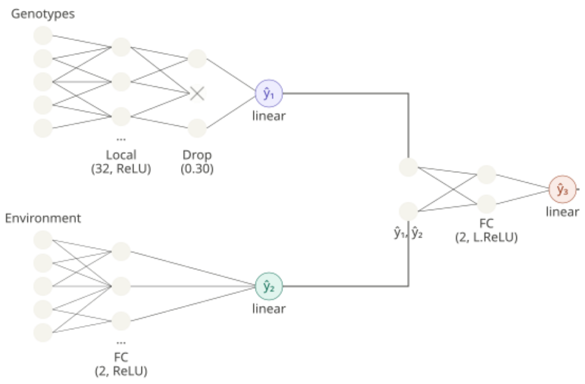
Dataset	SNPBLUP	ANN 1	ANN 2	ANN 3
Training	0.86	0.50	0.60	0.75
Validation	0.30	0.20	0.30	0.31
Test	0.22	0.16	0.17	0.30
Time (s)	741	12,488	12,114	220,859

➔Modelling environmental and genomic information separately improves the predictive ability

➔Some ANNs outperform SNPBLUP

Conclusions

- Set of ANNs
  - 1 ANN dedicated to each source of information
    - Environmental & Genomic
  - 1 ANN to combine the separate predictions into phenotypic predictions



→Might improve phenotypic predictions compared to linear models

WAGENINGEN UNIVERSITY & RESEARCH

Thank you for your attention!

To explore the potential of nature to improve the quality of life

WAGENINGEN UNIVERSITY & RESEARCH

BREED4FOOD

UNIVERSITY OF MIAMI