GenIALearn

EXPLORATORY PROJECT

2021-2023

Application of machine learning and deep learning to improve animal genomic selection

Coordination
Eric Barrey
UMR GABI
eric.barrey@inrae.fr
Didier Boichard
UMR GABI
didier.boichard@inrae.fr

Key words
Genomics
Gene interactions
Statistical learning
Machine learning
Deep learning

INRAE units involved
GABI
MIA Paris Saclay

Partnerships
UEVE Université Paris-Saclay

Context and challenges

The development of genomic selection - and other "omics" analyses such as metagenomics, transcriptomics, metabolomics and proteomics - now makes it possible to characterise animals using thousands of measurements. This massive data is integrated into models to predict production traits with the highest possible degree of accuracy.

The most commonly used models in genomic prediction (additive genetic model such as GBLUP) are very efficient in predicting the genetic value of animals on a few genetically correlated traits. On the other hand, this type of model does not allow the integration of a very large number of heterogeneous measurements, nor does it predict many output traits without knowing their genetic correlations. Moreover, this model is limited in its ability to take into account the many non-linear interactions that occur between regions of the genome or environmental factors.

In order to overcome these obstacles, we propose statistical learning (machine learning) and deep learning methods, derived from AI, to process both additive genetic information and non-linear genetic information present in massive genotyping data.

© kjpargeter, CS - freepik

Metaprogramme
DIGIT-BIO

digitbio@inrae.fr
www.inrae.fr/digitbio/

Goals

The GenIALearn project proposes to evaluate the performance of statistical and deep learning methods for the joint prediction of multiple complex traits, by integrating massive genotyping data. Two main families of methods will be compared altogether and versus the reference method GBLUP:

- on the one hand, ensemble learning methods (random forests, gradient boosting), coupled with a learning step to represent the input data, in order to propose reference prediction levels;
- on the other hand, deep learning methods of different architectures (neural networks), coupled with learning step on massive data base, which should produce predictive models adapted for animal genomic selection.

Research units involved and partnerships

INRAE scientific division	INRAE research units	Expertises
Animal genetics	GABI	Fine phenotyping of complex traits, multi-omics (genotyping, transcriptomics, metagenomics, metabolomics), genetic values evaluation and complex multi-trait predictions.
Mathematics, computer and data sciences, digital technologies	MIA Paris Saclay	Modelling, statistical learning, machine learning, large and heterogeneous data, application to life sciences.
External partners		**Expertises**
UEVE Université Paris-Saclay	UBISC	Neural network construction methods and deep learning, Applications for transcriptomic and image analysis