11. **Using machine learning to improve the prediction of animal performances based on genomics**

Jifan Yang[1*], Pascal Duenk[1], Mario P. L. Calus[1], Yvonne C. J. Wientjes[1]

[1] Animal Breeding and Genomics, Wageningen University & Research, The Netherlands

* Corresponding author. E-mail: jifan.yang@wur.nl

The widespread use of genomic information for predicting animal performances has largely boosted production efficiency of farming in recent years. Used to predict animal performances, conventional prediction methods assume linear relationships between genes and phenotypes. However, these models may not fully capture non-linear relationships (i.e., complex interactions between genes) which are known to be important for complex traits. To overcome this limitation, non-linear machine learning (ML) models like kernel methods and artificial neural networks have been proposed. Although ML looks promising, it is still unknown how to include prior information about causal genes in ML model and whether that is beneficial, which ML method has the highest prediction accuracy, and how the benefit of using ML models changes over generations. Therefore, the objectives of this research are 1) to investigate how to include causal genes as prior information into ML models and whether this results in a higher prediction accuracy; 2) to explore which ML method has the highest prediction accuracy for different combinations of the amount of data and heritabilities of the trait of interest; and 3) to observe whether the benefit of ML models over conventional methods changes across multiple generations of selection based on predicted performances. In order to investigate these objectives, we will use simulations to compare the accuracy of predicted performances of different ML and linear prediction scenarios.

Back to contents