Scientific report summarizing program 362-0691 (year of research publication 2024):

Application of machine learning to the prediction of GEBVs using SNP markers in the Israeli dairy population

Eyal Seroussi and Dror Sagi, Institute of Animal Science, Agricultural Research Organization.

Abstract: Predicting estimated breeding values (EBVs) is important in the dairy cattle industry. Genomic evaluations (GEBVs) are essential to the dairy cattle breeding program. In the past, efficient methods have been developed for genomic selection based on extensive determination of genotypes with the genetic chip technology (Illumina BovineSNP50™). However, effective implementation required a large training series, which is not available for the Israeli herd. When performing genomic predictions by regression of phenotypes to high-density SNPs arrays, the number of markers (p) far exceeds the number of available samples (n). Such extensive regression, when p>n becomes a statistical problem, so that this regression requires methods for selection or reduced representation of the number of markers. Due to the potential loss of information through these techniques, methods based on pattern recognition have become promising alternatives for prediction based on genomic data and are part of a wide variety of machine learning systems, which are able to detect and characterize gene interactions expressed as genotype patterns. In this context, neural nets are powerful tools for genomic prediction, utilizing the full dimension of the data. These networks are considered as a universal approximation of complex and representative functions of any function, thus able to approximate non-linear relationships between the genetic markers and the observed phenotypes, intuitively. This feature is achieved through a universal learning ability provided by certain algorithms, which are applied to train the function. E.g., it is not necessary to determine a genetic inheritance model in advance, so that there is a pre-match to any possible genetic architecture of the target traits. Our study compared the performance of these systems in predicting EBVs based on the local population only, to the performance obtained by the current method that uses a conservative algorithm on a larger sample. The study included the control of the genetic records in the herd book and their adjustment to what is required for predicting accurate breeding values (Gershoni *et al.,* 2022, Genes 13:485); Identifying patterns in the database as a means of reducing its complexity; and a preliminary test of algorithms for predicting accurate cultivation values with the help of the improved database. Next, we will examine algorithms for predicting grooming values based on machine learning. The experiments showed that the algorithm for predicting grooming values of young cattle with the help of an artificial neural network and the representation of the genome as an image reveals similar results to those of the existing system. To achieve improvement, it was decided to investigate in a follow-up program (362-0774), incorporation of existing genomic data for cows; and to propose (Leitner *et al.,* 2024, Genes 15:92) incorporation into the breeding index of an alternative trait for genetic evaluation of mastitis based on lifetime merit.