Genomic prediction with machine learning in sugarcane, a complex highly polyploid clonally propagated crop with substantial non-additive variation for key traits

Chensong Chen[1], Owen Powell[2], Eric Dinglasan[2], Elizabeth Ross[2], Seema Yadav[2], Xianming Wei[3], Felicity Atkin[3], Emily Deomano[3], and Ben Hayes[4]

[1]The University of Queensland
[2]Queensland Alliance for Agriculture and Food Innovation
[3]Sugar Research Australia Ltd
[4]University of Queensland

January 18, 2023

Abstract

Sugarcane has a complex, highly polyploid genome with multi-species ancestry. Additive models for genomic prediction of clonal performance might not capture interactions between genes and alleles from different ploidies and ancestral species. As such genomic prediction in sugarcane presents an interesting case for machine learning methods, which are purportedly able to deal with high levels of complexity in prediction. Here we investigate deep learning networks (DL), including Multilayer networks (MLP) and convolution neural networks (CNN), and Random Forest (RF) for genomic prediction in sugarcane. The data set was 2912 sugarcane clones, scored for 26,086 genome wide SNP markers, with final assessment trial (FAT) data for total cane harvested (TCH), Commercial cane sugar (CCS) and Fibre content. The clones in the latest trial (2017) were used as a validation set. We compared performances of these methods to GBLUP extended to include dominance and epistatic effects. The prediction accuracies from GBLUPs were 0.37 for TCH, 0.37 for CCS and 0.48 for Fibre, while the DL models had accuracies of 0.33 for TCH prediction, 0.38 for CCS prediction and 0.43 for Fibre. Optimised RF achieved a prediction accuracy of 0.35 for TCH, 0.38 for CCS and 0.48 for Fibre. Both DL and RF predictions were more accurate additive GBLUP but generally lower than extended GBLUP. Finally, we identified a partially shared distribution of SNP selections between RF and GBLUP models. We conclude RF may have some utility for genomic prediction for crops with highly complex genomes, particularly if non-additive interactions can be captured with clonal propagation.

Hosted file
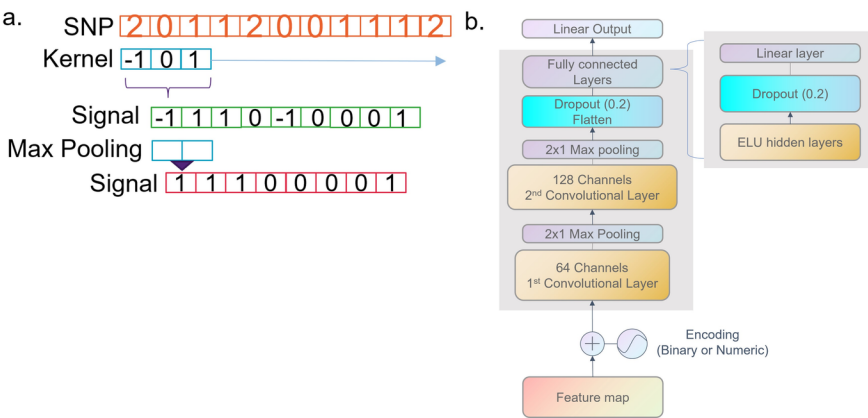
Hosted file

`tpg-2022-12-0186-File003.xlsx` available at https://authorea.com/users/566161/articles/619618-genomic-prediction-with-machine-learning-in-sugarcane-a-complex-highly-polyploid-clonally-propagated-crop-with-substantial-non-additive-variation-for-key-traits

Hosted file

`tpg-2022-12-0186-File005.xlsx` available at https://authorea.com/users/566161/articles/619618-genomic-prediction-with-machine-learning-in-sugarcane-a-complex-highly-polyploid-clonally-propagated-crop-with-substantial-non-additive-variation-for-key-traits