Gefördert durch:

Bundesministerium für Ernährung und Landwirtschaft

aufgrund eines Beschlusses des Deutschen Bundestages

Projektträger

Bundesanstalt für Landwirtschaft und Ernährung

KIBREED

Züchtung von standortangepassten Sorten mittels Algorithmen der Künstlichen Intelligenz

Projektziel (Project Aims)

The project aimed to first, curate winter wheat BigData with data variety including phenotypic, genotypic and enviromic data. Second, to use artificial intelligence approaches for classifying environments and studying causal variants for phenotypic variation.

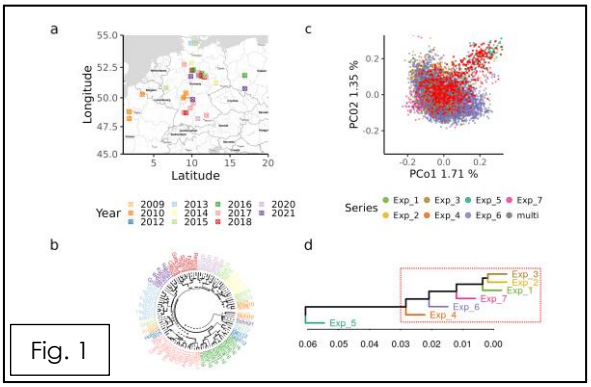

Fig. 1

Fig. 1: (a) Physical map of central Europe with 117 environments (combination of year and location) included in the study, (b) environment clusters derived from enviromic data, (c) genetic structure of experimental series with principle coordinates of genetic distance matrix (d) dendrogram showing population differentiation with F_{st} statistic (on x axis).

Förderbereich des Projektes: Agriculture (plant/animal)
KI-Bereich(e): Machine learning, Deep learning
Projektteilnahme an X-KIT Cluster: Data Analytics Algorithmen, Pflanzenzüchtung

Aktuelle Ergebnisse (Current Results)

1. Winter-wheat BigData was compiled from diverse sources (Fig. 1), with daily weather data during the winter wheat growth period. It includes grain yield information for around 14k genotypes, spanning narrow and wide hybrids, advanced breeding material, and commercial cultivars from seven experimental series. **[Use case 1]** GBLUP (Genomic best linear unbiased prediction) models were compared with a dynamic convolutional neural network (CNN) for predicting mean genotypic (grain yield) values, whereas the complex GBLUP model outperformed in 5-fold cross cross-validation test, the CNN showed a clear trend of rising performance with increasing training set size. Later, **[Use case 2]** these models were then expanded to predict environment specific genotypic (grain yield) values and here GBLUP based models clearly outperformed the CNN model. **[Use case 3]** We then selected a core set of 500 lines (red points in Fig. 1c) and using residuals of their predicted yield clustered the environments with K-nearest neighbors. **[Use case 4]** Lastly, the cluster information was combined with enviromic data and a random forest was used to derive feature importance scores for enviromic markers. The latter revealed air temperature, wind speed, long/short wave radiation and dew point as important markers for environment classification.

2. What was developed in the project: Wheat grain yield prediction models

3. Conclusion of the project: To increase the penetrance of deep learning for grain yield prediction in plant breeding, clear use cases and investments in optimal model hyperspace search are essential before these outcompete GBLUP-based models.

Laufzeit 01.06.2021 – 31.05.2024	**Koordination** Prof. Dr. Jochen C. Reif Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Corrensstraße 3 06466 Seeland / OT Gatersleben **Ansprechperson** Prof. Dr. Jochen C. Reif reif@ipk-gatersleben.de +49 39482 5 840	**Projektbeteiligung** ▪ Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) ▪ KWS SAAT SE & Co. KGaA

Gatersleben: 06.03.2024