Genetic algorithm for predicting performance of species soyabeans and recommending favorable phenotypes - GAPPS

Alja Eremić, Luka Jack Komljen, Aleksandra Zečević, Vedran Gvozderac
UP FAMNIT - Hackathon 4.0
18.05.2025.

Abstract

This study investigates the impact of soil and weather conditions on soybean phenotypes, utilizing historical data and projected climate scenarios extending 100 years into the future. By simulating environmental changes, we aim to identify soybean traits that confer resilience to climate variability. A Genetic Algorithm for Predicting Performance of Soybean (GAPPS) was developed to forecast optimal phenotypic traits under evolving conditions. Key findings highlight the significance of specific crop traits. The GAPPS model offers a predictive framework for breeding soybean varieties optimized for future climates.

Introduction

Climate change poses significant challenges to agriculture, particularly in crop yield and resilience. Soybeans, a globally vital crop, exhibit phenotypic plasticity in response to environmental factors. This research focuses on Ottawa, Kansas, analyzing how soil and weather influence soybean phenotypes. By integrating historical data and future climate projections, we aim to identify traits that enhance resilience and develop a predictive model to guide breeding programs.

Methodology

Data Collection: Utilized historical weather data (2004-2019) from Ottawa, Kansas, including temperature, precipitation, humidity, wind speed, and solar radiation. Phenotypic data encompassed plant height, number of pods, biological weight, sugar content, relative water content, chlorophyll A & B, protein percentage, leaf area index, and seed yield per area.

Climate Projection: Generated 100-year future climate scenarios using statistical models based on historical trends.

Simulation Framework: Developed a simulation environment to model soybean growth under projected climate conditions, assessing phenotypic performance.

GAPPS Model: Implemented a Genetic Algorithm for Predicting Performance of Soybean (GAPPS) to evolve optimal phenotypic traits over successive generations, selecting for resilience under changing climates.
Evaluation Metrics: Assessed model performance using metrics such as fitness scores, convergence rates, and trait stability across simulated generations.

Results

The neural network predicted key growth and yield parameters over a 30-year period. Plant height showed moderate variability (43–54 cm), while pod counts averaged around 79, occasionally peaking above 85. Biomass remained stable between 222 and 235 units. Sugar content and relative water content fluctuated, reflecting physiological responses or environmental influences. Chlorophyll A and B levels varied, with notable increases in Chlorophyll B during certain years. Protein percentage remained stable between 32% and 39%, consistent with healthy nutritional quality. Seed yield exhibited variability, ranging from 35 to 51 units, with several peak years indicating favorable conditions.

Discussion

The neural network model effectively captures the dynamics of key agronomic traits over time. Stable plant height and pod numbers suggest consistent growth patterns under simulated conditions. Variations in sugar content and relative water highlight physiological adjustments impacting biomass and yield. Fluctuations in chlorophyll content reflect photosynthetic efficiency changes, correlating with seed yield variations. Protein stability indicates consistent crop nutritional quality, essential for agricultural value. These outcomes validate the model’s utility for forecasting soybean performance and guiding agricultural management strategies.

Conclusion

This study confirms the effectiveness of the neural network model in predicting critical growth and yield parameters for soybeans over an extended period. The model’s potential ability to capture phenotypic variability and maintain trait stability under changing conditions supports its possible usage as a decision-making tool in crop management and breeding. Future work should integrate more environmental variables to further improve prediction accuracy and foster sustainable agriculture.

References

BMC Plant Biology. (2021). Environmental and genetic regulation of plant height in soybean. *BioMed Central.*

ScienceDirect. (2019). Effect of drought stress on sugar metabolism in leaves and roots of soybean. *Scientia Horticulturae, Elsevier.*

Frontiers in Plant Science. (2020). Characterization of Photosynthetic Phenotypes and Chloroplast Ultrastructure in Soybean Genotypes. *Frontiers in Plant Science.*

Frontiers in Plant Science. (2023). Effects of high night temperature on soybean yield and compositions. *Frontiers in Plant Science.*

Theoretical and Applied Climatology. (2023). Rainfall interception loss as a function of leaf area index and rainfall by soybean. *Springer.*

Data Sources

SoyURT: Soil and Weather Environment Playground Data (Ottawa, KS) – GitHub Repository for experimental design and field layout.

Advanced Soybean Agricultural Dataset (2025) – Kaggle dataset providing phenotypic and agronomic measurements.

FAOSTAT: Crops and Livestock Products (QCL) – *Food and Agriculture Organization (FAO)*: Provides comprehensive global data on crop production, yield, and harvested area, essential for agricultural research and analysis.

National Weather Service (NWS) – *National Oceanic and Atmospheric Administration (NOAA)*: Offers extensive weather data, forecasts, and historical climate information across the United States.