GAPPS

Alja Eremić, Luka Jack Komljen, Aleksandra Zečević, Vedran Gvozderac
Wizards of OS

About the name:

Genetic **A**lgorithm for **P**redicting **P**erformance of **S**pecies *Glycine Max*

and recommending favorable phenotypes

01 GAPPS Idea

Genetic Algorithm

Base Phenotypes

Plant Height
Bio weight
Relative water
Leaf Area Index

fitness

Yearly World Representation

Soil metrics
Temperature
Wind Speed

Future Phenotypes

Plant Height
Bio weight
Relative water
Leaf Area Index

Methodology

US Location of our study is in Ottawa Kansas

How GAPPS (**tries to**) Predicts the Future

Steps:

1. 📅 Gather past 15 years of weather & soil data
2. Simulate 30 years of future climate (linear past trends + randomization)
3. 🌿 Feed into genetic algorithm
4. 🧬 Select most resilient phenotype combinations for a predicted year via **fitness function**
5. Output: future-optimized soybean phenotype profile

Implementation

02

- Base Phenotypes
- World Representation
- Fitness Evaluation
- Genetic Algorithm

Base Phenotypes

Why start with Soybeans?

- Globally crucial: food, feed, fuel
- Rich research background & phenotype data
- Highly sensitive to environmental changes
- Perfect candidate for climate-smart breeding

Trait Name	Unit
Plant Height	cm
Number of Pods	count
Biological Weight	g
Sugar Percentage	%
Relative Water Content	%
Chlorophyll A	mg/g
Chlorophyll B	mg/g
Leaf Area Index	m^2/m^2
Seed Yield per Area	kg/ha

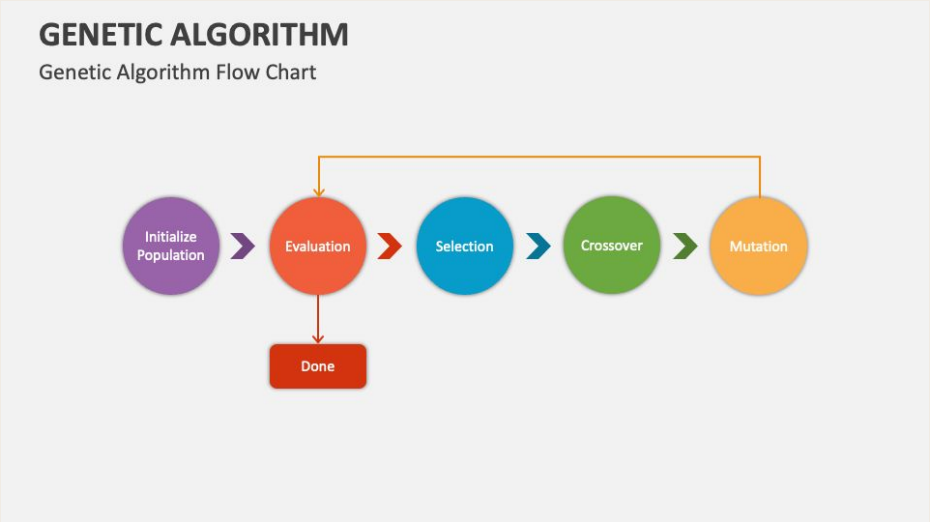
World Representation

🌦️ Weather Variables Table

Variable Name	Unit	Description
Average Temperature	°C	Daily mean air temperature
Maximum Temperature	°C	Highest recorded temperature in a day
Minimum Temperature	°C	Lowest recorded temperature in a day
Rainfall	mm	Total daily precipitation
Wind Speed	m/s	Average daily wind speed
Relative Humidity	%	Daily average humidity level
Solar Radiation	W/m²	Daily solar energy reaching surface

Genetic Algorithm

- Different plant phenotypes against changing conditions of future years
- LEADSM (Lightweight Evaluation Algorithm for Data-driven Selection Models)

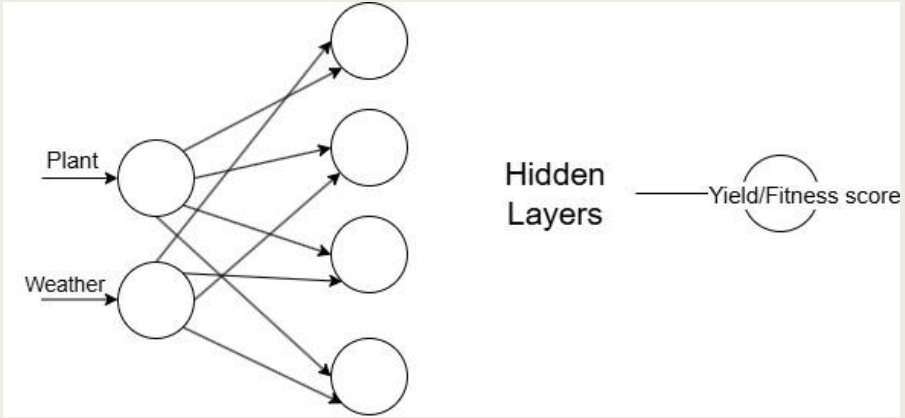


Fitness Evaluation

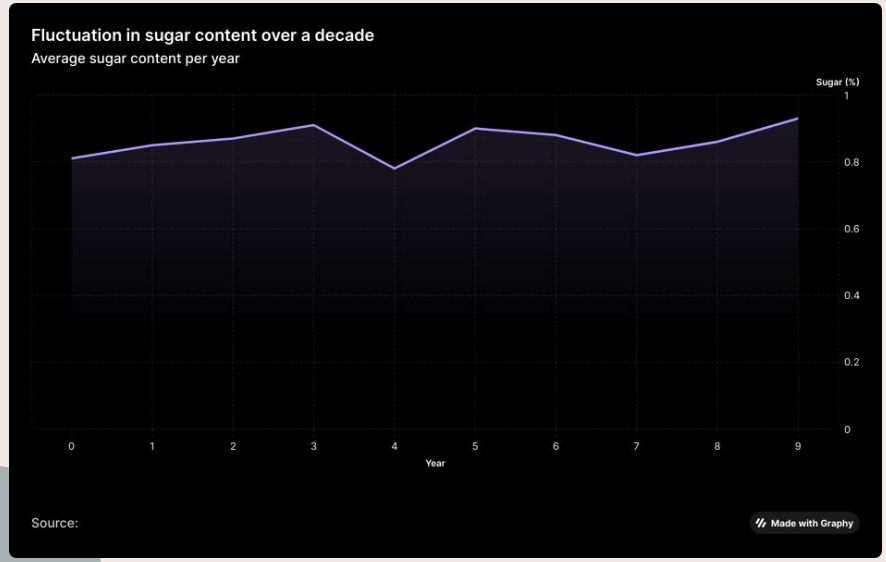
- Needs to model real world survivability of plants
- Implementations
  - Trained Neural Network (more data the better)
  - How do world parameters, and the plants phenotype affect its survival

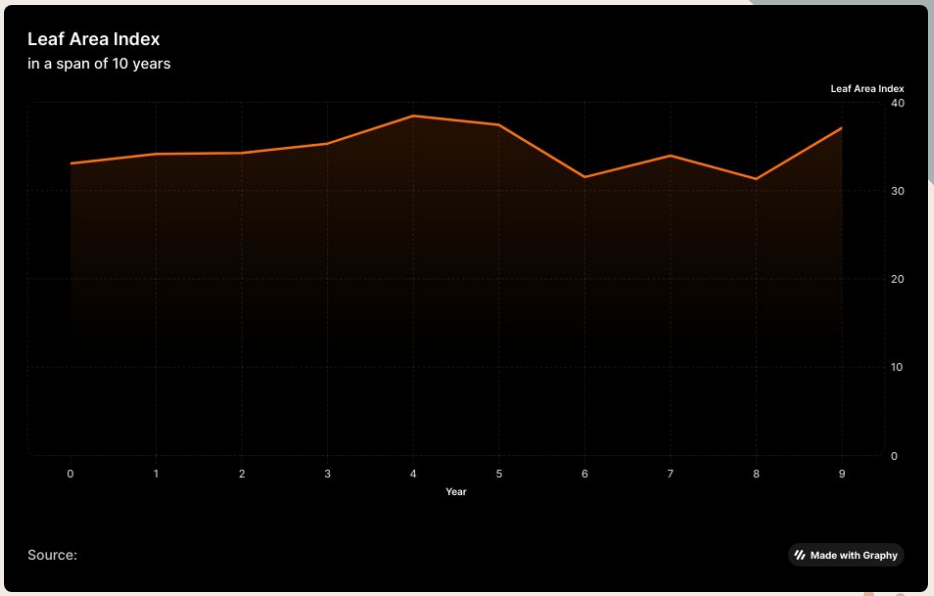
Trained fitness evaluator

- AI model trained using Deep Learning Java library with PyTorch as an engine
- Training data used plant values and weather conditions in that year as the input and the total yield that year for that plant for the output



03 Results


Fluctuation in sugar content over a decade
Average sugar content per year
Sugar (%)
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
Year
Source:
Made with Graphy


Leaf Area Index
in a span of 10 years
Leaf Area Index
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
Year
Source:
Made with Graphy

Prediction of favorable trades in next 5 years

Year	Plant Height	Number of Pods	Biological Weights	Sugar	Relative Water	Chloro A	Chloro B	Protein Percentage	Leaf Area Index	Seed Yield
0	53.3186	79.1753	227.3939	0.9839	0.7111	4.2415	3.7922	35.5941	39.6696	35.2828
1	51.6096	79.1753	227.3939	0.9839	0.7111	2.8579	1.8584	33.8983	39.6696	35.2828
2	51.6096	79.1753	233.3068	1.0610	0.8692	2.8579	3.3464	32.7704	39.6696	35.1495
3	51.6096	79.1753	227.3939	0.9839	0.7111	2.8579	3.7922	35.5941	32.3354	47.9049
4	51.6096	79.1753	227.3939	0.9839	0.4692	2.8579	1.8584	32.7704	39.6696	48.4300
5	47.0711	79.1753	227.3939	0.3084	0.7664	2.8579	7.0720	32.8206	39.6696	35.2828

04 Conclusion

Over the 30-year period, key plant growth parameters such as **Plant Height**, **Number of Pods**, and **Biological Weights** remained relatively stable, indicating consistent growth conditions.

Plant Height has shown a slight decrease, indicating a possibility that shorter plants may be more favourable in upcoming years.

Relative water has also shown a slight increase overall, indicating that water retention may be a favorable trade in the future as well, which may be caused by raising temperatures.

Seed Yield showed fluctuations, with noticeable peaks around Years 3, 6, and 14, indicating periodic increases in productivity that may relate to optimal weather conditions.

04 Conclusion

Despite these observations, predicting favorable future phenotypes remains **challenging**. This is primarily due to:

- Limited dataset size and duration.
- Poor correlation between variables.
- Difficulty in accurately simulating fitness evaluations.

Moreover, the relatively **stable** conditions over this 30-year span may not exert strong selective pressures, thus reducing the need for significant phenotypic shifts (and prediction of them) during this period.

Resources

- BMC Plant Biology, 2021
- ScienceDirect, 2019
- Frontiers, 2020
- SoyURT GitHub
- Kaggle Dataset (2025)
- FAOSTAT
- NWS

Thanks!

Does anyone have any questions?