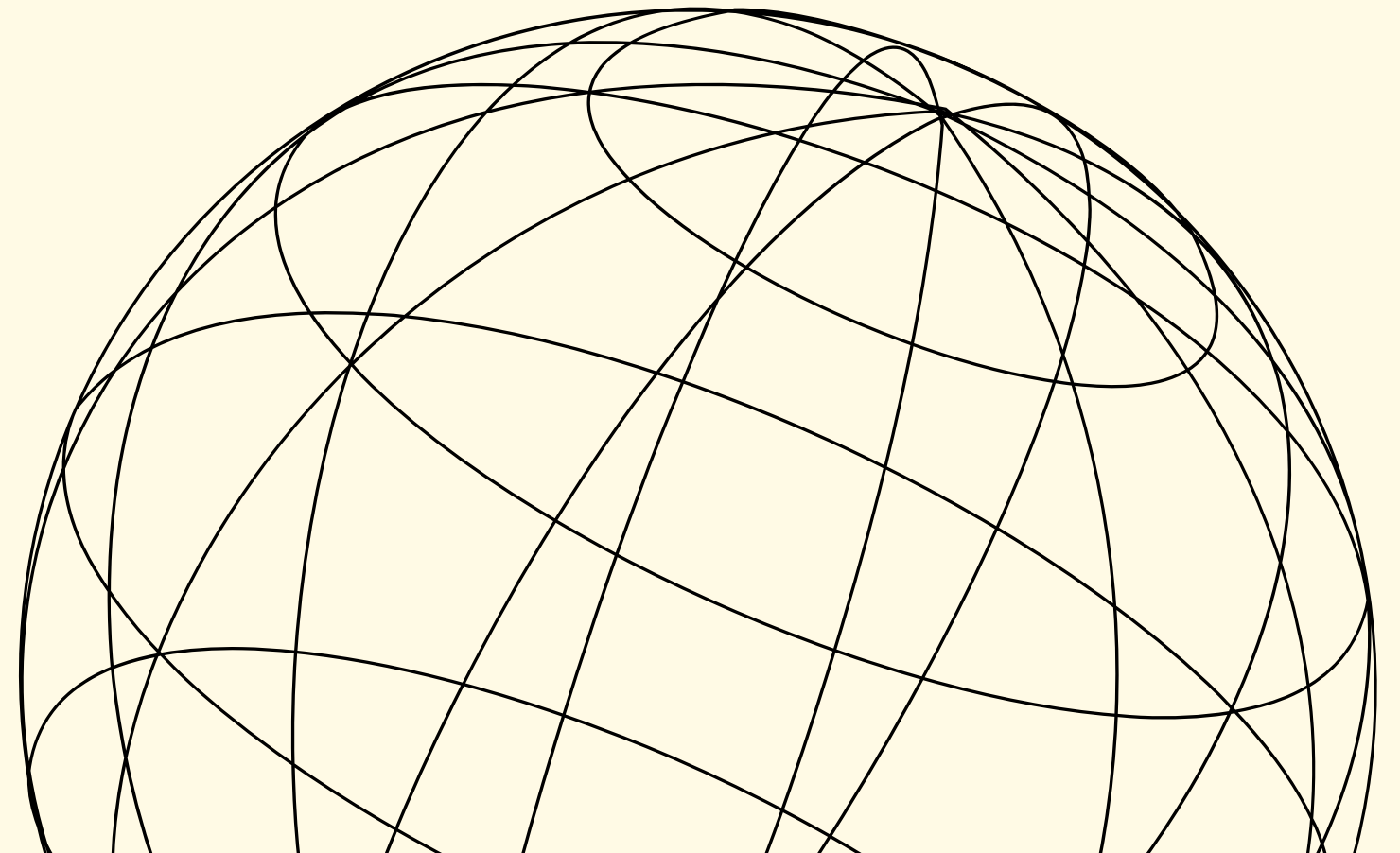


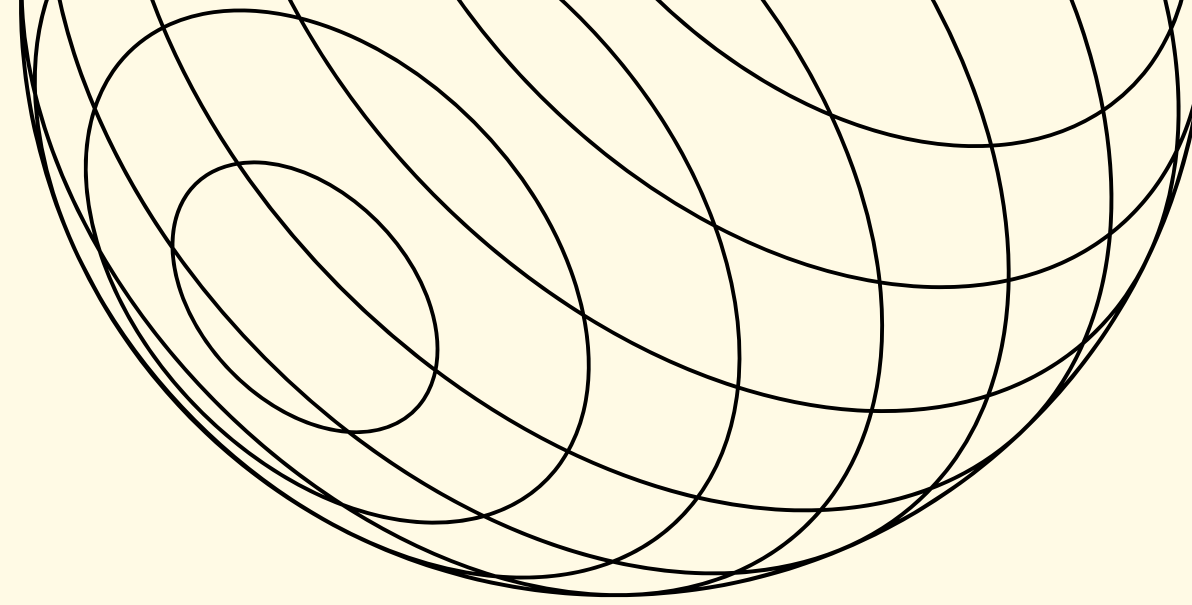
# SEED QUALITY CLASSIFICATION

Agricultural Data Science

Jacob Serfaty



# SEED QUALITY INDUSTRY



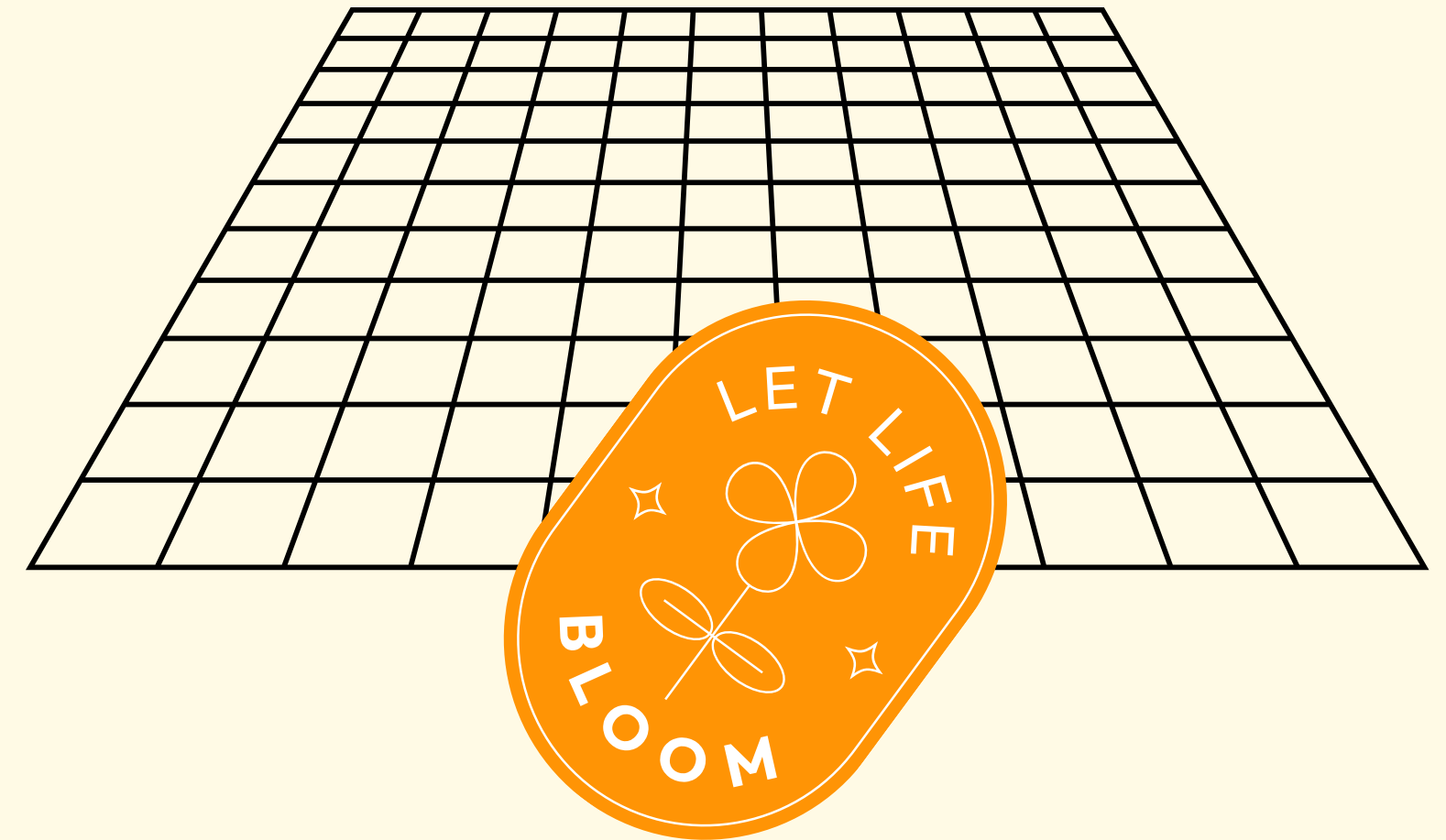
*“Good seed is the  
foundation of good  
crops.”*

*- Norman Borlaug*

- High-quality soybean seeds have the potential to result in increased crop yield.
- Quality seeds have a higher resistance to common soybean diseases and pests.
- High-quality seeds often contain desirable genetic traits, such as tolerance to specific environmental conditions (e.g., drought or heat tolerance)

# BUSINESS UNDERSTANDING

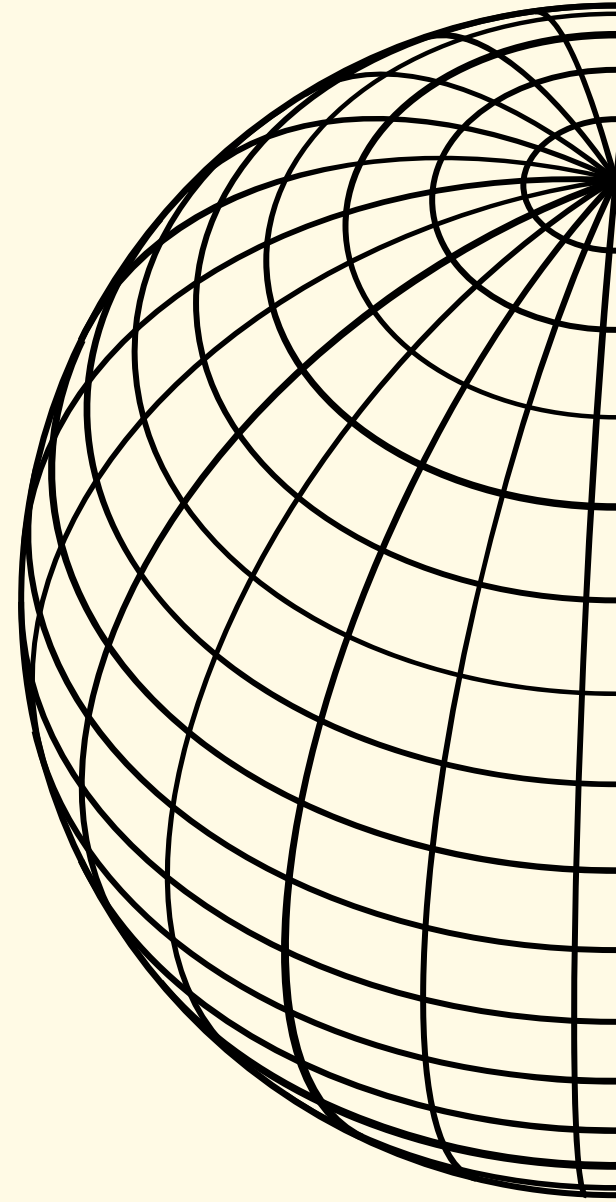
*Our goal is to improve the efficiency of agricultural tech businesses, by improving the quality of the soybean seeds in which they use.*



- Collect soybean seed images from agricultural tech businesses in the soybean market to put into a classification model
- Create a model that can distinguish between viable and non-viable seeds
- Integrate this model into agricultural hardware for sorting seeds

# DATA UNDERSTANDING

- The dataset consists of 5513 images of soybean seeds
- The data was equally balanced
- Each image is 227x227 pixels
- There are 5 classes
  - Intact
  - Broken
  - Immature
  - Skin-damaged
  - Spotted





# DATA UNDERSTANDING

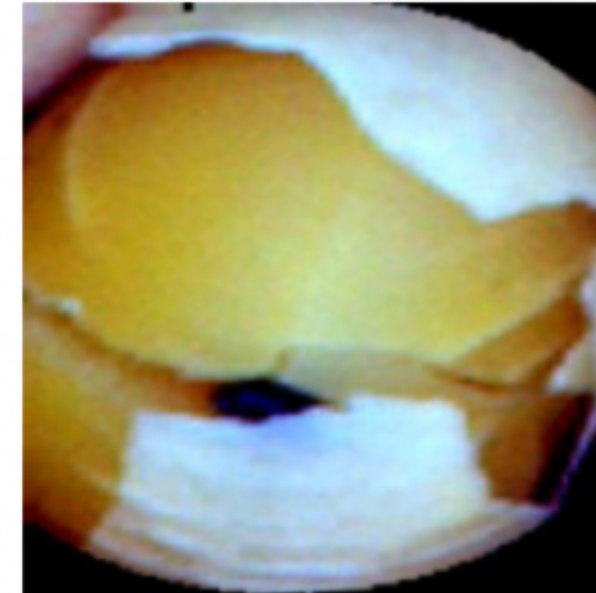
True: Immature Predicted: Immature



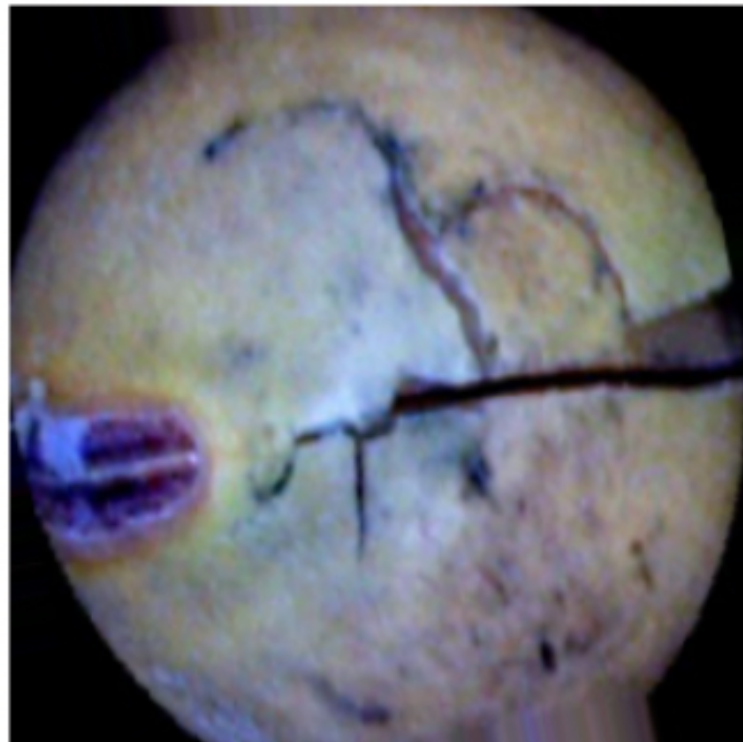
True: Intact Predicted: Intact



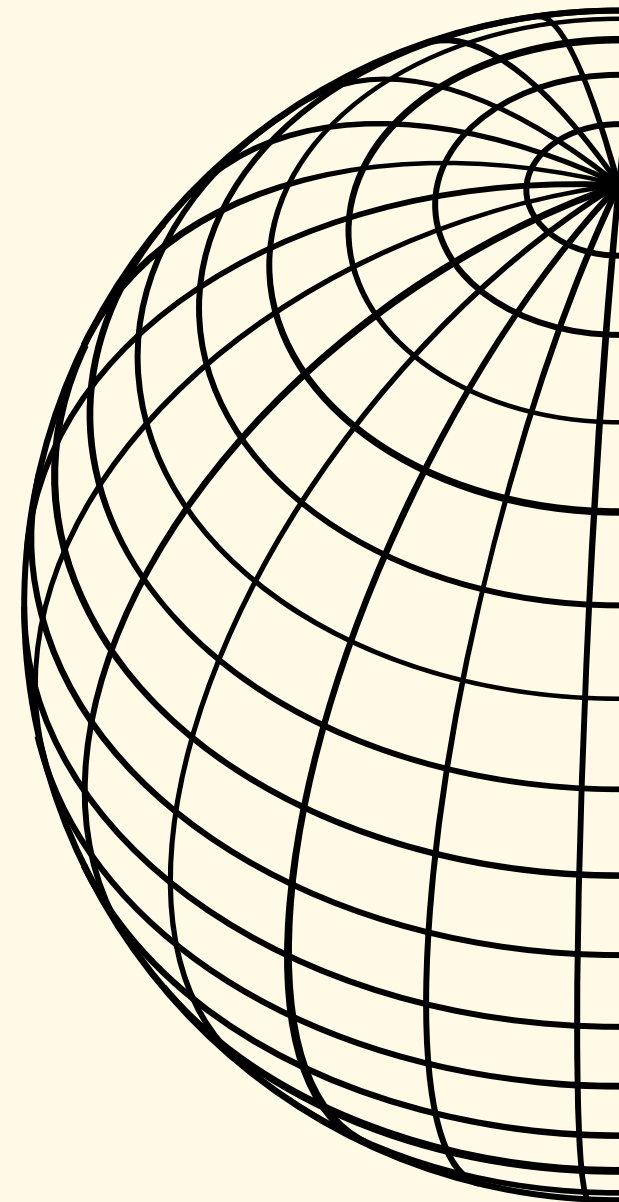
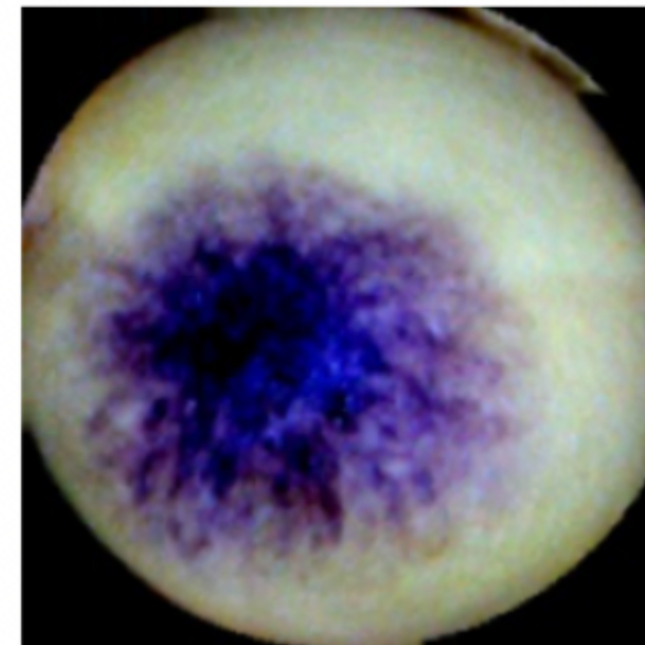
True: Skin-damaged Predicted: Skin-damaged



True: Broken Predicted: Broken

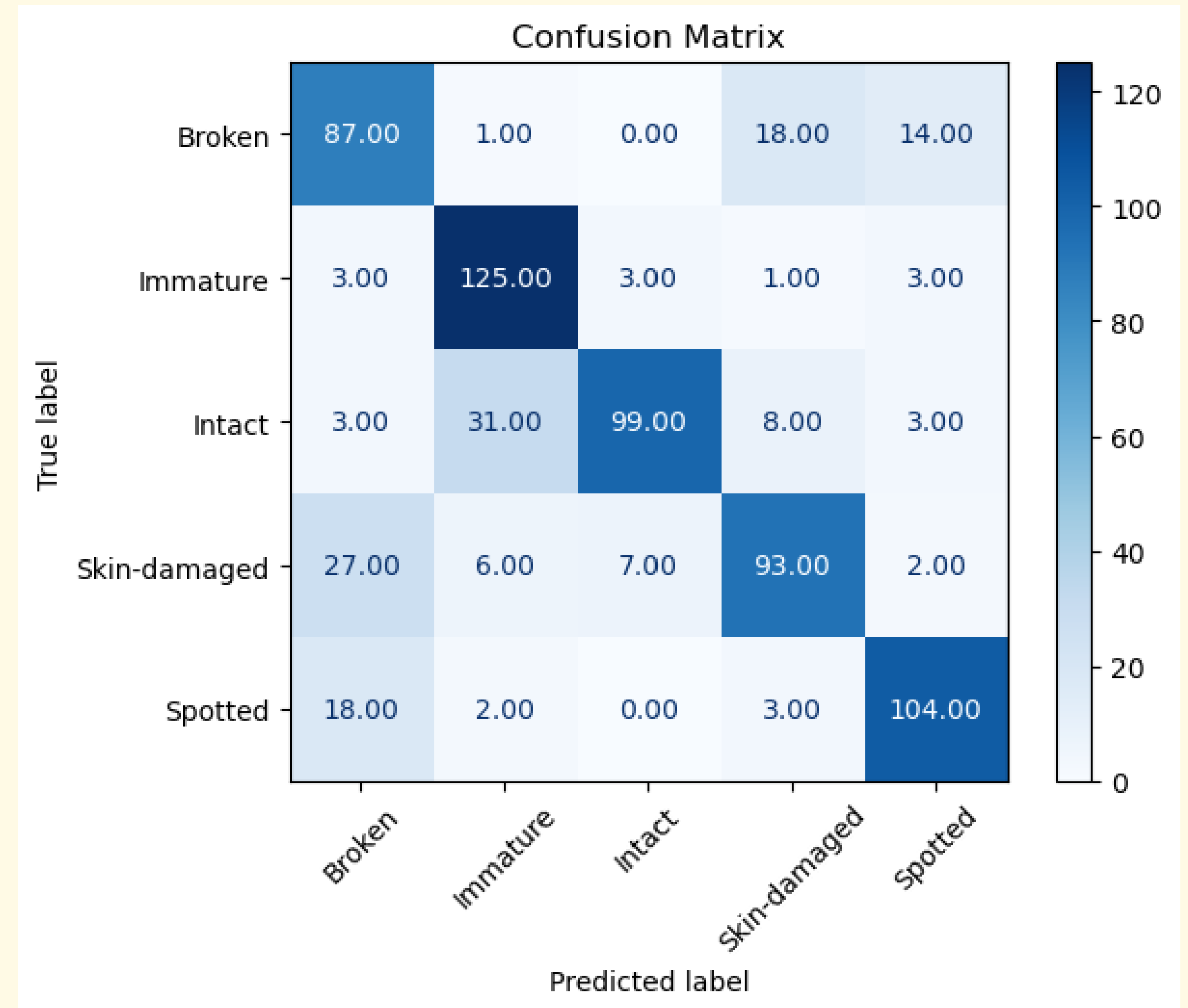


True: Spotted Predicted: Spotted



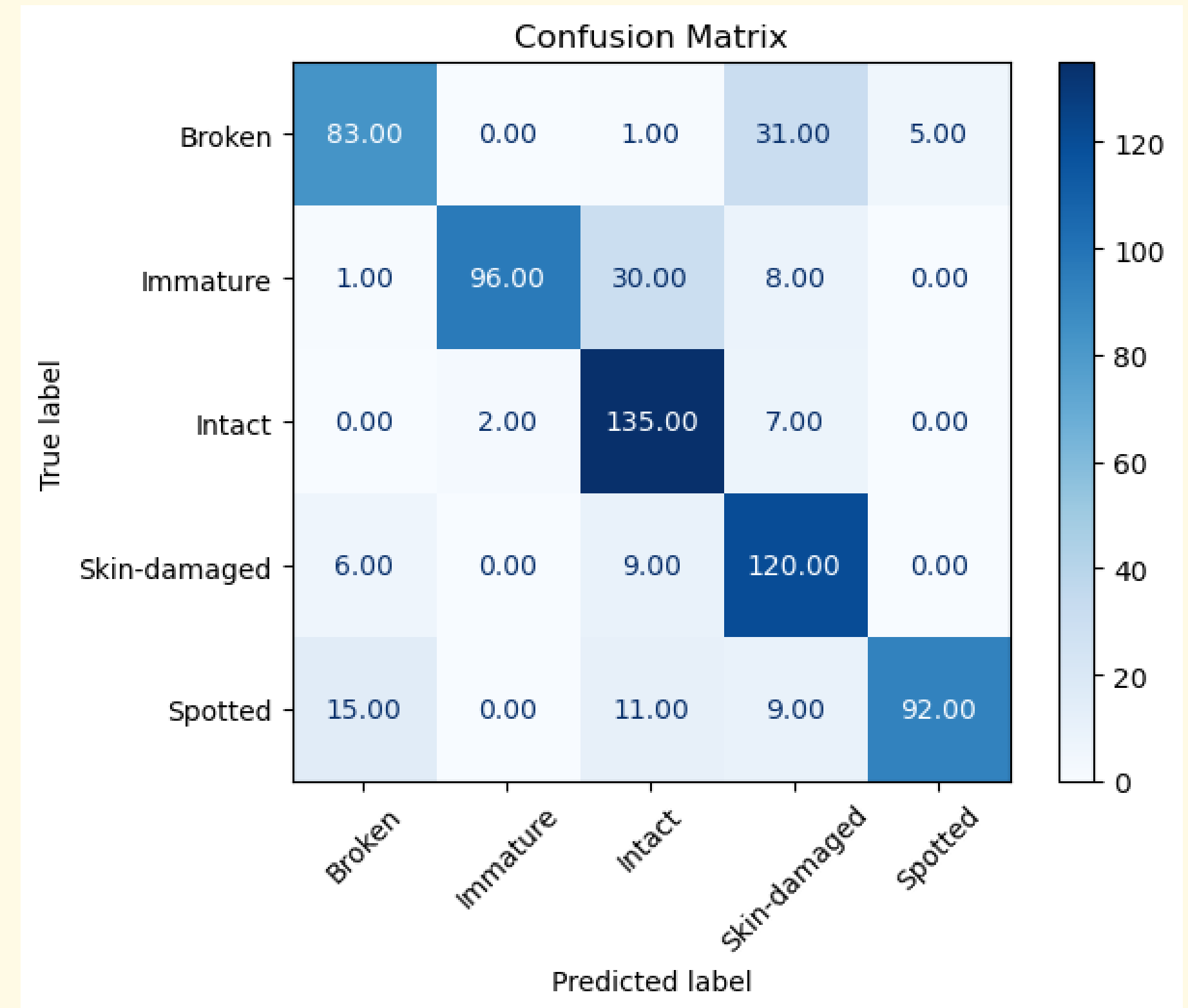
# MODELLING

- All of the models were Convolutional Neural Networks (CNNs)
- The base model layout:
  - 3 2D-convolutional layers
  - 3 max-pooling layers
  - 1 hidden layer



# MODELLING

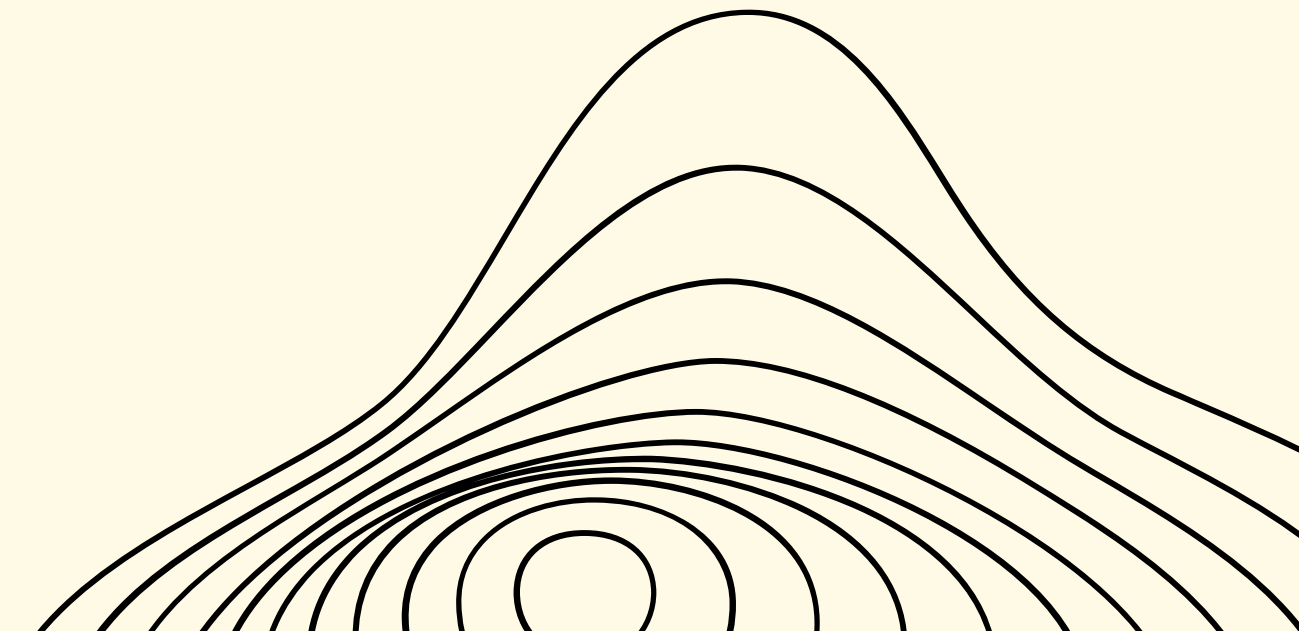
- The final model layout:
  - 3 2D-convolutional layers
  - 3 max-pooling layers
  - 3 hidden layers
  - 2 dropout layers
- This model also included:
  - Bias
  - Padding
  - L2 regularization



# BUSINESS RECOMMENDATIONS

- The model that should be used for soybean seed identification is the final classification model
- The model should be integrated into a seed-sorting mechanism that can classify each seed in real-time.

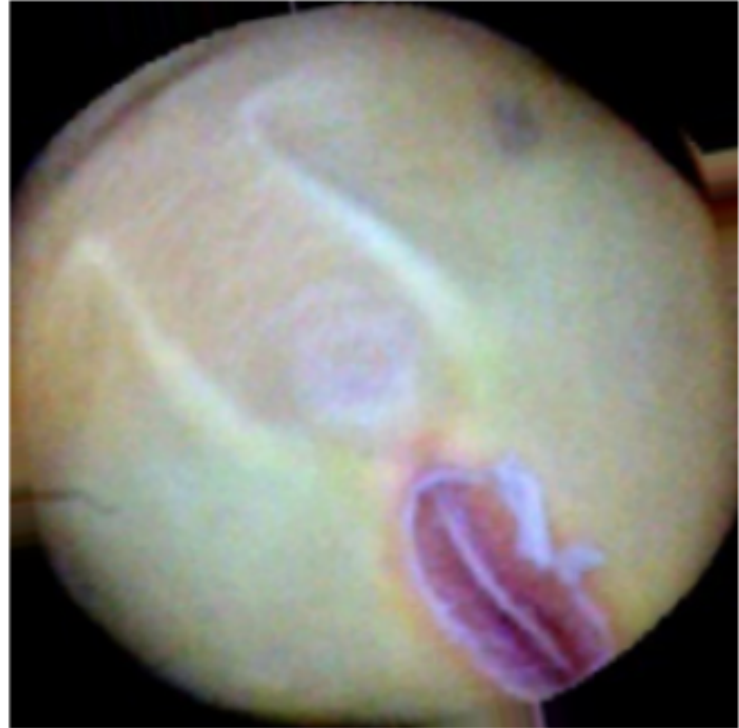
- *The model needs data with higher-quality color images*
- *The model needs data that is not mislabelled*





# MODEL IMPROVEMENTS

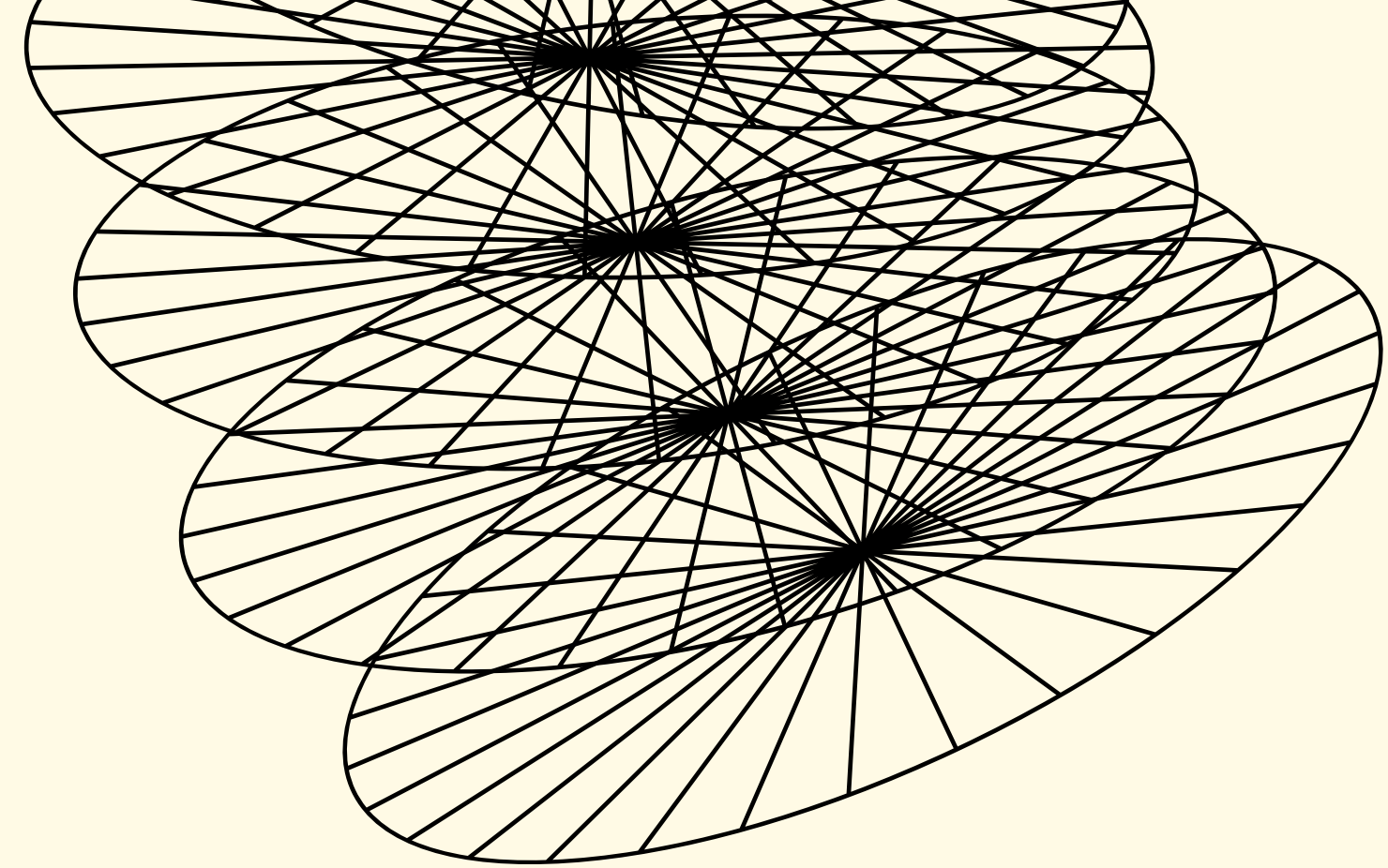
True: Immature Predicted: Intact



True: Broken Predicted: Skin-damaged



- *The model can be improved through transfer learning models trained for seed identification*
- *The model can be improved by comparing classes, which are more heavily misclassified into a binary classification model*





# THANK YOU

[GITHUB.COM/JACOBSERFATY](https://github.com/JACOBSERFATY)

[LINKEDIN.COM/IN/JACOB-SERFATY/](https://www.linkedin.com/in/JACOB-SERFATY/)