PlantVillage Image Classification

ZEHRA KEZER

Aygaz Introduction to Artificial Intelligence Bootcamp

INTRODUCTİON

Today's agricultural sector needs new technologies in the fight against plant diseases and pests.
Artificial intelligence and deep learning have significant potential to increase efficiency and sustainability in agriculture.

- In this project, we developed an artificial intelligence model that can detect healthy and unhealthy conditions of pepper plants.
- Our model aims to provide farmers with an important diagnostic tool by accelerating disease detection processes in agriculture.

METHODOLOGY

Librarys

- Numpy
- Pandas
- Seaborn
- Matplotlib
- Sklearn
- Tensorflow

Data Collection

- Kaggle - PlantVillage Dataset

Algorithms

Deep Learning algorithms

- Convolutional Neural Network

Machine learning algorithms

- K-Nearest Neighbor(KNN)
- Random Forest
- Support vector machine
- Decision Tree

EXPLORE THE DATA

Dataset includes 880 healthy pepper leaves, and 748 unhealthy pepper leaves

Our dataset consists of two categories: visual and label. The label is divided into healthy and unhealthy.

```
DatasetDict({
    train: Dataset({
        features: ['image', 'label'],
        num_rows: 1628
    })
})
```

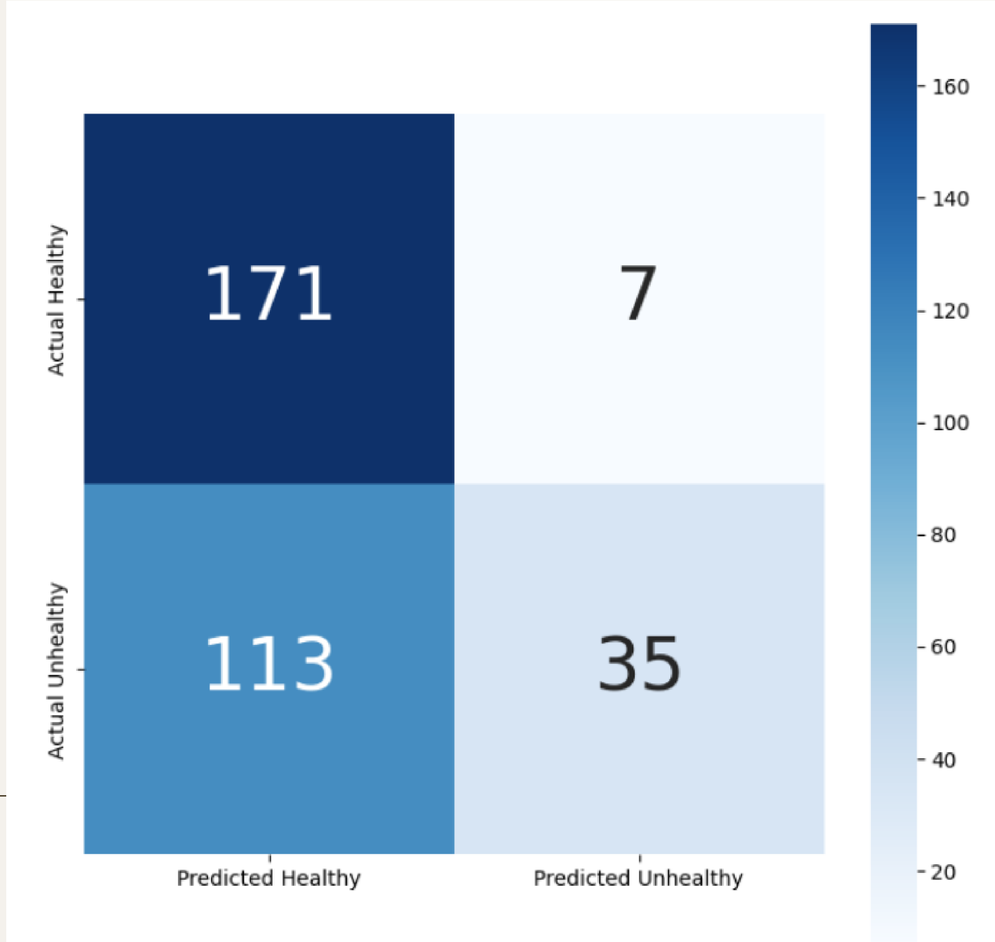
MODEL TRAİNİNG

After separating our images into train and test, we normalized them and trained them with different algorithms to check which one fits our model.

K-Nearest Neighbor(KNN)

Our model trained with KNN has an accuracy of 0.63. As we see in the confusion matrix, although it knows the healthy ones well, it predicts the unhealthy ones very poorly.

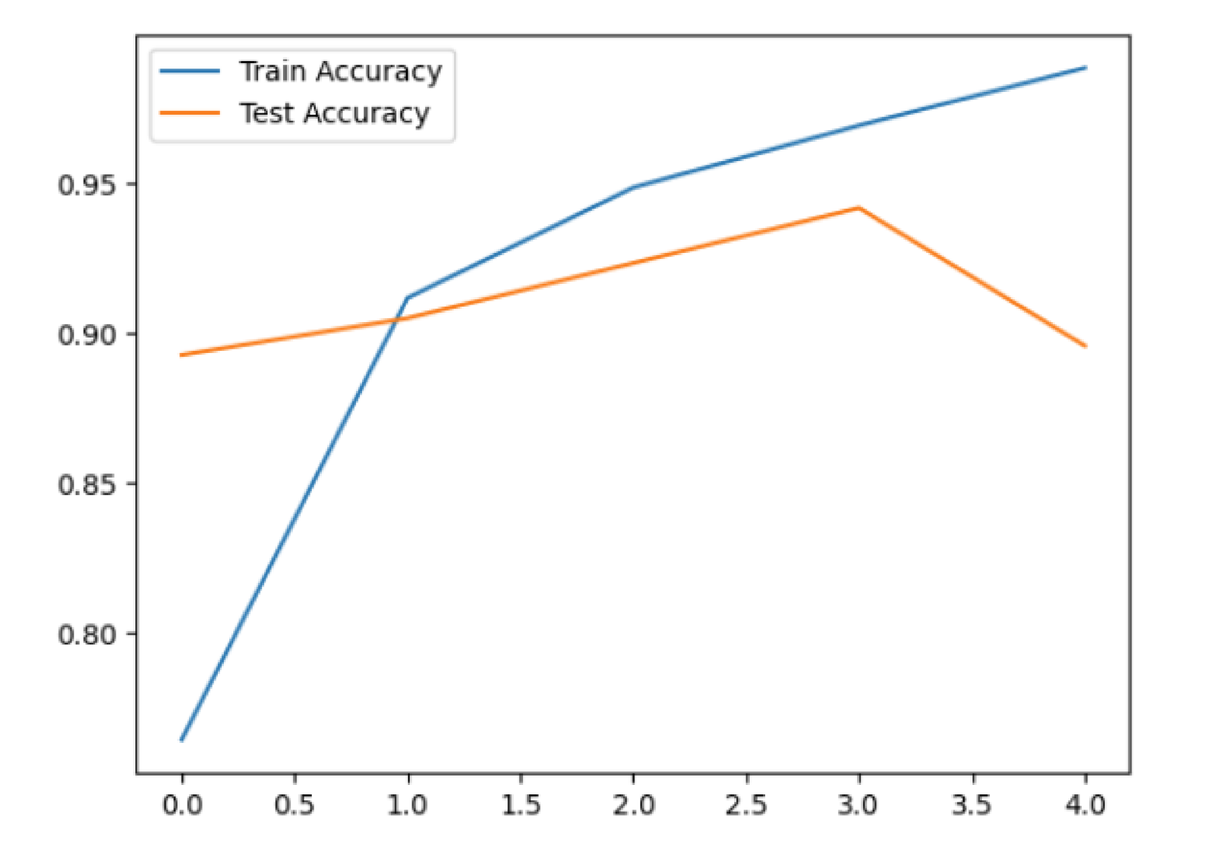
Accuracy: 0.6319
F1 Score: 0.5714
Recall: 0.6319
Precision: 0.7071



Convolutional Neural Network(CNN)

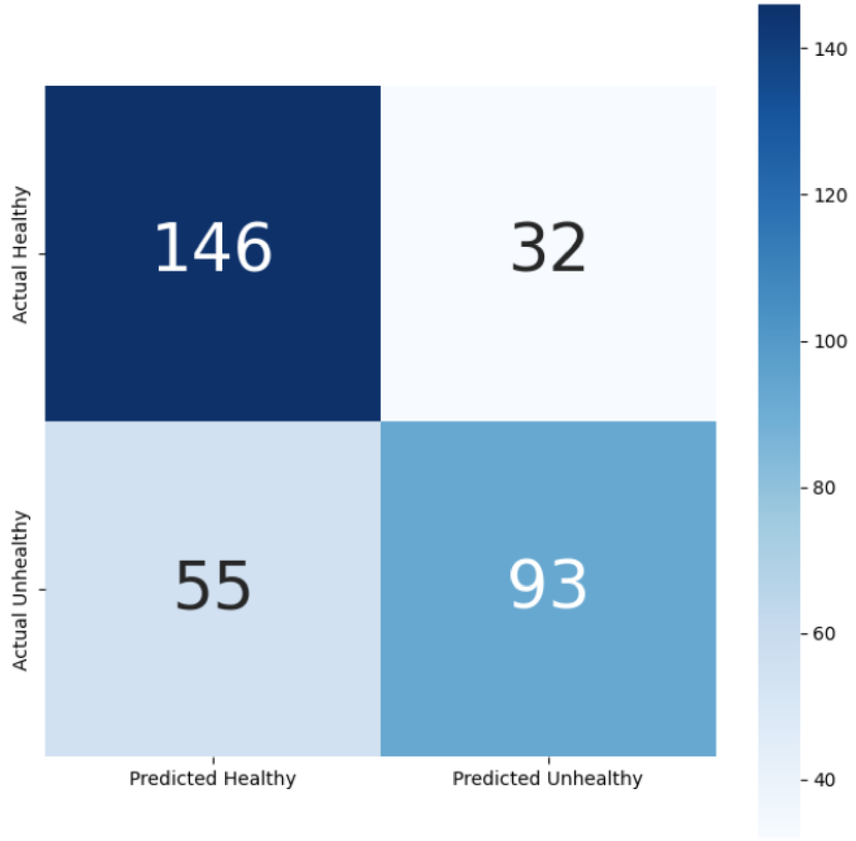
Epoch 1/5 loss: 1.1098 - accuracy: 0.7642 - val_loss: 0.2528 - val_accuracy: 0.8926
Epoch 2/5 loss: 0.2232 - accuracy: 0.9117 - val_loss: 0.2020 - val_accuracy: 0.9049
Epoch 3/5 5s/step - loss: 0.1398 - accuracy: 0.9485 - val_loss: 0.1604 - val_accuracy: 0.9233
Epoch 4/5 4s/step - loss: 0.0938 - accuracy: 0.9693 - val_loss: 0.1285 - val_accuracy: 0.9417
Epoch 5/5 loss: 0.0501 - accuracy: 0.9885 - val_loss: 0.2510 - val_accuracy: 0.8957

Increasing the number of Epochs too much may cause the model to overfit the training data. 4 epochs seems ideal for this data.



Decision Tree

Accuracy: 0.7331
F1 Score: 0.7300
Recall: 0.7331
Precision: 0.7344

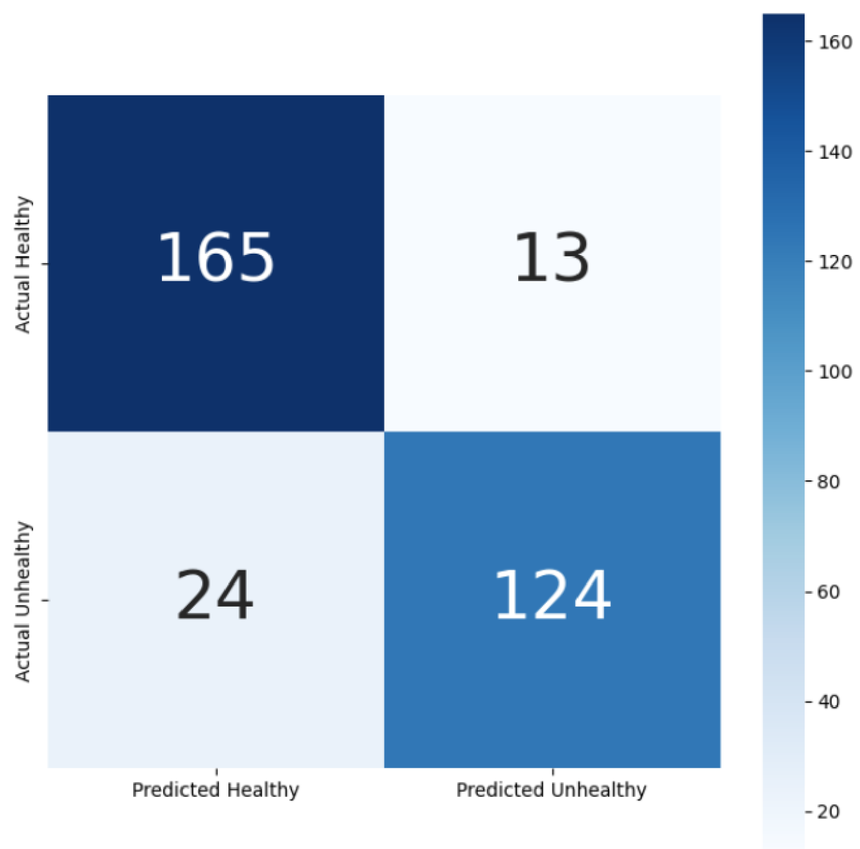


The higher accuracy, F1 score, recall, and precision indicate that the Random Forest model is more reliable and effective in distinguishing between healthy and unhealthy cases.

This is visually supported by the confusion matrices, where the Random Forest has fewer false positives and false negatives compared to the Decision Tree and KNN.

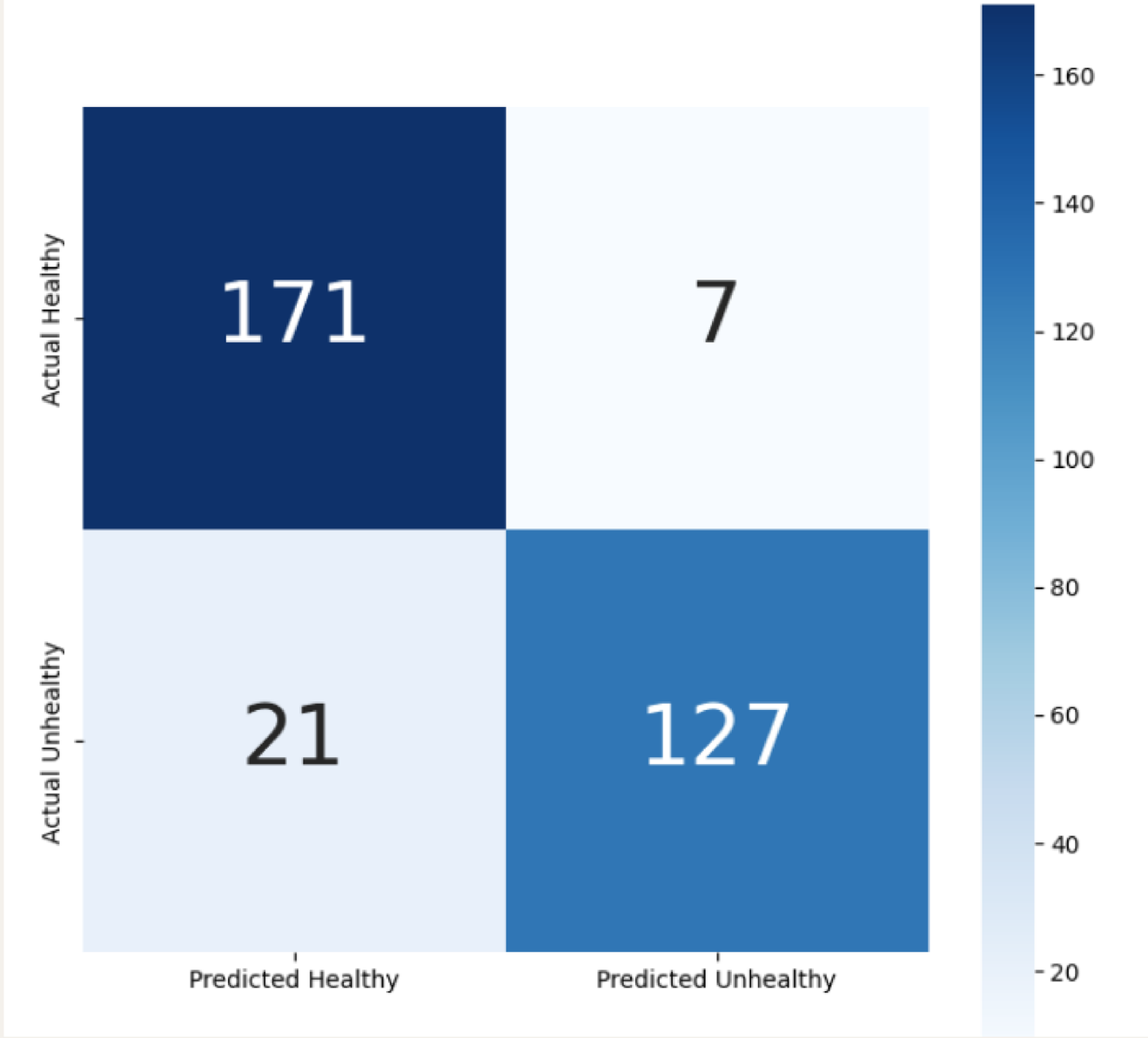
Random Forest

Accuracy: 0.8865
F1 Score: 0.8860
Recall: 0.8865
Precision: 0.8876



Support vector machine

Accuracy: 0.9141
F1 Score: 0.9136
Recall: 0.9141
Precision: 0.9166



- High Performance: our model's accuracy, precision, recall and F1 score values are all over 90%. This indicates that our model performed quite well on our dataset and succeeded in the classification task.
- The best confusion matrix result is in SVM

The Support Vector Machine model has the potential to be used in disease diagnosis and improving product quality in the field of agriculture.

CONCLUSION

- In this project, an artificial intelligence model that can detect healthy and unhealthy conditions of pepper plants was developed. The developed model was evaluated using various performance metrics and satisfactory results were obtained.

- We achieved the best result in CNN, but if we want to use a machine learning method, SVM will be the best option.

- The performance of the model is directly related to the quality and diversity of the data set. Future work may aim to further improve the overall performance of the model by using larger and more diverse data sets.


THANKS FOR LISTENING