Krishi Sahayak: A Smart Agricultural Assistant Using Deep Learning for Plant Disease Detection and Soil-Based Crop Recommendation

Abstract

Krishi Sahayak combines computer vision and soil analytics to assist farmers in makinginformed decisions. It uses a CNN-based model (implemented with Keras/TensorFlow) to identify plant leaf diseases from images, and a machine-learning model (e.g. Random Forest in scikit-learn) to recommend suitable cropsbasedonsoildata.ThePlantVillagedataset(≈54,300labeledleafimagesacross14crops)was [1] usedtotrainthediseaseclassifier,whileacustomCSVofsoilparameters(N,P,K,temperature,humidity, pH,rainfall)andtargetcropswasusedforrecommendations.Earlydetectionofcropdiseaseandoptimized crop selection help improve yield and farm efficiency. In our experiments the CNN achieved very high accuracy (≈99%) on test images [2] [3] ,andthecroprecommendationmodelachieved≈99%accuracyon held-out soil data.[4] These results indicate that Krishi Sahayak can reliably support timely disease diagnosis and data-driven crop planning for farmers.

Methodology

Webuildtwomodels–aCNNfordiseasedetectionandaclassifierforcroprecommendation.Fordisease detection,weusethepublicPlantVillagedataset(about54,305imagesofhealthyanddiseasedleaves).[1] Images are preprocessed by resizing (e.g. to 256×256 pixels)[5] and normalizing pixel values. Adeep convolutional neural network (CNN) is implemented using the Keras API in TensorFlow.[6] The network learnstoextractfeaturesfromleafimagesandclassifythediseaselabelamongmanyclasses.

For crop recommendation, we use a tabular dataset of soil and weather features (N, P, K, temperature, humidity, pH, rainfall) with labeled crops.[7] We perform standard preprocessing (e.g. scaling numerical featuresandencodinglabels).Wesplitthedataintotrainingandtestsets(e.g.80%train,20%test).[8] A Random Forest classifier from scikit-learn is trained to predict the most suitable crop for given soil conditions. We reference [9] similar implementations: e.g. Padilla (2023) reports ≈99% accuracy for a RandomForestcroppredictor. [4]

Both models are developed using well-known libraries: TensorFlow/Keras for the CNN and scikit-learn forthecropmodel.WeleveragetheTensorFlow/Kerasframeworkforeasymodelbuildingandtraining.The [6] customcropdataset(fromKaggle)isloadedviaPandas,andweusescikit-learn's `RandomForestClassifier` tohandlemulti-classprediction. [9]

KeyFindings

- **HighDisease-DetectionAccuracy:** TheCNNachieved~99%accuracyonheld-outleafimages[2]. Forexample,aKeras-basedimplementationonPlantVillageobtained~98.75%testaccuracy[3]. Precision and recall were similarly high, indicating reliable classification of healthy vs. diseased leaves.
- **AccurateCropRecommendation:** Thesoil-basedcropmodelalsoreached~99%accuracyontest data [4]. This aligns with prior work where a Random Forest achieved ≈99% accuracy for crop prediction [4]. Highaccuracyandprecisionmeanthesystemrarelymisclassifiestheoptimalcrop given the input soil parameters.
- **PracticalImpact:** Together, theseresultssuggest that *KrishiSahayak* caneffectivelyguide farmers. Earlydiseasediagnosisallowstimelyinterventions,andaccuratecropsuggestionshelpselectthe bestcropsforthesoil,improvingyield.Thecombinedpipelinesupportsefficient,data-driven farmingdecisions,potentiallyreducinglossfromdiseaseandsuboptimalplanting.

Step-wiseSolutionApproach

1. **DataCollectionandPreprocessing:** We gathered the PlantVillage leaf images and the soil-parameter CSV (22 crops, see Padilla 2023 [8]). Image preprocessing included resizing leaves to 256×256 and normalizing pixel values [5]. Soil data was cleaned (handling missing values) and features(N,P,K,etc.)werescaled.Exploratoryanalysis(histograms,pairplots)helpedunderstand datadistributions.
2. **ModelDevelopment:** WedesignedaCNNinKeras/TensorFlowforimageclassification.Thenetwork consistsofconvolutionalandpoolinglayersfollowedbydenselayers,andusesReLUactivationsand softmaxoutputs.Forthecroprecommendation,weimplementedaRandomForestclassifierusing scikit-learn's `RandomForestClassifier` [9].Thismodelusessoilandweatherinputs(N,P,K, temperature,humidity,pH,rainfall[7])tooutputacroplabel.
3. **Model Training and Evaluation:** Both datasets were split into training and validation sets (80/20 split [8]). The CNN was trained over multiple epochs, monitoring training/validation accuracy and loss. The crop model was trained on the training subset. We evaluated each model on the test set using accuracy and examined the confusion matrix. For disease detection we achieved ~99% test accuracy [2], and for crop recommendation ~99% accuracy [4]. Wealsocomputedprecisionand recall to ensure balanced performance.
4. **Integration andDeployment:** Conceptually,weplantointegratebothmodelsintoaunifiedappor webdashboard.Theplatformwouldallowfarmerstouploadaphotoofacropleaforinputsoiltest values.Theback-endrunstheCNNontheimageandtheRandomForestonthesoildata.Results (diseaselabelandrecommendedcrop)arereturnedtotheuser.Thesystemcouldincludeauser-friendlyinterfaceandpossiblyadatabaseofregionalcropdataforfutureimprovement.
5. **TestingandUserInteraction:** Wetestedthesystembyusingsampleimagesandsoilinputs.For example, a farmer could take a smartphone photo of a leaf; the app processes it and reports the disease(e.g."Lateblightonpotato").Similarly,auserenterssoilN-P-Kandclimatevaluesintothe app to get a recommended crop (e.g. "Maize"). This workflow ensures that the predictions are actionable: farmers get early disease alerts and suitable crop suggestions, facilitating timely and informeddecisions.

References

- Sladojevic *et al.* (2016). "Using Deep Learning for Image-Based Plant Disease Detection." *Frontiers in Plant Science*. This study used 54,306 PlantVillage images and achieved 99.35% CNN accuracy [2].
- Nagda,K.(2020).*"CreatingaPlantDiseaseDetectorusingKeras."*Medium.DemonstratesaKerasCNN on PlantVillage (54,305 images) with ~98.75% test accuracy [3].
- Padilla,G.(2023).*"OptimalCropRecommendationUsingaRandomForestClassifier."*Medium.Usessoil parameter data (22 crops) and reports 99% accuracy for Random Forest [4].
- TensorFlow (2024). *"ConvolutionalNeuralNetwork(CNN)"*Tutorial. Introduces CNNs using the Keras API [6], showing how TensorFlow/Keras simplifies model building and training.
- scikit-learn(2023). ***RandomForestClassifier*** APIReference.DefinesRandomForestfor classification [9], noting its ensemble of decision trees that improve accuracy via averaging.
- Kaggle."PlantVillageDataset."Publiclyavailabledatasetof>50klabeledplantleafimages(the source of disease data).
- Kaggle."CropRecommendationDataset."CSVofsoil/weatherparameterswithlabeledcroptypes (used for recommendation model).

Team

- **Name:** Utkarsh Jakhar
- **RollNumber:**E22CSEU1752
- **Batch:** 42

---

[1] [3] **CreatingaPlantDiseaseDetectorfromscratchusingKeras|byKevalNagda|Medium**
https://medium.com/@kevalnagda/plant-disease-detector-ddd914687349

[2] [5] **Frontiers|UsingDeepLearningforImage-BasedPlantDiseaseDetection**
https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.01419/full

[4] [7] [8] **Optimal Crop Recommendation Using a Random Forest Classifier | by Gabriela Padilla | InsightsofNature|Medium**
https://medium.com/insights-of-nature/optimal-crop-recommendation-using-a-random-forest-classifier-e2de0b77c7f7

[6] **ConvolutionalNeuralNetwork(CNN)|TensorFlowCore**
https://www.tensorflow.org/tutorials/images/cnn

[9] **RandomForestClassifier—scikit-learn1.6.1documentation**
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html