Multimodal Fusion of Leaf Images and Geospatial Data for Early Soybean Disease Prediction

Suri, Abhay (School: Linn Mar High School)

Soybean diseases cause substantial annual yield losses. Existing detection methods rely on experts or lab examinations of fields, which can be labor intensive and inaccurate, or image classification models that fail to generalize beyond controlled settings. Early warning systems for regional disease outbreaks are also limited. This project proposes a multimodal machine learning approach that integrates geospatial environmental data and soybean leaf image data to estimate soybean disease risk. A geospatial model was developed using PRISM climate data, SSURGO soil data, and Soybean Cyst Nematode and Soybean Rust outbreak data, and evaluated across Logistic Regression, Support Vector Machines, an Artificial Neural Network, and Random Forests, with a LightGBM model achieving the strongest performance, reaching approximately 0.95 across accuracy, F1-Score, precision, and recall. Additionally, a ResNet-34 Convolutional Neural Network was trained on a merged dataset of three publicly available soybean image datasets, ASDID, PDDB, and an Amity University Chhattisgarh soybean dataset, to reduce bias. This results in an approximately 0.25 improvement in accuracy, F1-Score, precision, and recall when evaluated on unseen data from a fourth dataset, SoyNet, compared to control models trained on each dataset. The outputs are then combined using a Bayesian Fusion model. The probability histogram of the fused model shows a strong peak near 1 for diseased samples that is separated from healthy predictions, indicating strong predictions with low overlap. This work gives an accessible method of detecting soybean disease, potentially aiding in timely disease treatment and reducing yield loss.

Awards Won: