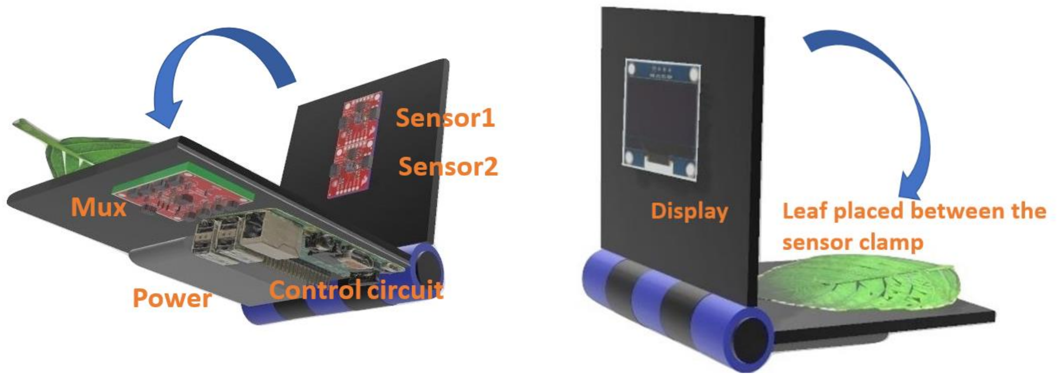
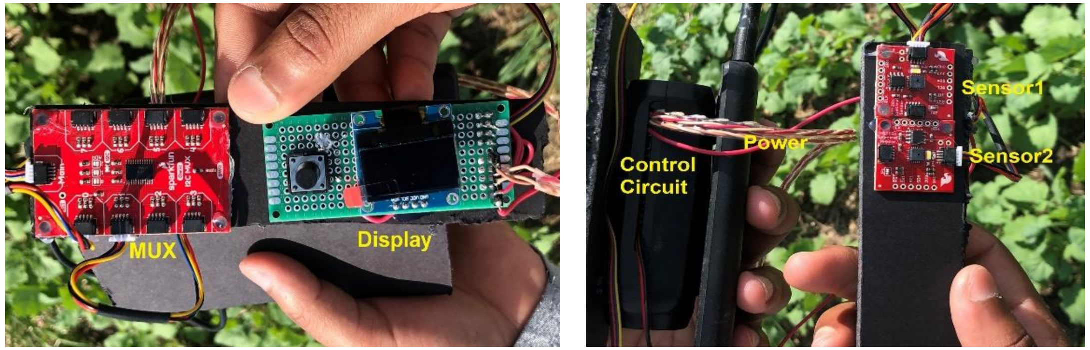
Machine Learning Based Low-Cost Multispectral Sensor for Leaf Nitrogen and Phosphorus Level Classification

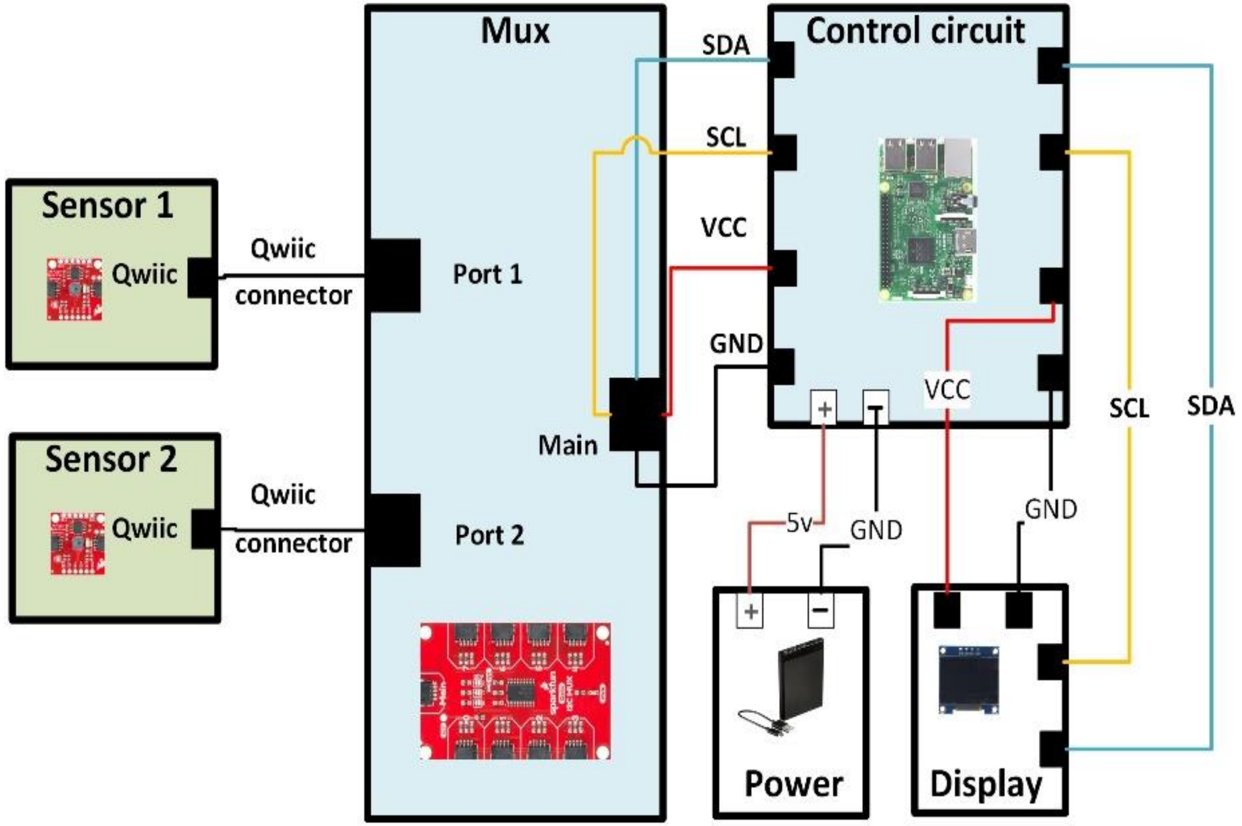
Mohammad Habibullah

Methodology

<u>Overall setup of the System:</u>







Methodology

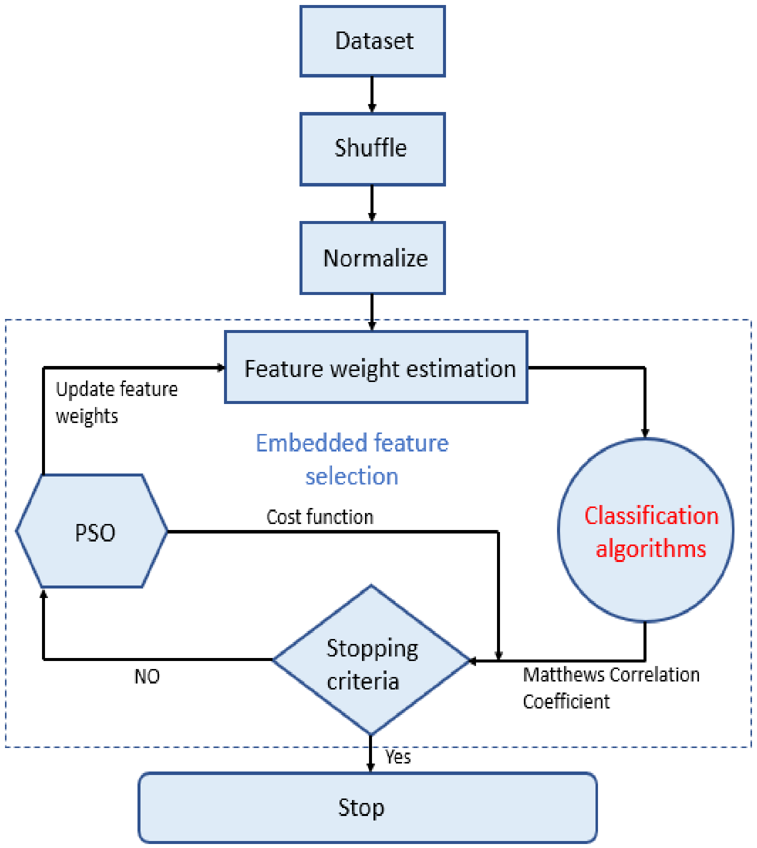
State-of-the-art Algorithms:

- K- means clustering:
  - Used in preprocessing stage
  - Euclidean distance is used
  - The center of the cluster (centroid) was determined
- Ensemble Bagged Decision Tree (EBT):
  - Used to classify N and P levels in leaves
- Support Vector Machine (SVM):
  - Used to classify N and P levels in leaves
  - Radial basis function (RBF) is used as kernel
  - RBF kernel parameters are optimized using the Wang [66]
- K-Nearest Neighbor (KNN):
  - Used to classify N and P levels in leaves
- Decision Tree (DT):
  - Used to classify N and P levels in leaves
  - C4.5 decision tree is used

Methodology

Feature Engineering:

- The features are normalized using z-score
- An independent-sample t-test is used to identify statistically discriminative features Also, embedded method is used, in which features are weighted based on the Particle Swarm Optimization (PSO) algorithm during learning
- Swarm size of the PSO is set to 200, and the maximum number of iterations is 500.
- Also, the range of the weights of the features is -5 to 5, and the tolerance limit is set to 10^{-12}.
- Matthews Correlation Coefficient (MCC) is used as a cost function to optimize the feature weights



Methodology

Model Validation:

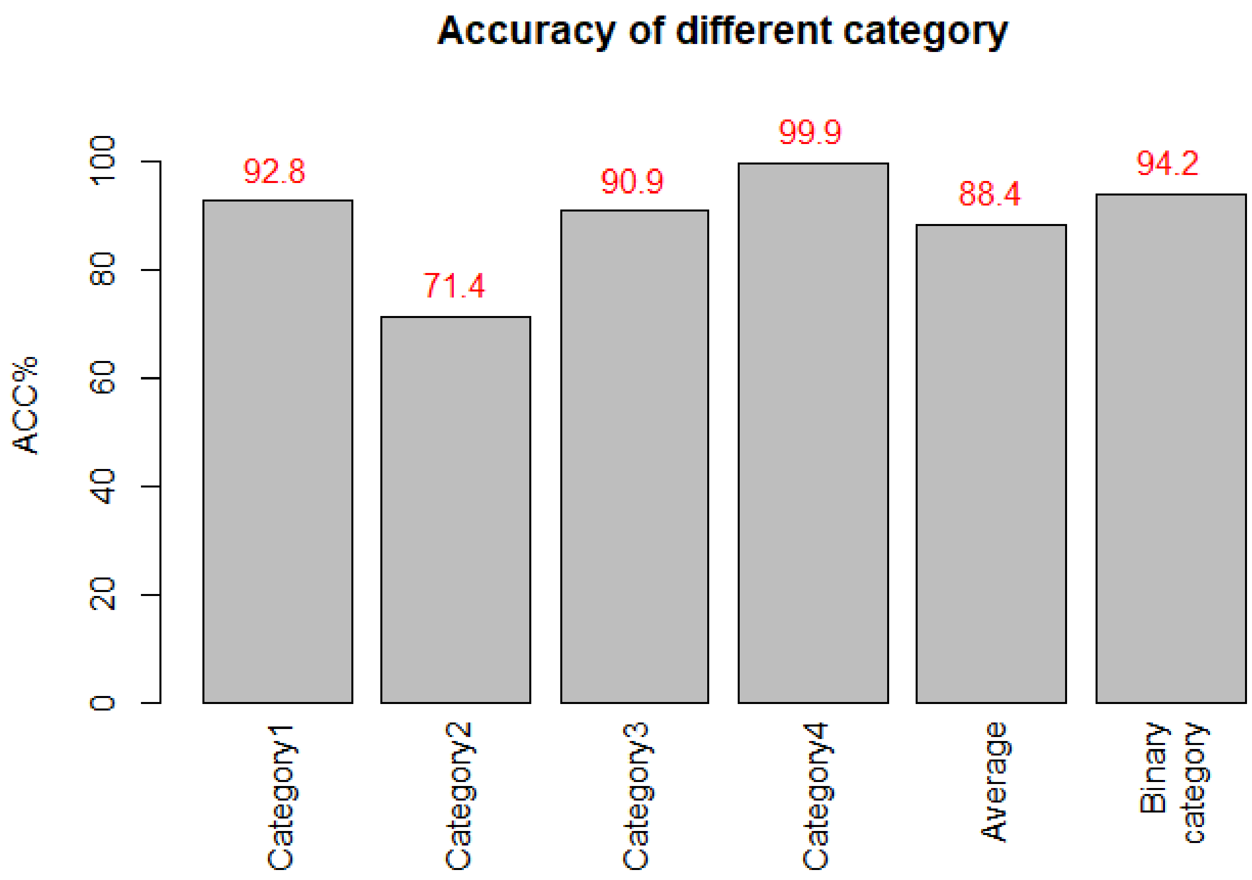
- Hold-out method: 75%-25% split
- K-fold cross validation: 5-fold cross validation

Model Validation:

- Classification metrics

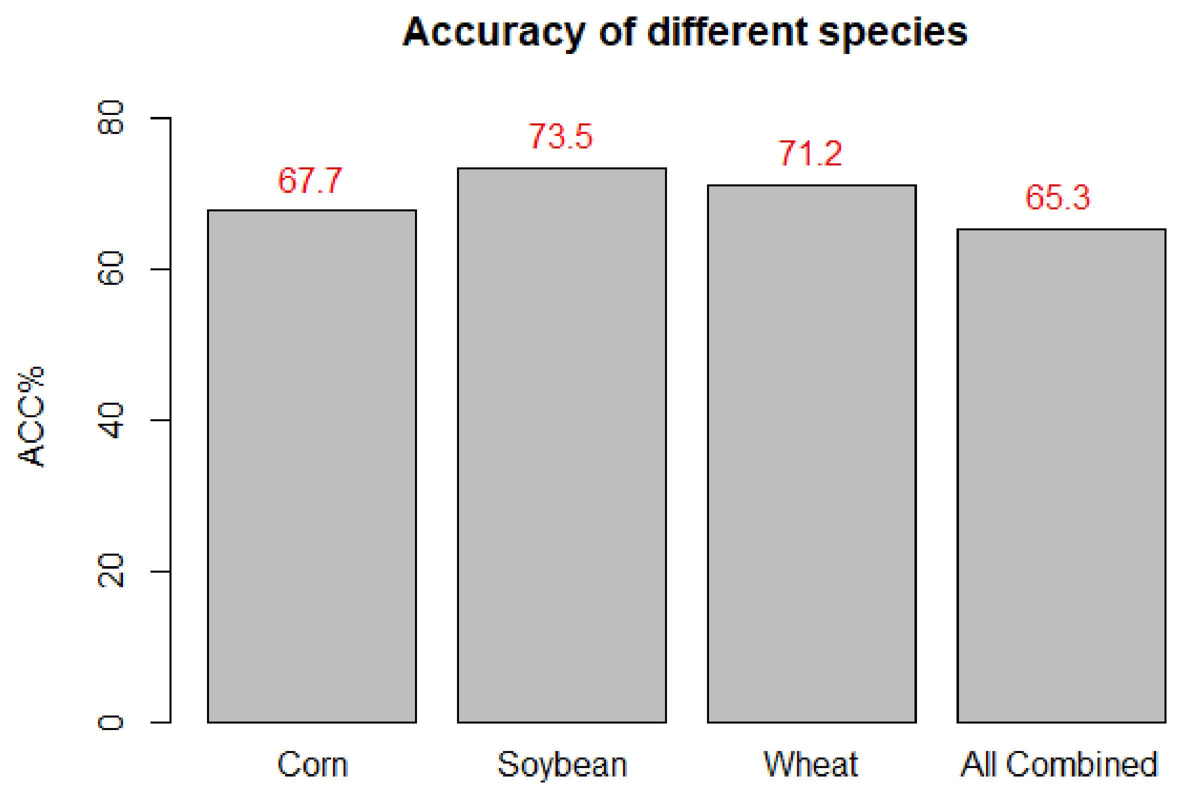
Parameter	Evaluation focus	Definition
$\mathrm{Accuracy}_\mathrm{i}$	Effectiveness of a classifier for i-th class	$\frac{\mathrm{TP_i}+\mathrm{TN_i}}{\mathrm{TP_i}+\mathrm{TN_i}+\mathrm{FN_i}+\mathrm{FP_i}}$
$\mathrm{Sensitivity}_\mathrm{i}$ $/\mathrm{Recall}_\mathrm{i}$	Effectiveness of a classifier to identify positive labels for i-th class	$\frac{\mathrm{TP_i}}{\mathrm{TP_i}+\mathrm{FN_i}}$
$\mathrm{Specificity}_\mathrm{i}$	Effectiveness of a classifier to identify negative labels for i-th class	$\frac{\mathrm{TN_i}}{\mathrm{TN_i}+\mathrm{FP_i}}$
$\mathrm{Precision}_\mathrm{i}$	Class agreement of the data labels with the positive labels for i-th class	$\frac{\mathrm{TP_i}}{\mathrm{TP_i}+\mathrm{FP_i}}$
$\mathrm{F1-Score}_\mathrm{i}$	Relations between positive labels and those given by a classifier for i-th class	$\frac{2\times\mathrm{Precision_i}\times\mathrm{Sensitivity_i}}{\mathrm{Precision_i}+\mathrm{Sensitivity_i}}$
$\mathrm{Accuracy}_\mathrm{m}$	The average per-class effectiveness of a classifier	$\frac{\sum_{\mathrm{i}=1}^{\mathrm{l}}(\frac{\mathrm{TP_i}+\mathrm{TN_i}}{\mathrm{TP_i}+\mathrm{FN_i}+\mathrm{FP_i}})}{\mathrm{l}}$
$\mathrm{Sensitivity}_\mathrm{m}$ $/\mathrm{Recall}_\mathrm{m}$	Effectiveness of a classifier to identify class labels if calculated from sums of per-category decisions	$\frac{\sum_{\mathrm{i}=1}^{\mathrm{l}}\mathrm{TP_i}}{\sum_{\mathrm{i}=1}^{\mathrm{l}}(\mathrm{TP_i}+\mathrm{FN_i})}$
$\mathrm{Specificity}_\mathrm{m}$	The average per class effectiveness of a classifier to identify negative labels	$\frac{\sum_{\mathrm{i}=1}^{\mathrm{l}}\mathrm{TN_i}}{\sum_{\mathrm{i}=1}^{\mathrm{l}}(\mathrm{TN_i}+\mathrm{FP_i})}$
$\mathrm{Precision}_\mathrm{m}$	Agreement of the data class labels with those of a classifiers if calculated from sums of per-category decisions	$\frac{\sum_{\mathrm{i}=1}^{\mathrm{l}}\mathrm{TP_i}}{\sum_{\mathrm{i}=1}^{\mathrm{l}}(\mathrm{TP_i}+\mathrm{FP_i})}$
$\mathrm{F1-Score}_\mathrm{m}$	Relations between data's positive labels and those given by a classifier based on a per-class average	$\frac{2\times\mathrm{Precision_m}\times\mathrm{Sensitivity_m}}{\mathrm{Precision_m}+\mathrm{Sensitivity_m}}$

N-Sensing



- Category 1, Category 2, Category 3, and Category 4 represent four N treatments (0, 6, 12, and 20 g/L) in the greenhouse experiment.
- <u>Binary category</u>: (Category1+Category2) & (Category3+Category4)

P-Sensing



Accuracy of different species
ACC%
0
20
40
60
80
67.7
73.5
71.2
65.3
Corn
Soybean
Wheat
All Combined