CLASSIFICATION OF THE MATURITY OF TEA LEAVES USING HYPERSPECTRAL IMAGING

Dr.K.Subramanian[1], Kangkan Talukdar[2], Mamidi H Varshith[3], Krtyush Kumar[4]

[1]Assistant Professor, Department of IT and Analytics, Xavier Institute of Management and Entrepreneurship (XIME), Bangalore, India

[2,3,4]Research Student, Department of IT and Analytics, Xavier Institute of Management and Entrepreneurship (XIME), Bangalore, India

E-mail: subramanian@xime.org, kangkan29blr@xime.org, mamidi29blr@xime.org, krtyush29blr@xime.org

Dr K. Subramanian,

ABSTRACT

Technologies for the food industry to enhance product quality through precise and efficient analysis methods have been driven by advancements in the food industry, which have in turn led to the adoption of cutting edge technologies. HSI, which was first developed for remote sensing, has also been used for food commodity classification because it is a non-invasive technology that can analyze multiple chemical properties at a single time. Black tea production, the tender leaves (bud and first two leaves) must be harvested, but since manual plucking involves mature leaves, the quality of the product is reduced. A machine learning based classification model using hyperspectral imaging to identify tender tea leaves is presented in this research with the aim of improving productivity and minimizing plucking errors.

Spectral data of tea leaves were captured using a relatively inexpensive in-house hyperspectral camera using hyperspectral imaging. About six images, each containing 5–6 leaves, were used to get spectral information and were further divided into two maturity classes: tender and mature. The data was split into training and testing sets for the study. The classification performance of six algorithms: CART, LR, LDA, KNN, NB, and SVM was evaluated through confusion matrices.

The study demonstrated that among the six machine learning models tested, KNN and SVM achieved the highest accuracies of approximately 75%. When tested with the validation dataset, the models showed better performance in identifying tender leaves compared to mature leaves, which is particularly beneficial for tea manufacturers who primarily seek tender leaves. The research also successfully proved that a portable, low-cost hyperspectral imager could be constructed and calibrated for practical use, although software limitations prevented full utilization of the custom-built device. Despite using third-party hyperspectral tea leaf data for the final analysis, the study conclusively showed that tender leaves can be distinguished using HSI and machine learning, offering a promising solution for tea gardens to automate and improve their leaf selection process.

Keywords: Hyperspectral Imaging, HIS in tealeaves

INTRODUCTION

The food industry is focused on developing high-quality products using technologies that ensure quick, accurate analysis with minimal losses. Hyperspectral imaging, originally used in remote sensing, has become a valuable tool for classifying and identifying food commodities. It offers advantages such as minimal sample preparation, non-destructive analysis, quick acquisition, and the ability to map multiple chemical compositions simultaneously.

In black tea production, tender leaves (bud and first two leaves) are crucial for quality. However, human error during plucking often leads to the inclusion of mature leaves. Hyperspectral imaging, combined with AI and machine learning, can address this by analyzing leaf color and texture to classify maturity. Immature leaves are lighter green and smoother, while mature leaves are darker and coarser. Studies, such as Wicaksono et al. (2019), demonstrate the use of color features like YCbCr for maturity classification.

This study proposes a machine learning-based classification model leveraging hyperspectral imaging to accurately identify tender tea leaves, enhancing productivity and reducing errors in tea plucking.

(I) HYPERSPECTRAL IMAGING:

A spatial map of spectral variation is produced using the HSI approach, which often produces hundreds of spectral bands. In most cases, the combination of those colours is enough for people to recognise images. Integrating spectroscopic and imaging techniques to enable direct identification of various components and spatial distribution of samples was the main driving force behind the development of an HSI system.

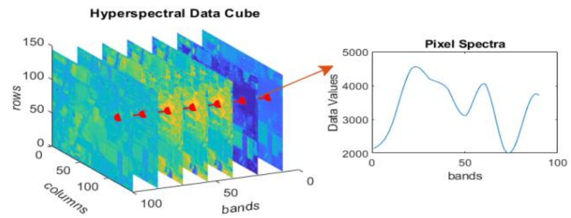

Fig. 1.1 (Data mapping for hyperspectral data)

(II) MACHINE LEARNING:

Computers can now automatically learn from historical data thanks to a quickly developing technology called machine learning. A range of algorithms are used in machine learning to build mathematical models and forecast outcomes based on knowledge or data from the past.

The creation of methods that enable a computer to learn independently from data and prior experiences is the main focus of machine learning, a branch of artificial intelligence. Without explicit programming, machine learning algorithms use training data—samples of past data—to build a mathematical model that helps with predictions or decision-making. To create predictive models, machine learning blends statistics and computer science. Algorithms used in machine learning are developed or implemented using historical data.

(III) STRUCTURE OF AN IMAGE CLASSIFICATION TASK:

1. Image Preprocessing
2. Detection of an object
3. Feature extraction and Training
4. Classification of the object

RESEARCH GAP

Although this area of study of HSI is relatively modern there has been few research analyses on hyperspectral imaging in the food related industry. The researchers found a way to classify fruit ripeness status using machine learning techniques. Najeeb and Safar (2018) used colour and texture to define fruit development. Mubin et al. (2019) used deep learning to classify oil palm trees into juvenile and mature categories with a 95.11 percent accuracy rate. through the elimination of the texture and colour components. Behera et al. (2020) classified papaya fruit maturity state using an Artificial Neural Network (ANN). Pourdarbani et al. (2020) classified Fuji apples and mangoes using K Nearest Neighbour (KNN) by removing texture and colour characteristics. The classification of mangosteen fruit maturity status using a Convolutional Neural Network (CNN) was explained by Sudana et al. (2020).

Agricultural robots powered by artificial intelligence (AI) are now used by farmers. Over the past five years, agri-informatics has made extensive use of contemporary technologies like machine learning. In a number of agri-informatics applications, including soil texture analysis, soil pH prediction, chlorophyll prediction, and plant disease detection, researchers have tried to apply machine learning methods.

Also, this project is being done with low-cost equipment, usually we see researchers using the HSI technology use conventional or lab-based HSI equipment, which is very costly and is not portable. We want to prove that expensive equipment is not required to perform these researches, we can DIY our own equipment which makes the whole project cheaper and cuts unnecessary costs.

RESEARCH METHODOLOGY

IMAGE ACQUISITION

The images for testing were provided by a nearby tea garden owner with permission, with an in-house Hyperspectral Camera.

CREATING DATASETS

Around 6 hyperspectral images of tea leaves were acquired, in which every image had around 5-6 leaves from which data can be extracted. After extracting spectral data from the leaves individually, these were split into two classes based on the leaf's maturity, i.e., tender and matured. Matured leaves were marked as 0 and tender leaves were marked as 1. Then, the whole dataset was divided into two, training set and testing/validation set. The data is stored in the form of excel sheet (.xlsx).

EXTRACTING SPECTRAL DATA

Extraction of spectral data of hyperspectral images is done on MATLAB (R2022). The Hyperspectral Image Library of the Image Processing Toolbox was made use to open the hypercube in MATLAB and extract its spectral data. Here is the code that is used to import a hypercube to MATLAB and view it.

```
>> hcube = hypercube('ENVI_cube_20211215_160400.raw')

hcube =

  hypercube with properties:

      DataCube: [1024×1280×10 uint16]
    Wavelength: [10×1 double]
      Metadata: [1×1 struct]

>> hyperspectralViewer(hcube)
```

Fig 2.7 (Code in MATLAB that opens the Hyperspectral Viewer)

The above code finally opens the following screen:

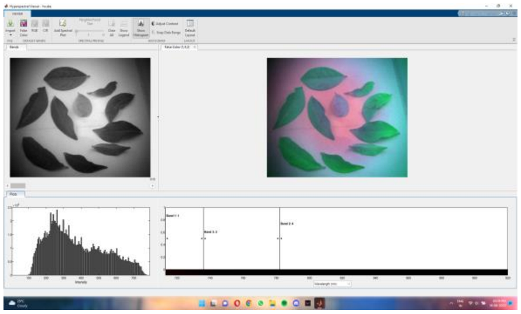

Fig 2.8 (Interface for extracting spectra data in MATLAB)

MODEL EVALUATION

We will be firstly training the model using 6 ML algorithms. The algorithms are: Classification and Regression Trees (CART), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbours (KNN), Naive Bytes (NB), and Support Vector Machine (SVM). Conceptual Descriptions of the models are mentioned in the upcoming sections.

At first, these models are used to train the model. This is done with the training dataset. Then, the trained model is validated with the same dataset to see how accurate the trained model is.

```
[ ] models =[]
    models.append(('LR', LogisticRegression()))
    models.append(('LDA', LinearDiscriminantAnalysis()))
    models.append(('KNN', KNeighborsClassifier()))
    models.append(('CART', DecisionTreeClassifier()))
    models.append(('NB', GaussianNB()))
    models.append(('SVM', SVC()))

[ ] results =[]
    names = []
    scoring = 'accuracy'
    for name, model in models:
        kfold= KFold(n_splits=2, random_state=None)
        cv_results= cross_val_score(model, X, Y, cv=kfold, scoring=scoring)
        results.append(cv_results)
        names.append(name)
        msg= "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
        print(msg)
```

Fig 2.9: (Code to train the model with each algorithm.)

The most accurate algorithm(s) is/are then used to test the respective model with the testing/validation dataset. In the end, we then represent the accuracy of the tested model using confusion matrices.

PERFORMANCE EVALUATION USING A CONFUSION MATRIX

A table that is commonly used to explain how well a classification model (also known as a "classifier") performs on a set of test data with known true values is called a confusion matrix. It is represented by a table with four distinct combinations of actual and anticipated classes. The four different combinations are:

- True Positive: The positive class is correctly predicted by the model.
- True Negative: The negative class is correctly predicted by the model.
- False Positive: The model predicts the positive class incorrectly.
- False Negative: The model makes an incorrect prediction about the negative class.

RESEARCH FINDINGS

SPECTRAL DATA

Here is the spectral data for each leaf from the hyperspectral images. The spectral data is actually the reflectance at a ROI in an image for each band of the whole hypercube. MATLAB only allowed taking 10x10 pixels of selection, and the spectral data shown below is an average of these points. Since the data is split into two parts, one for training and the other is for testing, we will first be showing the training data, followed by the testing data.

Tender leaves:

Bands	leaf 1	leaf 2	leaf 3	leaf 4	leaf 5	leaf 6	leaf 7	leaf 8	leaf 9	leaf 10	leaf 11	leaf 12
1	290	294	206	218	187	204	333	270	228	241	360	575
2	372	350	337	321	266	269	389	328	344	291	348	470
3	277	255	303	253	223	245	313	284	300	238	255	304
4	173	147	196	149	142	149	179	161	176	138	140	177
5	94	76	118	88	88	88	96	84	104	82	95	115
6	65	61	74	55	54	61	75	58	76	59	77	90
7	47	44	44	42	44	39	56	43	48	44	54	74
8	41	46	46	35	32	35	35	31	32	35	54	61
9	30	33	34	36	26	28	39	25	36	24	39	74
10	37	36	30	37	27	34	39	28	40	30	52	78

Table: Spectral data of all the tender leaves (train set)

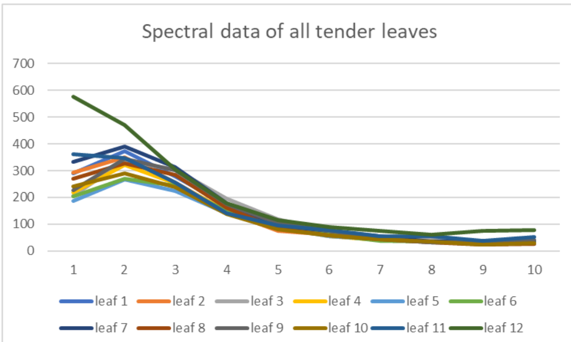

Fig 3.2 Line graph of all the spectral data of all tender leaves (train set)

Matured leaves:

Bands	leaf 1	leaf 2	leaf 3	leaf 4	leaf 5	leaf 6	leaf 7	leaf 8	leaf 9	leaf 10	leaf 11	leaf 12
1	287	564	292	696	211	532	606	449	323	236	588	573
2	389	409	347	416	297	505	471	433	333	330	456	482
3	312	277	256	268	258	342	275	279	248	252	284	290
4	175	165	149	148	158	199	168	147	131	169	164	181
5	100	104	85	104	100	127	106	87	90	95	112	110
6	72	82	68	96	66	86	87	63	64	57	92	98
7	53	66	48	92	48	77	80	57	52	50	80	80
8	42	69	34	72	37	66	61	40	47	37	68	70
9	30	55	25	84	31	61	65	47	39	35	84	62
10	38	64	36	92	37	79	73	42	51	37	72	70

Table: Spectral data of all matured leaves (train set)

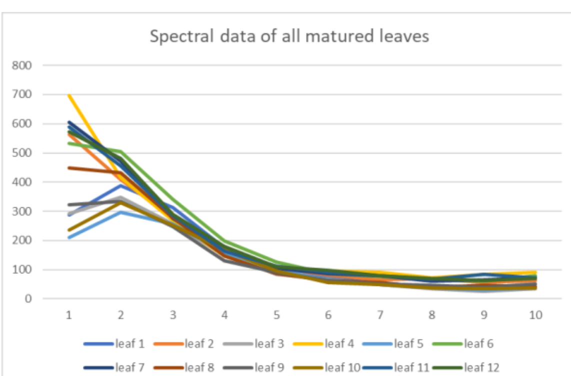

Fig 3.4 Line graph of all the spectral data of all matured leaves (train set)

The following are the spectral data for each leaf (containing both tender and matured leaves) for the testing/validation set.

Tender leaves:

Bands	leaf 1	leaf 2	leaf 3	leaf 4	leaf 5	leaf 6	leaf 7	leaf 8	leaf 9	leaf 10
1	328	280	211	219	224	227	214	299	200	246
2	388	322	328	311	344	253	226	275	259	272
3	328	264	316	241	292	167	159	157	195	163
4	168	151	192	146	196	99	84	86	109	86
5	92	91	106	84	100	57	57	57	69	56
6	68	63	78	50	72	44	39	43	38	39
7	52	40	41	40	56	27	30	39	34	33
8	36	34	36	37	40	31	27	36	22	26
9	32	31	33	25	28	31	27	36	22	26
10	40	31	40	29	36	34	32	35	33	30

Table: Spectral data of all the tender leaves (Test set)

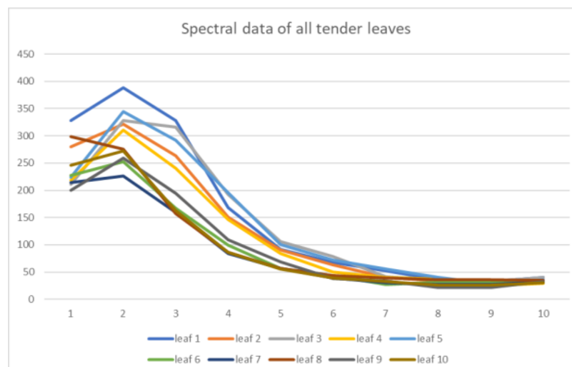

Fig 3.7 Line graph of all the spectral data of all tender leaves (test set)

Matured leaves

Bands	leaf 1	leaf 2	leaf 3	leaf 4	leaf 5	leaf 6	leaf 7	leaf 8	leaf 9	leaf 10
1	220	232	224	173	176	254	240	263	223	248
2	268	240	248	252	243	311	312	319	327	340
3	208	180	156	190	222	248	256	246	272	280
4	126	100	88	107	142	111	136	129	137	145
5	63	56	60	72	89	76	80	86	74	81
6	40	36	40	50	58	87	76	80	77	84
7	39	28	28	33	39	39	32	43	43	49
8	32	32	20	28	38	33	32	34	38	38
9	28	24	28	27	29	20	24	40	30	36
10	34	24	32	28	39	28	24	45	38	39

Table: Spectral data of all the matured leaves (Test set)

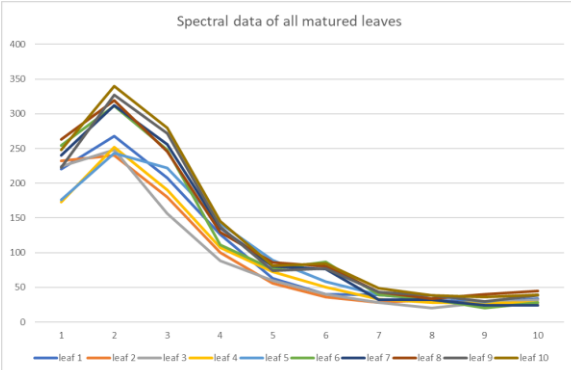

Fig 3.9 Line graph of spectral data of all matured leaves (test set)

TRAINING ACCURACY

As mentioned in the methodology, we trained 6 different models of different algorithms with the training set. Here are the accuracies for each model:

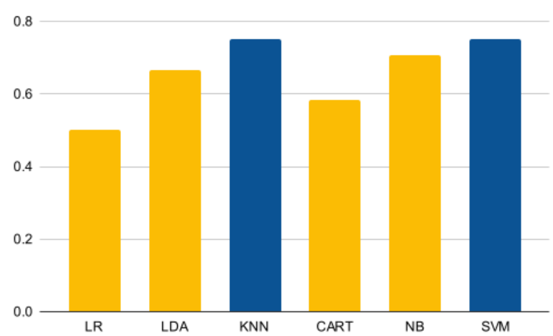

Fig 3.11 Accuracies of all the models in a bar graph

From the figure, we can see that KNN and SVM have the highest accuracies of them all (around 75%). Therefore, we will be testing the model with our testing/validation dataset with these two algorithms

TESTING THE MODELS

Finally, we tested the most accurate models found earlier using the testing dataset. The most accurate models came out to be KNN and SVM, and therefore we used these models to test. We shall discuss how each of the models performed in the next section.

CONFUSION MATRICES

To know how each of the models performed, we plot confusion matrices for each model. Here are confusion matrices for each model that was tested.

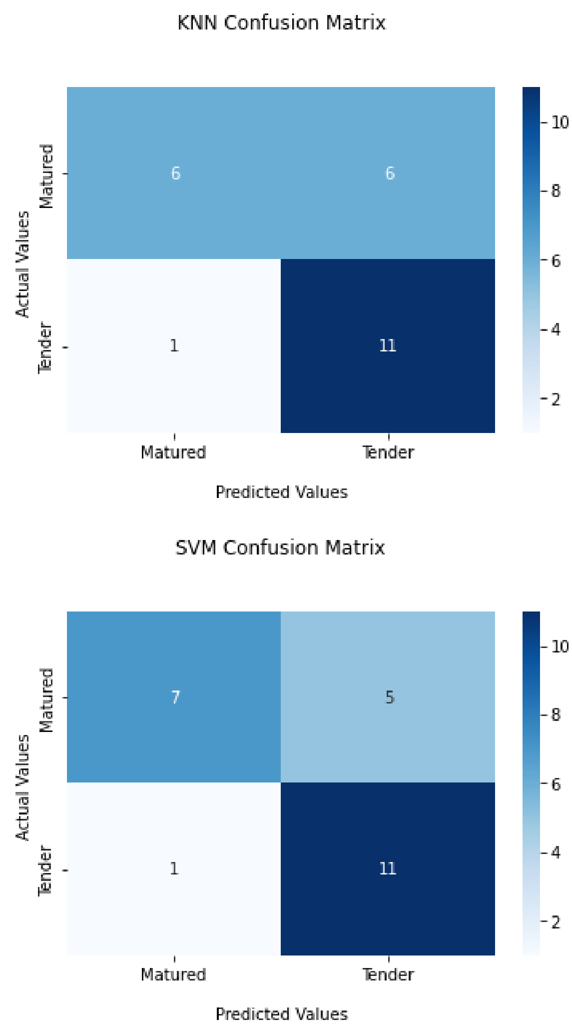

Fig 3.13 (Confusion matrix of K-Nearest Neighbour & Support Vector Machine)

CONCLUSION

We were able to construct a portable and low-cost Hyperspectral Imager, take images and even calibrate it for our use. However its software limitations couldn’t allow us to utilise it further as it couldn’t export the required hypercubes. It is a small software issue, and if given enough time, it could be rectified and therefore, it is possible to deploy a low cost and portable Hyperspectral Imager and use it instead of commercial and heavy counterparts for small-scale uses.

Due to this problem, we used some 3rd party hyperspectral tea leaves for our use. We then extracted its spectral data using a software named MATLAB. This information from the tea leaves is necessary so that we are able to analyse and classify them.

Out of the 6 ML models that were used, 2 of them, namely KNN and SVM proved to be the most accurate and therefore these were used for the testing phase of the ML modelling.

While testing the model, it was found that the model was able to differentiate tender leaves better than matured leaves from the confusion matrix. This is not bad, as for tea manufacturers, tender leaves are needed rather than matured leaves.

Upon executing this experiment, we came to know that tender leaves can be distinguished using HSI and Machine Learning. This can prove to be very helpful for tea gardens to identify tender leaves in their garden using a portable HSI camera and then proceed accordingly.

REFERENCES

“Benelli, A., Cevoli, C., & Fabbri, A. (2020). In-field hyperspectral imaging: An overview on the ground-based applications in agriculture. *Journal of Agricultural Engineering, 51*(3), 129-139.”

“ElMasry, G., & Sun, D. W. (2010). Principles of hyperspectral imaging technology. In *Hyperspectral imaging for food quality analysis and control* (pp. 3-43). Academic Press.”

“Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology, 18*(12), 590-598.”

“Jia, Y., Huang, W., Guo, L., Wu, J., & Fan, S. (2018). Apple surface pesticide residue detection method based on hyperspectral imaging. In *International Conference on Intelligent Science and Big Data Engineering* (pp. 27-63). Springer, Cham.”

“Jian, W., Xianyin, Z., & ShiPing, D. (2010). Identification and grading of tea using computer vision. *Applied Engineering in Agriculture, 26*(4), 639-645.”

“Li, X., Wei, Y., Xu, J., Feng, X., Wu, F., Zhou, R., & Jin, J. (2017). Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. In *Hyperspectral imaging in agriculture, food and environment* (pp. 27-63).”

“Salazar-Vazquez, J., & Mendez-Vazquez, A. (2020). A plug-and-play hyperspectral imaging sensor using low-cost equipment. *HardwareX, 7*, e00087.”