Machine Learning of Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 20
Applications of Classification and Clustering Methods in Agriculture (contd.)

Welcome friends to this twentieth lecture of NPTEL online certification course of Machine Learning for Soil and Crop Management. And this is the last lecture of week 4. And in this week, we are discussing the Application of Different Classification and Clustering Methods in Agriculture specifically focusing on soil and crop.

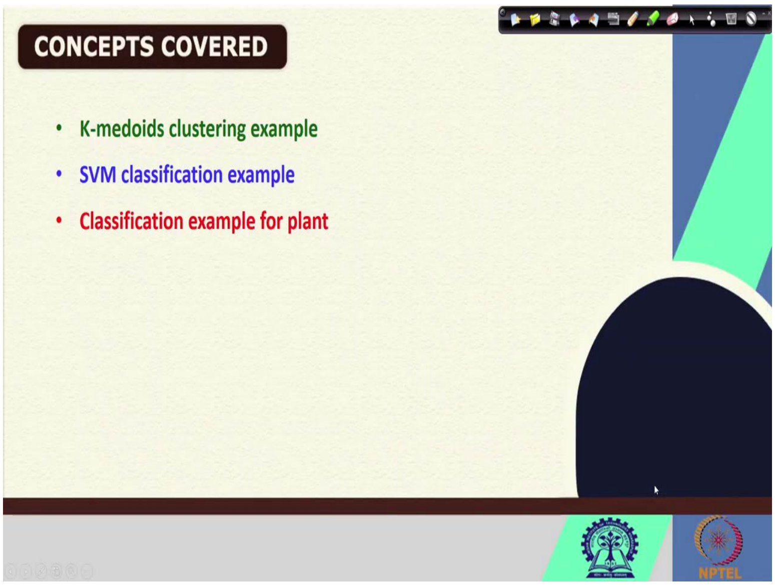
So, in our first 4 lectures of this week, we have already discussed multiple classification and clustering methods, we have discussed in detail the linear classification methods, kernel based classification methods we have discussed. So, we have discussed linear discriminant analysis, then we have also discussed the logistic regression, then k nearest neighbor based classification we have discussed, we have also discussed the classification tree based clustering and classification to this classification actually.

And also we have discussed support vector machine, artificial neural network, random forests and their application for classification we have seen for both soil and crop, we have also discussed the clustering, different dissimilarity measures, how to calculate different dissimilarity measures in case of clustering, what are the importance of selection of proper dissimilarity measures in clustering we have also discussed, also we have seen the performance metrics which you can calculate from a confusion metrix of any classification problem.

And we have also discussed in details K-means clustering, K-medoids clustering and also, we have discuss the agglomerative hierarchical clustering, we have seen that Dendrogram, we have discussed the Dendrogram in details. So, in this lecture instead, so, we have covered most of the theories, as far as the classification and clustering methods are concerned there are also other classification and clustering methods.

Of course, it is not possible to cover all of them in a single, week, but we will try to discuss them in our upcoming weeks. I have decided to discuss some of the application instead of discussing the theory in this twentieth lecture. So, in this lecture, we are going to discuss or see, some of the examples and we will explain the different trends, classification trends and clustering trends in details using some of the published examples.

(Refer Slide Time: 3:39)

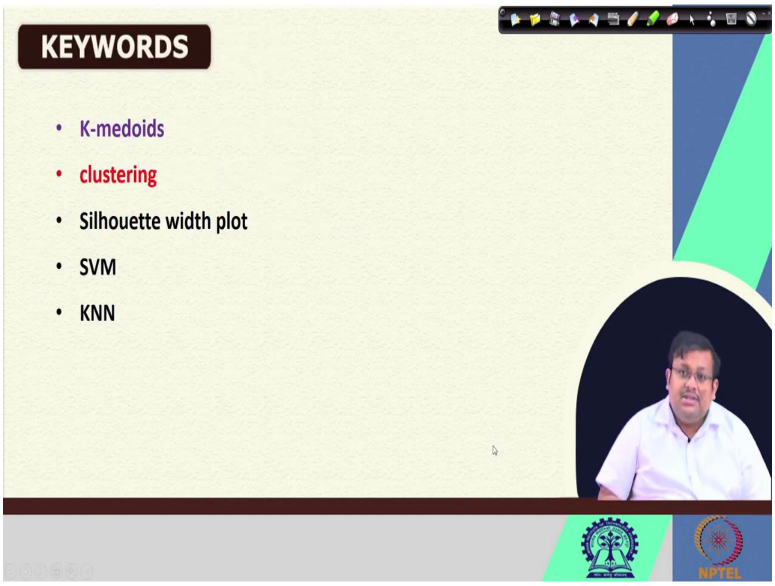


CONCEPTS COVERED

- K-medoids clustering example
- SVM classification example
- Classification example for plant

So, these are the concepts which we are going to cover, we are going to start with an example of K-medoids clustering and then we are going to discuss the SVM classification and finally, we are going to discuss one example of classification for plant.

(Refer Slide Time: 4:08)

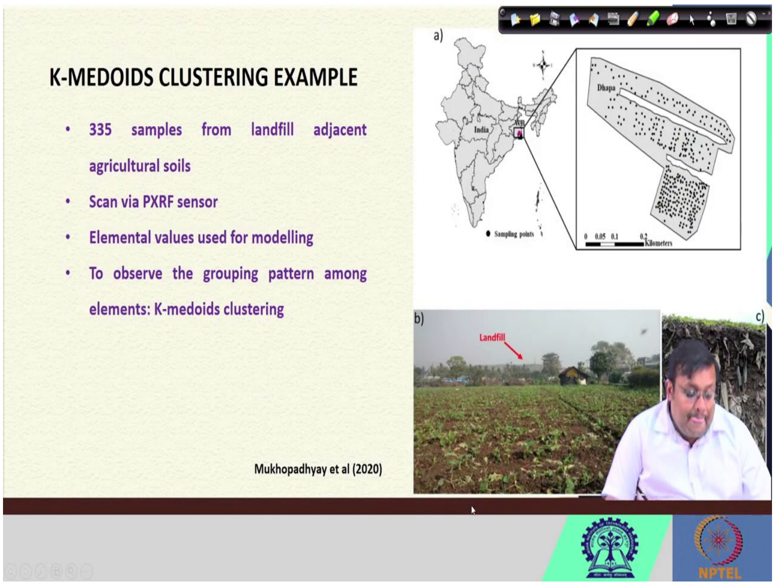


KEYWORDS

- K-medoids
- clustering
- Silhouette width plot
- SVM
- KNN

So, these are the key words which we are going to discuss K-medoids, then clustering, then silhouette width plot, SVM, kNN. So, these are some key words for this lecture.

(Refer Slide Time: 4:26)



K-MEDOIDS CLUSTERING EXAMPLE

- 335 samples from landfill adjacent agricultural soils
- Scan via PXRF sensor
- Elemental values used for modelling
- To observe the grouping pattern among elements: K-medoids clustering

Mukhopadhyay et al (2020)

So, let us see some one example of K-medoids clustering. So, in case of K-medoids clustering these paper which was published in 2020 by Mukhopadhyay et al gives a very good application of K-medoids clustering method. Now, the basic principle of K-medoids clustering in case of K-medoids clustering, the observations are assigned to each of these K clusters and then they are used to calculate the distance from one of the sample from the cluster.

So, that is called K-medoids clustering and this algorithm iterates until the lowest within class variability is reached. So, this is called the K-medoids clustering which is somewhat different than K-means clustering because K-means clustering, the distance is calculated from the center of the class which is calculated by taking a mean and of course, the Euclidean distance is a square in case of K-means, we use the squared Euclidean distance.

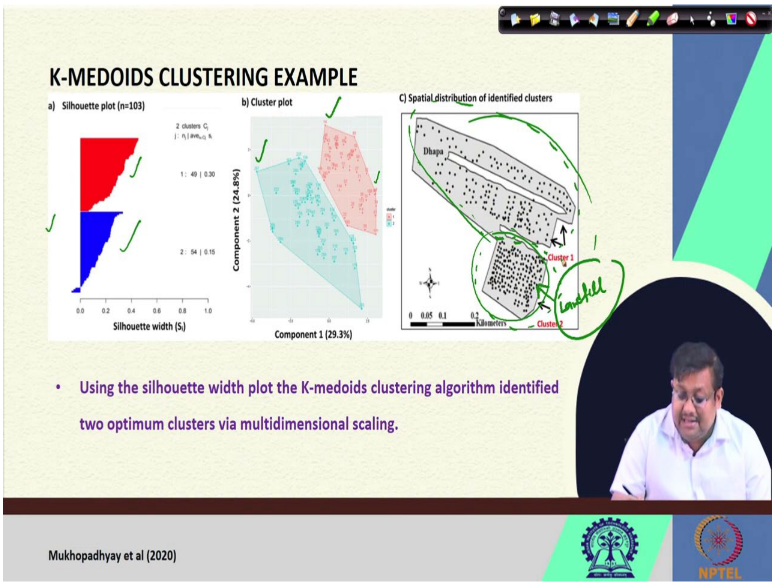
So, K-means clustering is less robust to outliers, whereas, the K-medoids clustering is more robust towards the outlier. So, in this example, they have calculate they have collected around 335 samples from a landfill adjacent agricultural soil and this landfill is situated near Kolkata city of India and there is a big landfill area called Dhapa.

And surrounding that landfill area, there is a big agricultural area where farmers used to go different types of vegetables. Now, these vegetables they are hyper accumulator of the heavy metals and being in close proximity to this landfill, there is a high chance of getting heavy metal pollution in this area. So, the idea of the research was to use some approximate soil sensor to get the special variability map of heavy metals for this area, and also to explore the heavy metal contamination of this area for future operations.

So, these are the, this is the area and these are the samples, soil samples 335 soil samples collected. After collecting the samples, they were scanned using the portable XRF, we will discuss the portable XRF in details in our upcoming weeks. So, remember that this is a handheld equipment, which used to measure or which used to calculate the elemental content total elemental content of any powdered material.

So, we scan the soil samples using the portable XRF to get their elemental content and those elemental values were used for subsequent modeling. So, to observe, so, in the subsequent machine learning applications, we try to observe the grouping pattern among the elements using the K-medoids clustering.

(Refer Slide Time: 8:41)



K-MEDOIDS CLUSTERING EXAMPLE

- Using the silhouette width plot the K-medoids clustering algorithm identified two optimum clusters via multidimensional scaling.

Mukhopadhyay et al (2020)

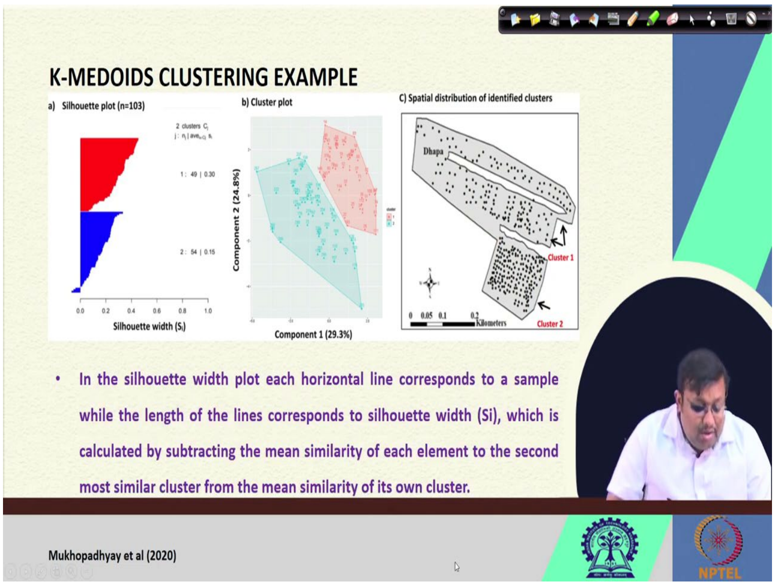
So, K-medoids clustering this is the K-medoids clusters and this is called a silhouette width plot. So, using the silhouette width plot, the K-medoids clustering algorithm identified 2 optimum cluster you can see here 2 optimal clusters via multi-dimensional scaling. And surprisingly, these 2 clusters belong to these 2 different regions.

And it has been found that these cluster these 2 clusters, cluster 1 and cluster 2, so this red one is cluster 1, this is assigned to this upper patch of this area, whereas, this cluster 2 is assigned to this lower patch of this area. What is the reason? Because it has been found that the original landfill is situated in the close proximity from this lower patch.

So, of course, we can expect higher amount of contamination in this lower patch as compared to this upper patch which is farther away from this landfill site active landfill dump site. So, the elemental content as well as our statistical measurement of clustering was able to clearly identify the special variation of the heavy metal contamination using these K-medoids clustering.

Now, this K-medoids clustering is based on the silhouette width plot which are going to discuss. So, using the silhouette width plot we have calculated 2 clusters these 2 clusters shows can be can be also seen visually in the special there is can be represented by the special variation which you see in these upper patch and lower patch.

(Refer Slide Time: 11:04)



K-MEDOIDS CLUSTERING EXAMPLE

- In the silhouette width plot each horizontal line corresponds to a sample while the length of the lines corresponds to silhouette width (Si), which is calculated by subtracting the mean similarity of each element to the second most similar cluster from the mean similarity of its own cluster.

Mukhopadhyay et al (2020)

So, in this silhouette plot, the silhouette width plot each horizontal line so, these are basically each horizontal line. So, each horizontal line correspond to a sample while the length of the line correspond to silhouette width. So, this length of the line is indicative of the silhouette width which is calculated by subtracting the mean similarity of each element to the second most similar cluster from the mean similarity of in own cluster.

So, again please try to understand these height of this individual sample or silhouette width is calculated by subtracting the mean similarity of each element to the second most similar cluster from the mean simulator of its own cluster. So, of course, the more close a sample to its corresponding cluster the silhouette width will be high. So, based on these they are being sorted, we can see 2 clear separation, one is red cluster another is blue cluster.

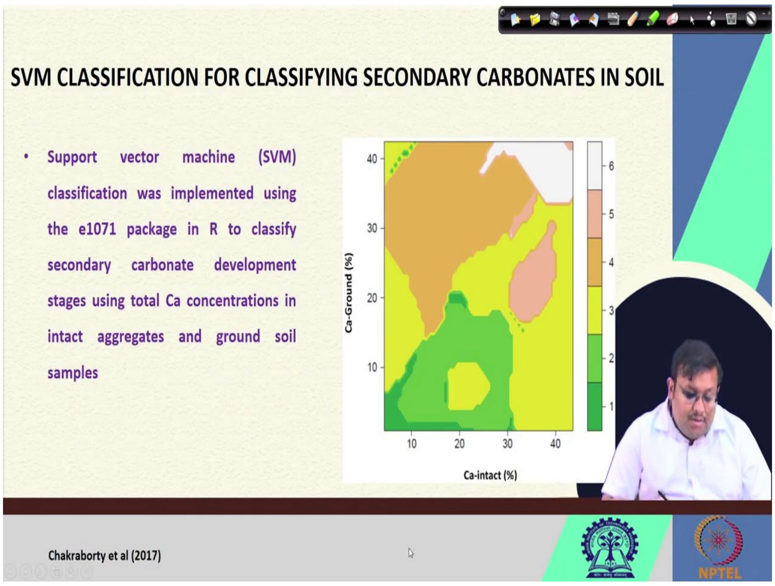
Now, generally these Si lies between minus 1 to plus 1. So, we can see either negative values to positive values. So, while clustering, the sample indicated a large, so, when there is when a sample is well clustered this like here, we can see there is these Si value should be closer to plus 1 and a value of 0 around 0 will indicate a sample placed between 2 different clusters.

So, they are not clearly classified they are somewhat intermediate between 2 clusters. However, if the Si value is negative, then it signifies that it is placed in a wrong cluster. So, you can see most of the samples in each of these 2 clusters are correctly classified only a very minute number of samples in the second cluster which might be wrongly classified. So, overall we can see that the K-medoid clustering are more or less highly accurate.

So, interestingly the sample separated in the 2 clusters also exhibited special trips just like we have seen in case of special trips we have seen perhaps due to special difference in heavy metal concentration in the study of course, because of the close proximity from the active dumping side, those areas which are close will have higher concentration and those area which are farther away will have lower concentration.

So, we can see the difference clearly their spatial variation and our statistical measure of clustering was also able to identify the spatial variation.

(Refer Slide Time: 14:13)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

- Support vector machine (SVM) classification was implemented using the e1071 package in R to classify secondary carbonate development stages using total Ca concentrations in intact aggregates and ground soil samples

Chakraborty et al (2017)

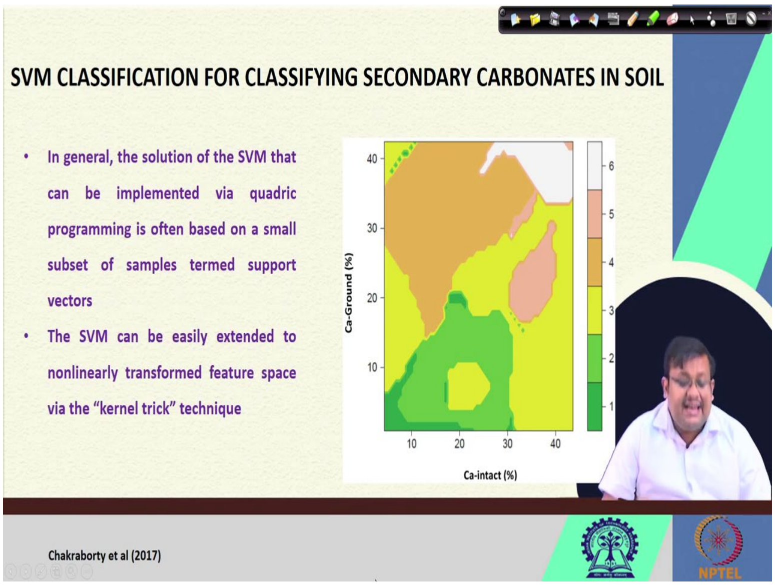
So, that goes one application of K-medoids clustering for soil. Let us see another example, where support vector machine classification was applied for classifying secondary carbonates in soil. Now, secondary carbonates are very important as far as the soil petrology is concerned we need to identify which the secondary carbonates in the field. So, they experience petrology, so, soil scientists can visually identify the soil and then they identify what is the stage of development of secondary carbonate.

So, stage of the development of the secondary carbonate is so, far being operated by the experienced metallurgist. So, back in 2017, our group thought that let us start let us apply this portable XRF for you portable excited based we just let us scan the soil samples we using the portable XRF and used those values of elements to classify different stages of the secondary carbonates development.

So, here you can see the support vector machine classification was implemented using these e1071 package in R to classify secondary carbonate development stage using total calcium concentration in intact aggregates and ground soil samples. So, we have the intact calcium in the intact samples, we calculate the calcium and also we ground the soil samples and we calculate the calcium using the portable XRF.

And using these 2 features, we try to classify the 6 stages of a secondary carbonates secondary 6 stages secondary carbonates, stage 1, 2, 3, 4, 5, 6 using the support vector machine.

(Refer Slide Time: 16:43)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

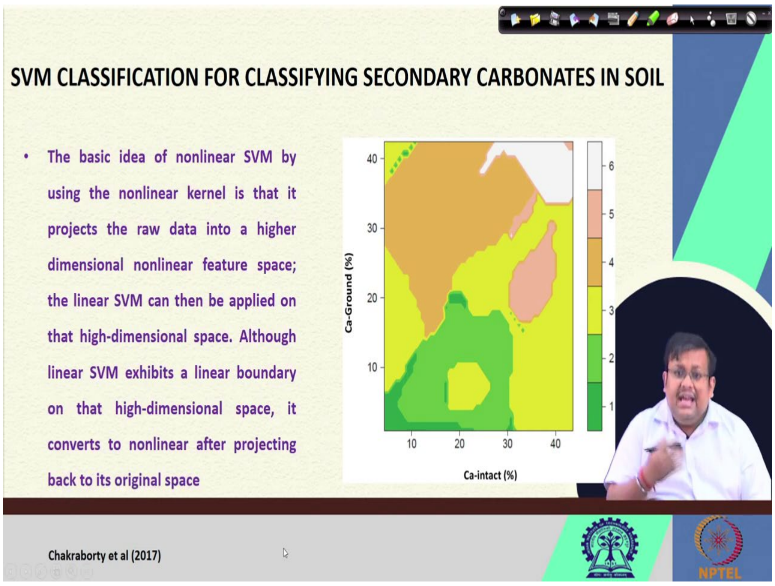
- In general, the solution of the SVM that can be implemented via quadric programming is often based on a small subset of samples termed support vectors
- The SVM can be easily extended to nonlinearly transformed feature space via the "kernel trick" technique

Chakraborty et al (2017)

So, in general the solution of these SVM that can be implemented via quadratic program is often based on a small subset of sample termed as support vectors we know that for the, what are the support vectors what are the, these are the closest point to the margin of the hyperplane in any support vector machine.

Now, the support vector machine can be easily extended to non-linearity, non-linearly transformed feature space via the kernel trick technique, we know what is kernel. Now, basically the kernel trick technique is to re-project the data from nonlinear space to linear space and then fixing the linear hyperplanes.

(Refer Slide Time: 17:26)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

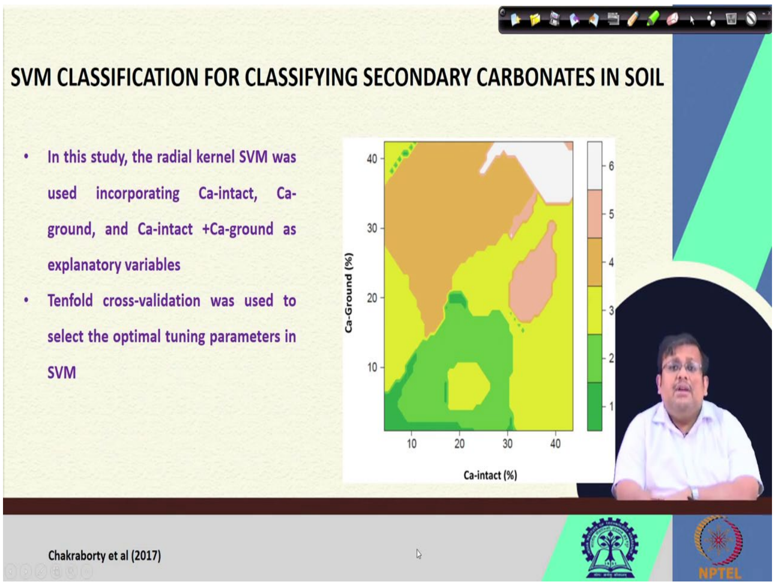
- The basic idea of nonlinear SVM by using the nonlinear kernel is that it projects the raw data into a higher dimensional nonlinear feature space; the linear SVM can then be applied on that high-dimensional space. Although linear SVM exhibits a linear boundary on that high-dimensional space, it converts to nonlinear after projecting back to its original space

Chakraborty et al (2017)

Now, the idea again, the idea of these nonlinear is SVM was, to project these raw data into higher dimensional nonlinear feature space. So, these linear SVM can be, so, the linear SVM can then be applied to these high multi-dimensional space although linear SVM exhibits a linear boundary on that high dimensional space. It converts to nonlinear after projecting back to its original space.

So, in the kernel trick what happens? High dimensional data projects to the from the nonlinear feature space to linear feature space then we fit the support vector linear boundary in each of these high dimensional space and then after we project back to its original space from this linear to nonlinear original feature space.

(Refer Slide Time: 18:35)



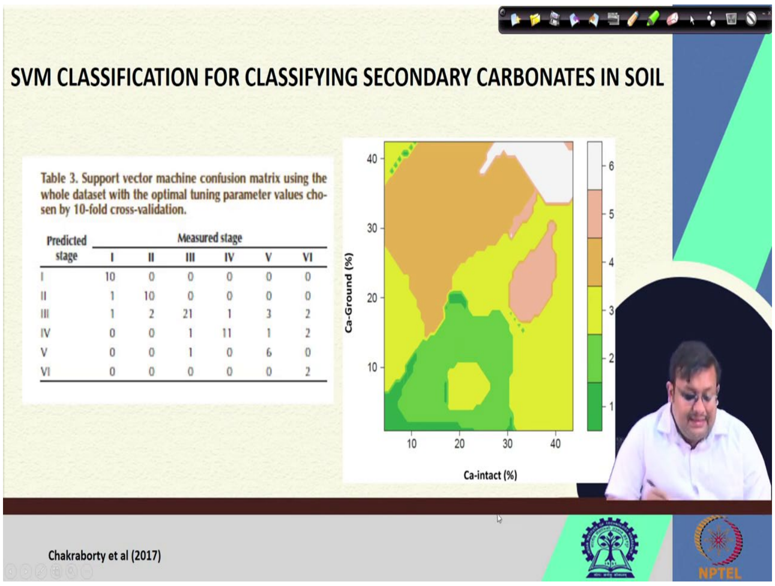
SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

- In this study, the radial kernel SVM was used incorporating Ca-intact, Ca-ground, and Ca-intact +Ca-ground as explanatory variables
- Tenfold cross-validation was used to select the optimal tuning parameters in SVM

Chakraborty et al (2017)

So, in this study generally it was radial kernel SVM was used incorporating the calcium content in the intact sample, calcium content in the ground samples and both combining calcium intact plus calcium ground samples as explanatory variables. So, we also did that tenfold cross validation to select the optimal tuning parameters of SVM.

(Refer Slide Time: 19:06)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

Table 3. Support vector machine confusion matrix using the whole dataset with the optimal tuning parameter values chosen by 10-fold cross-validation.

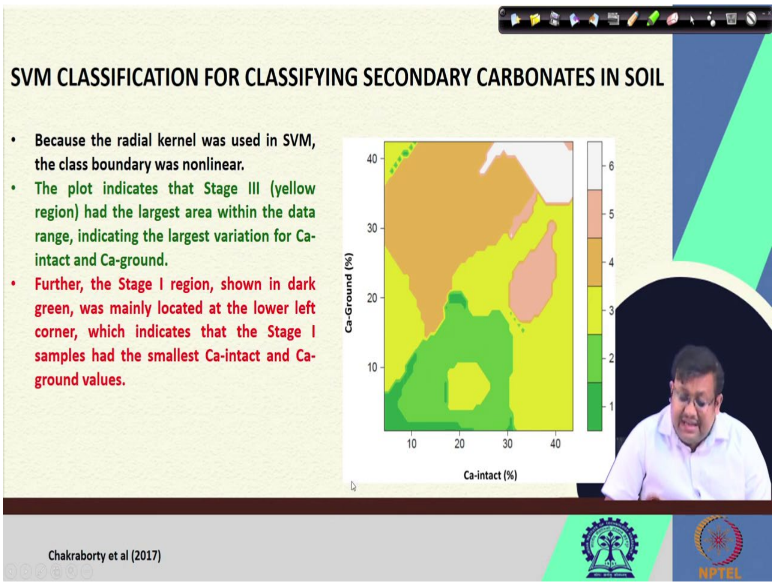
Predicted stage	Measured stage					
	I	II	III	IV	V	VI
I	10	0	0	0	0	0
II	1	10	0	0	0	0
III	1	2	21	1	3	2
IV	0	0	1	11	1	2
V	0	0	1	0	6	0
VI	0	0	0	0	0	2

Chakraborty et al (2017)

So, we can see this is the support vector machine confusion metrix using the whole subset with the optimal tuning parameter values choosen by tenfold cross validation we can see this this is the confusion matrix and by seeing the confusion matrix it is quite clear that the misclassified samples are quite less.

So, from there we can see that SVM classification was quite accurate in our case for classifying the stages or developmental stages of secondary carbonates in the soil. So, here you can see these are the SVM boundaries, 2d boundaries so, we are also going to discuss this thing.

(Refer Slide Time: 19:59)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

- Because the radial kernel was used in SVM, the class boundary was nonlinear.
- The plot indicates that Stage III (yellow region) had the largest area within the data range, indicating the largest variation for Ca-intact and Ca-ground.
- Further, the Stage I region, shown in dark green, was mainly located at the lower left corner, which indicates that the Stage I samples had the smallest Ca-intact and Ca-ground values.

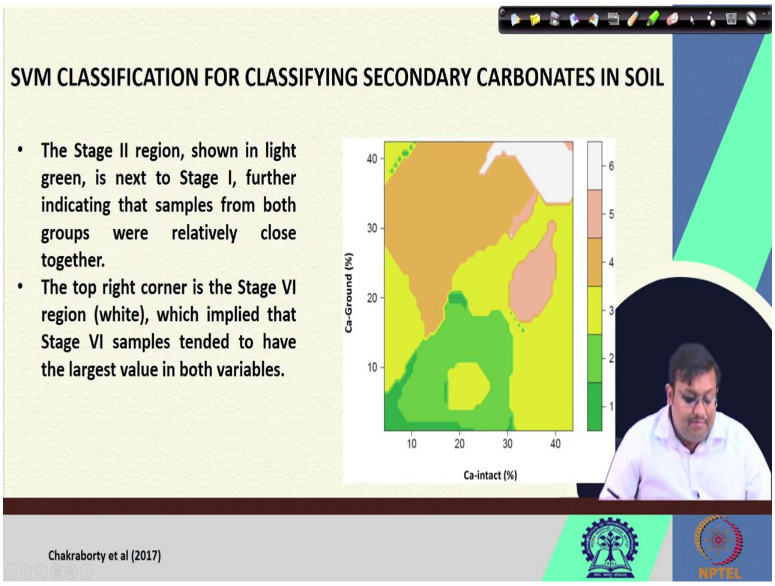
Chakraborty et al (2017)

But in a nutshell, SVM classification was highly accurate. Now, let us see what is the meaning of this is SVM classification plot? So, we can see that the radial kernel was used in SVM the class boundary was nonlinear, so, we cannot see these bounded linear class boundaries. So, these are not linear class boundaries these are nonlinear class boundaries.

So, this plot indicates that the stage 3 which is denoted by these yellow color is had the largest area within the data range indicating the largest variation for calcium in the intact samples and calcium in the ground samples okay. So, we use the calcium from the intact samples and calcium values in the ground samples.

So, we saw the largest variation for this stage 3. Further in stage 1 region which is denoted by these by this deep green color. So, the stage one is insured in dark green color was mainly located at the lower left corner as you can see in the lower left corner and had the smallest calcium intact and color ground values as of course, this is having the smallest calcium intact and calcium ground values. So, this is one another interpretation from this plot.

(Refer Slide Time: 21:26)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

- The Stage II region, shown in light green, is next to Stage I, further indicating that samples from both groups were relatively close together.
- The top right corner is the Stage VI region (white), which implied that Stage VI samples tended to have the largest value in both variables.

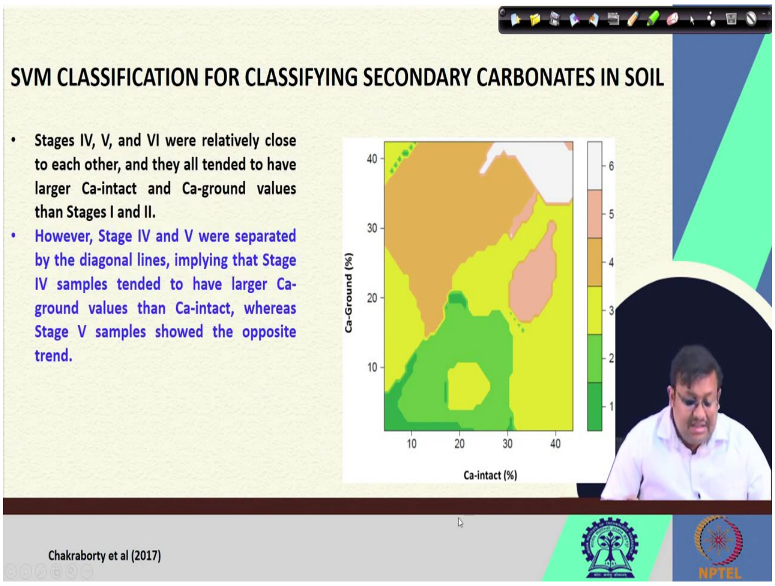
Chakraborty et al (2017)

Also, we can see that the stage 2 region shown in the light green, this is the light green stage 2 region is next to stage 1 further indicating that samples from both the groups are relatively close together. So, we can see they are almost in close proximity.

So, we can assume that the samples from this to close these groups are very much close to each other, then the top right corner is the stage 6, which is denoted by these white patch, which implied that the state 6 samples tended to have the largest values of both the variables calcium intact and calcium ground. So, we can see the largest values. So, that is why they are present this class is present at this top right corner.

So, this is another interpretation and also we can see the narrow nature of the diagonal stage 5. So, this narrow nature of this diagonal stage 5 region can be attributed to the 2 samples from stage 5 which are close to the diagonal area. So, it must be these diagonal, these narrow diagonal region for class 5 represents 2 samples from test stage 5, which are close to this diagonal area.

(Refer Slide Time: 22:57)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

- Stages IV, V, and VI were relatively close to each other, and they all tended to have larger Ca-intact and Ca-ground values than Stages I and II.
- However, Stage IV and V were separated by the diagonal lines, implying that Stage IV samples tended to have larger Ca-ground values than Ca-intact, whereas Stage V samples showed the opposite trend.

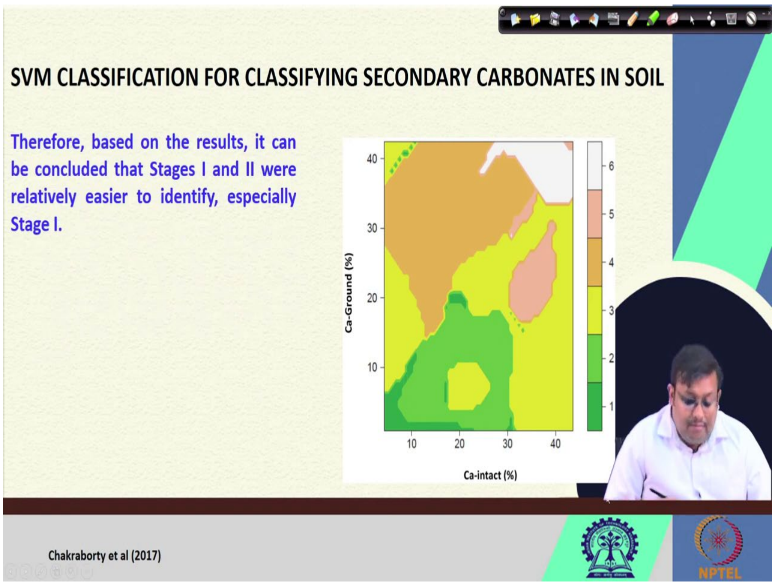
Chakraborty et al (2017)

And stage 4, 5 and 6, so, here these 4, so, this stage 4, stage 5 and stage 6 were relatively close to each other we can see here. They are relatively close to each other, they are all tended to have large calcium intact and calcium ground values. Then stage 1, 2, we know that we can this is quite evident from this graph.

And however, stage 4 and 5 were separated by the diagonal lines implying that stage 4 samples tended to have larger calcium ground values then calcium intact. So, of course, they are having the larger calcium ground values then calcium intact whereas stage 5 samples showed the opposite trend. So, they have higher calcium intact values then calcium ground values. So, they although they are close together they are showing some opposite trends.

So, from the, so, one thing is clear that by seeing this type of SVM classification plot, we can interpret a lot about the data, we can see a lot of relationship among these data by seeing these boundaries of and then that these classes in this SVM classification plot.

(Refer Slide Time: 24:33)



SVM CLASSIFICATION FOR CLASSIFYING SECONDARY CARBONATES IN SOIL

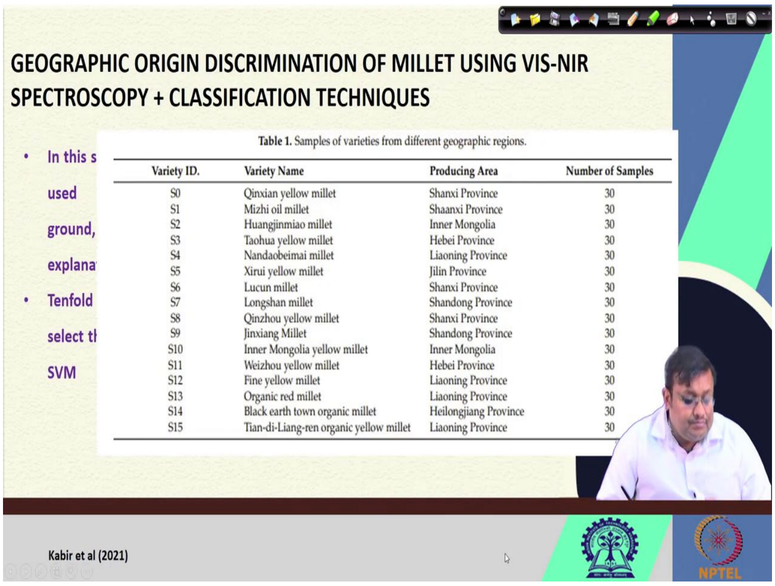
Therefore, based on the results, it can be concluded that Stages I and II were relatively easier to identify, especially Stage I.

Chakraborty et al (2017)

So, let us go ahead and see. So, based on these results overall we can see that it can be concluded that stage 1 and stage 2 were relatively easier to identify, especially stage 1 because stage 1 were highly concentrated in these lower values. So, we can, predict the stage 1 relatively easy.

So, in a nutshell, this shows that any classification algorithm is able to classify complex data with nonlinear boundaries to be very, interpretable outcome. So, we can interpret the data using this classification algorithm and using this type of classification plot in case of SVM and we have also seen the support vector SVM confusion matrix.

(Refer Slide Time: 25:41)



GEOGRAPHIC ORIGIN DISCRIMINATION OF MILLET USING VIS-NIR SPECTROSCOPY + CLASSIFICATION TECHNIQUES

- In this s
  used
  ground,
  explana
- Tenfold
  select tl
  SVM

Table 1. Samples of varieties from different geographic regions.

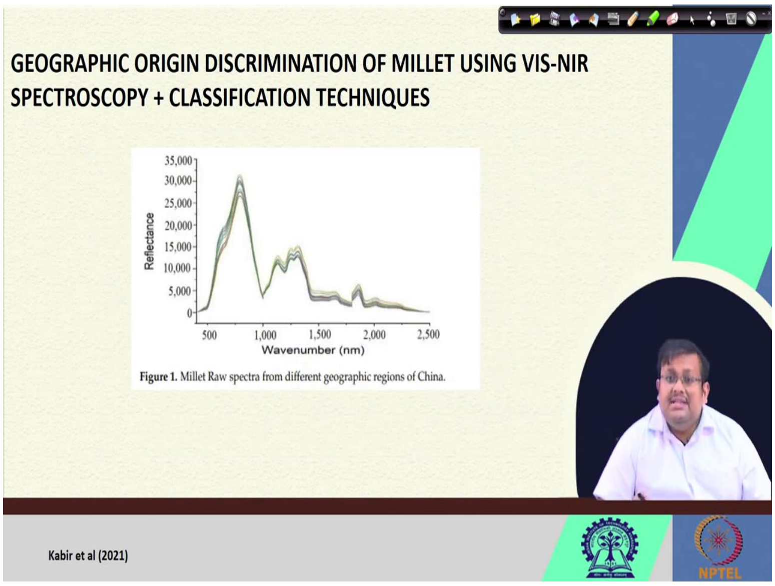
Variety ID.	Variety Name	Producing Area	Number of Samples
S0	Qinxian yellow millet	Shanxi Province	30
S1	Mizhi oil millet	Shaanxi Province	30
S2	Huangjinmiao millet	Inner Mongolia	30
S3	Taohua yellow millet	Hebei Province	30
S4	Nandaobeimai millet	Liaoning Province	30
S5	Xirui yellow millet	Jilin Province	30
S6	Lucun millet	Shanxi Province	30
S7	Longshan millet	Shandong Province	30
S8	Qinzhou yellow millet	Shanxi Province	30
S9	Jinxiang Millet	Shandong Province	30
S10	Inner Mongolia yellow millet	Inner Mongolia	30
S11	Weizhou yellow millet	Hebei Province	30
S12	Fine yellow millet	Liaoning Province	30
S13	Organic red millet	Liaoning Province	30
S14	Black earth town organic millet	Heilongjiang Province	30
S15	Tian-di-Liang-ren organic yellow millet	Liaoning Province	30

Kabir et al (2021)

Now, in this example, we can see that this is the geographic origin discrimination of millet using vis-nir spectroscopy and classification techniques. So, here are the millets where we 16 varieties of millets were classified were collected from or were collected. And in each of these variety there are 30 number of samples.

So, they are all collected from different areas. So, the idea of this paper by carburetor was to classify the samples based on their origin using their spectral data.

(Refer Slide Time: 26:31)



GEOGRAPHIC ORIGIN DISCRIMINATION OF MILLET USING VIS-NIR SPECTROSCOPY + CLASSIFICATION TECHNIQUES

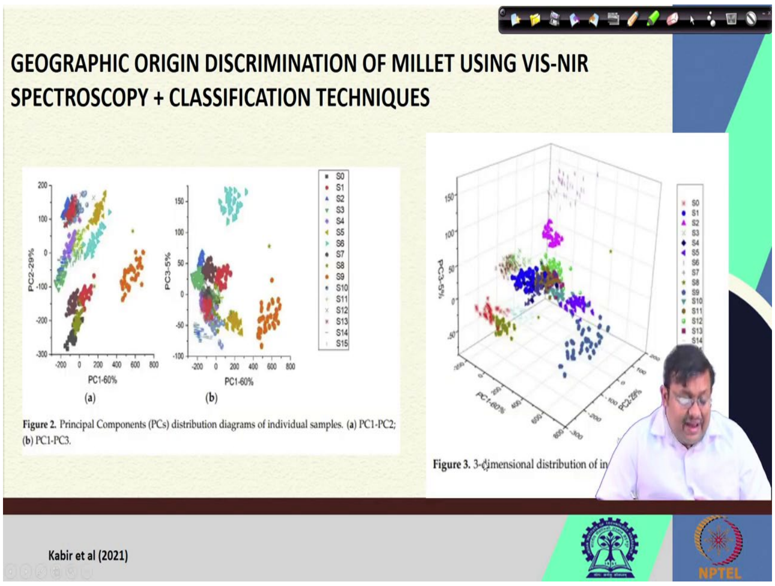
Figure 1. Millet Raw spectra from different geographic regions of China.

Kabir et al (2021)

So, once they have collected the millets they took the spectral reflectance values using the vis-nir spectroscopy, we are going to discuss what is vis-nir spectroscopy in our upcoming weeks and once the gate so, this is the millet raw spectra from different geographic regions of China. And once they get this raw spectra, they have calculated some they have done some spectral pretreatment, some spectral peated point is necessary to improve the signal to noise ratio.

And to get more information from the spectral data, we are going to discuss these in details in our upcoming weeks what is spectral people seeing, what are the different types of spectral preprocessing, we are going to discuss. But, so, we can see here this is the reflectance data and using the reflectance data of the millet and subsequent classification techniques, they try to classify the geographic origin of those millets varieties.

(Refer Slide Time: 27:38)



GEOGRAPHIC ORIGIN DISCRIMINATION OF MILLET USING VIS-NIR SPECTROSCOPY + CLASSIFICATION TECHNIQUES

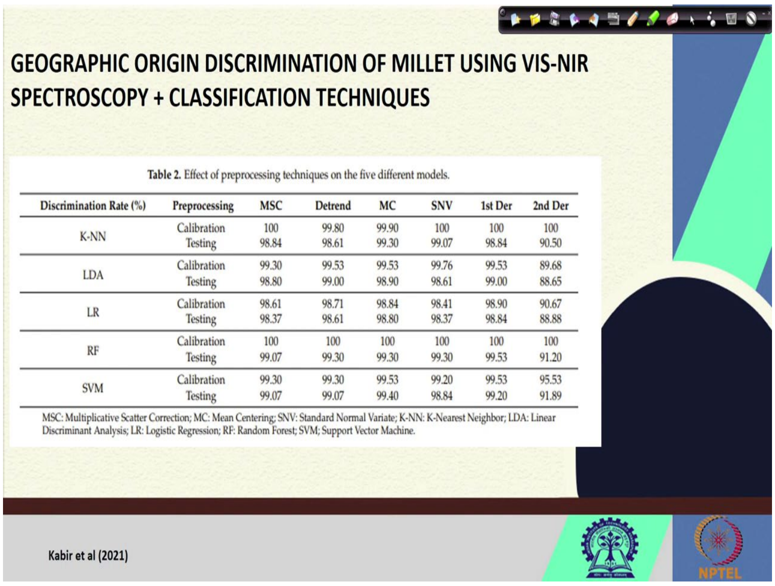
Figure 2. Principal Components (PCs) distribution diagrams of individual samples. (a) PC1-PC2; (b) PC1-PC3.

Figure 3. 3-dimensional distribution of in

Kabir et al (2021)

So, this is a principal component analysis which is you can see here that using PC 1 versus PC 2 and PC 1 versus PC 3, they try to cluster the samples into multiple classes, multiple classes and we can see the classes as they are originating in this PC space. And also this rightmost figure is showing the 3 dimensional distribution of the individual samples using the PC 1, PC 2, and PC 3, so that these shows the unsupervised classification of the samples on unsupervised clustering of the samples into different groups.

(Refer Slide Time: 28:25)



GEOGRAPHIC ORIGIN DISCRIMINATION OF MILLET USING VIS-NIR SPECTROSCOPY + CLASSIFICATION TECHNIQUES

Table 2. Effect of preprocessing techniques on the five different models.

Discrimination Rate (%)	Preprocessing	MSC	Detrend	MC	SNV	1st Der	2nd Der
K-NN	Calibration	100	99.80	99.90	100	100	100
	Testing	98.84	98.61	99.30	99.07	98.84	90.50
LDA	Calibration	99.30	99.53	99.53	99.76	99.53	89.68
	Testing	98.80	99.00	98.90	98.61	99.00	88.65
LR	Calibration	98.61	98.71	98.84	98.41	98.90	90.67
	Testing	98.37	98.61	98.80	98.37	98.84	88.88
RF	Calibration	100	100	100	100	100	100
	Testing	99.07	99.30	99.30	99.30	99.53	91.20
SVM	Calibration	99.30	99.30	99.53	99.20	99.53	95.53
	Testing	99.07	99.07	99.40	98.84	99.20	91.89

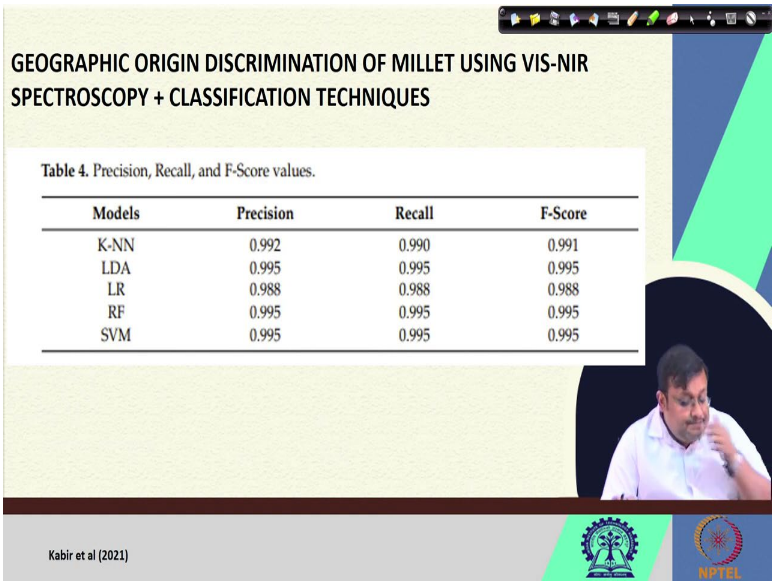
MSC: Multiplicative Scatter Correction; MC: Mean Centering; SNV: Standard Normal Variate; K-NN: K-Nearest Neighbor; LDA: Linear Discriminant Analysis; LR: Logistic Regression; RF: Random Forest; SVM; Support Vector Machine.

Kabir et al (2021)

Finally, they have used different classification and clustering techniques like k nearest neighbor, then LDA the then logistic regression, then random forest, and then they have also used the support vector machine. So, they try to here these are the results for different types of spectral pre-processing.

So, the pre-processing of like multiple scatter characterization, D trained then MC stand for the mean scattering, then standard normal variate first derivative, second derivative, different types of pre-processing they have done and then they have used this classification clustering algorithm. And this shows the discrimination did in the calibration, as well as in the testing samples.

(Refer Slide Time: 29:27)



GEOGRAPHIC ORIGIN DISCRIMINATION OF MILLET USING VIS-NIR SPECTROSCOPY + CLASSIFICATION TECHNIQUES

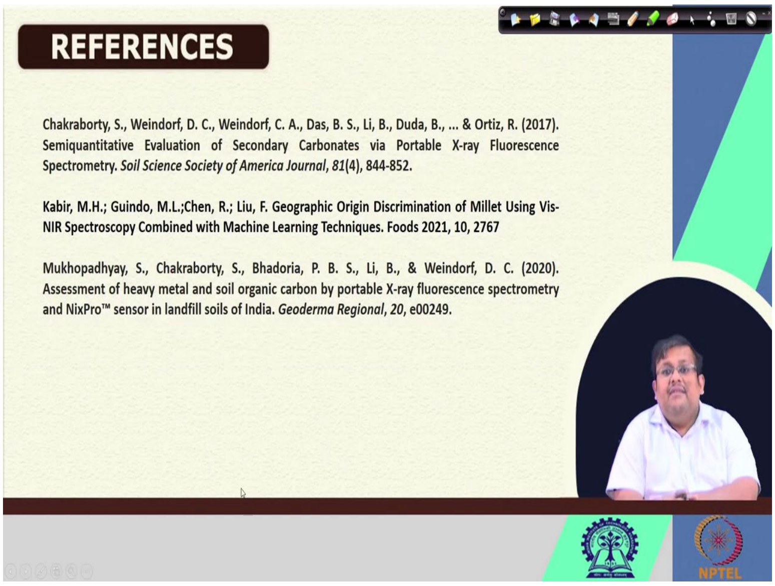
Table 4. Precision, Recall, and F-Score values.

Models	Precision	Recall	F-Score
K-NN	0.992	0.990	0.991
LDA	0.995	0.995	0.995
LR	0.988	0.988	0.988
RF	0.995	0.995	0.995
SVM	0.995	0.995	0.995

Kabir et al (2021)

And the precision recall and F scores were calculated which are the important performance metrics were calculated for all these different a classification and clustering algorithms. So, this shows the application of different types of clustering and classification algorithm for both soil and crop domain. We will see some more examples in our upcoming weeks.

(Refer Slide Time: 30:04)



REFERENCES

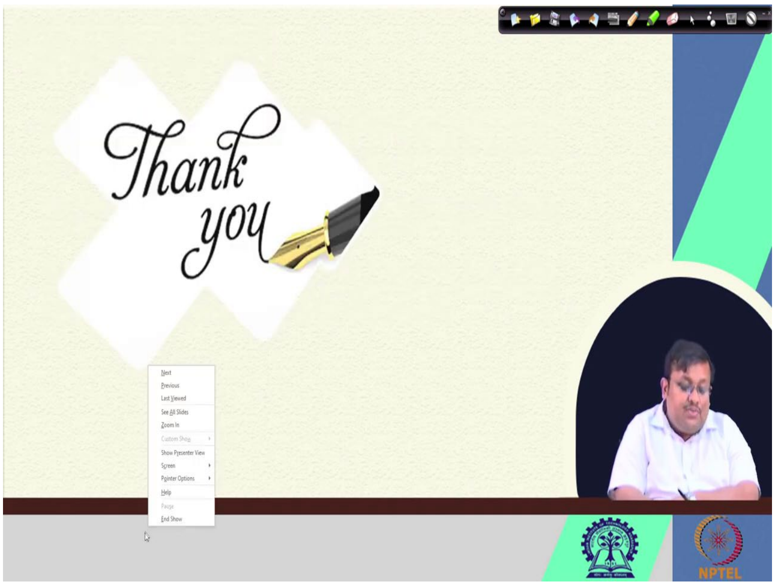
Chakraborty, S., Weindorf, D. C., Weindorf, C. A., Das, B. S., Li, B., Duda, B., ... & Ortiz, R. (2017). Semiquantitative Evaluation of Secondary Carbonates via Portable X-ray Fluorescence Spectrometry. *Soil Science Society of America Journal*, *81*(4), 844-852.

Kabir, M.H.; Guindo, M.L.;Chen, R.; Liu, F. Geographic Origin Discrimination of Millet Using Vis-NIR Spectroscopy Combined with Machine Learning Techniques. Foods 2021, 10, 2767

Mukhopadhyay, S., Chakraborty, S., Bhadoria, P. B. S., Li, B., & Weindorf, D. C. (2020). Assessment of heavy metal and soil organic carbon by portable X-ray fluorescence spectrometry and NixPro™ sensor in landfill soils of India. *Geoderma Regional*, *20*, e00249.

But I hope in this week, we have learned some important details for classification and clustering and some of their applications. So, these are the references which are in this lecture. If you have any queries, just let me know, just send me an email, I will be more than happy to answer your queries, you post your question in the forum, so that we can answer your queries.

(Refer Slide Time: 30:30)



And let us meet in our next week to discuss more in details about, other aspects of machine learning. So, the next week, we will be discussing a new aspect of this course. That is machine learning for soil and crop management. Thank you guys, let us meet in our next week lectures to discuss new new interesting topics. Thank you.