Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 44
Hyperspectral Remote Sensing and ML Applications in Agriculture (Contd.)

(Refer Slide Time: 0:22)

Welcome to this fourth lecture of week 9 of NPTEL Online Certification Course of Machine Learning for Soil and Crop Management. And in this week, we are discussing about Hyperspectral Remote Sensing and Machine Learning for agriculture and their applications. In our first 3 lectures, we have discuss the differences between hyperspectral remote sensing and multispectral remote sensing, characteristics of hyperspectral remote sensing we have discussed.

We have also discussed the different types of hyperspectral sensors which are which are developed and utilized by NASA, different types of spaceborne hyperspectral sensors as well as satellite hyperspectral sensors we have discussed and also we have discussed the hyperspectral sensors developed by Indian Space Research Organization.

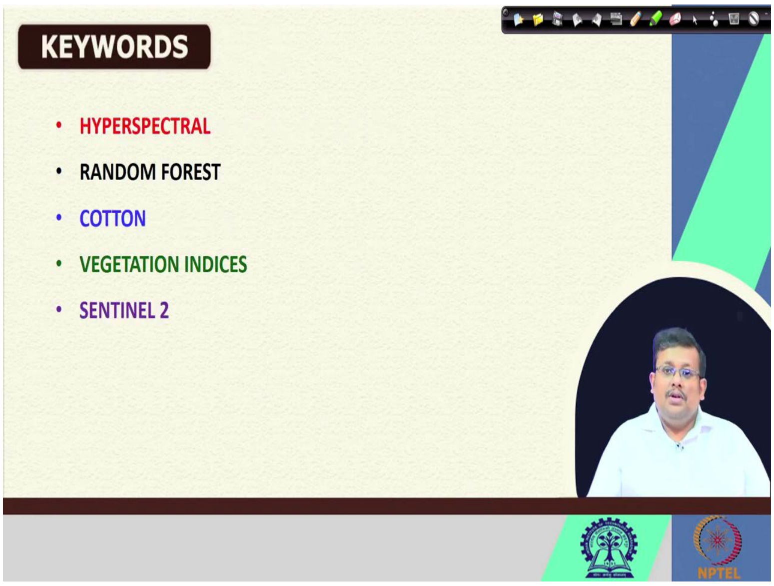
And also we have seen different types of data handling methods, how to download the data, from where to download the data and what are the softwares needed for handling the hyperspectral data we have discussed. And also we have discussed different types of scanner configuration, pushroom configuration and then whisk broom configuration we have discussed.

So, in this lecture, we are going to discuss the hyperspectral remote sensing application focusing on crop characterization and crop nitrogen estimation specifically cotton estimation and we are going to discuss one case study which was published in a reputed journal.

(Refer Slide Time: 2:23)

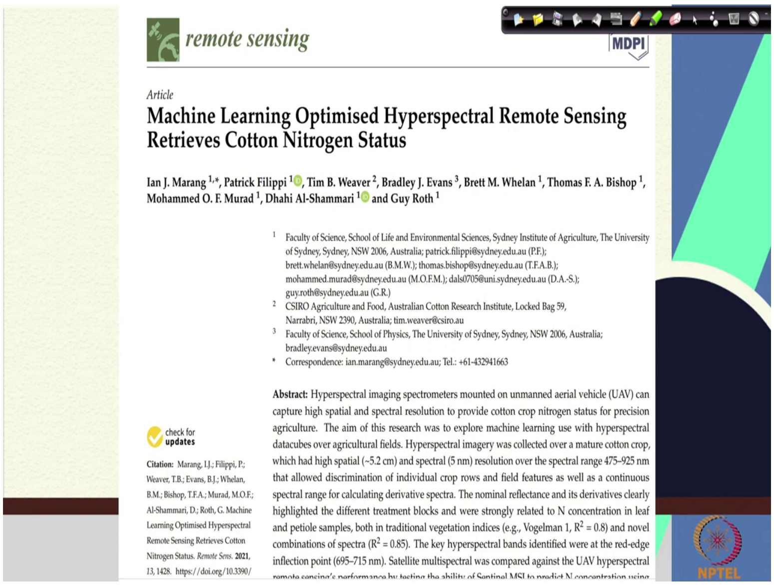
So, in this lecture, we are going to discuss these hyperspectral remote sensing crop application a case study. Also we are going to discuss what is red edge absorption and why it is important and also we are going to see different types of vegetation indices which were calculated in this case study.

(Refer Slide Time: 2:46)



So, these are the keywords which we are going to discuss, we are going to discuss hyperspectral, random forest then cotton, vegetation indices, Sentinel 2. These are some of the keywords for this lecture.

(Refer Slide Time: 3:04)



remote sensing

MDPI

Article

Machine Learning Optimised Hyperspectral Remote Sensing Retrieves Cotton Nitrogen Status

Ian J. Marang [1,*], Patrick Filippi [1], Tim B. Weaver [2], Bradley J. Evans [3], Brett M. Whelan [1], Thomas F. A. Bishop [1], Mohammed O. F. Murad [1], Dhahi Al-Shammari [1] and Guy Roth [1]

1 Faculty of Science, School of Life and Environmental Sciences, Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW 2006, Australia; patrick.filippi@sydney.edu.au (P.F.); brett.whelan@sydney.edu.au (B.M.W.); thomas.bishop@sydney.edu.au (T.F.A.B.); mohammed.murad@sydney.edu.au (M.O.F.M.); dals0705@uni.sydney.edu.au (D.A.-S.); guy.roth@sydney.edu.au (G.R.)
2 CSIRO Agriculture and Food, Australian Cotton Research Institute, Locked Bag 59, Narrabri, NSW 2390, Australia; tim.weaver@csiro.au
3 Faculty of Science, School of Physics, The University of Sydney, Sydney, NSW 2006, Australia; bradley.evans@sydney.edu.au
* Correspondence: ian.marang@sydney.edu.au; Tel.: +61-432941663

check for updates

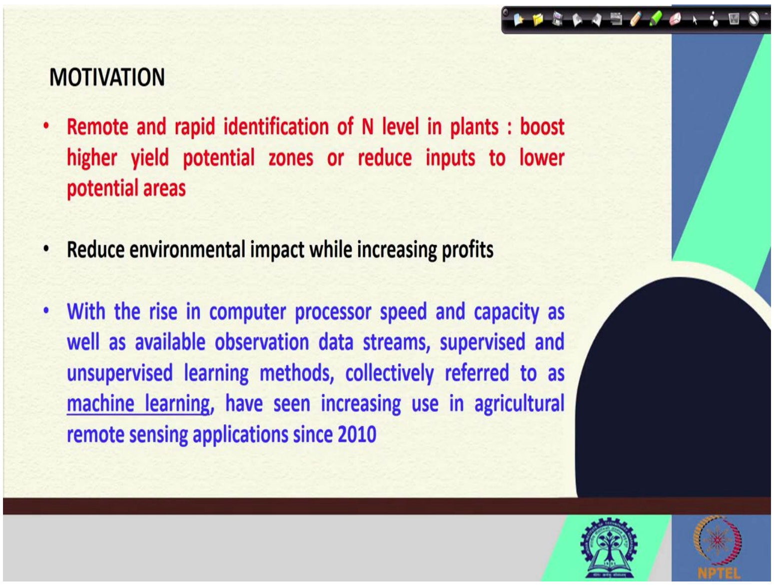
Citation: Marang, I.J.; Filippi, P.; Weaver, T.B.; Evans, B.J.; Whelan, B.M.; Bishop, T.F.A.; Murad, M.O.F.; Al-Shammari, D.; Roth, G. Machine Learning Optimised Hyperspectral Remote Sensing Retrieves Cotton Nitrogen Status. *Remote Sens.* **2021**, 13, 1428. https://doi.org/10.3390/

Abstract: Hyperspectral imaging spectrometers mounted on unmanned aerial vehicle (UAV) can capture high spatial and spectral resolution to provide cotton crop nitrogen status for precision agriculture. The aim of this research was to explore machine learning use with hyperspectral datacubes over agricultural fields. Hyperspectral imagery was collected over a mature cotton crop, which had high spatial (~5.2 cm) and spectral (5 nm) resolution over the spectral range 475–925 nm that allowed discrimination of individual crop rows and field features as well as a continuous spectral range for calculating derivative spectra. The nominal reflectance and its derivatives clearly highlighted the different treatment blocks and were strongly related to N concentration in leaf and petiole samples, both in traditional vegetation indices (e.g., Vogelman 1, $R^2 = 0.8$) and novel combinations of spectra ($R^2 = 0.85$). The key hyperspectral bands identified were at the red-edge inflection point (695–715 nm). Satellite multispectral was compared against the UAV hyperspectral

So, this is the case study or the other published paper in remote sensing journal. You can see it is a machine learning optimized Hyperspectral Remote Sensing Retrieve Cotton Nitrogen Status. So, in this research, the researcher tried to estimate the cotton nitrogen status and also the map cotter nitrogen status using hyperspectral remote sensing data in combination with different machine learning approaches.

So, this is an open source journal. So, you can search it in the internet and you can get it free and I would request you to go through it in details to have more comprehensive information, comprehensive knowledge about the steps which they have followed. So, in this lecture, we are going to briefly discuss the important points or the takeaway points from this study.

(Refer Slide Time: 4:15)



So, first question comes what was the motivation of the study? Now, you all know that remote and rapid identification of nitrogen level in plants can help in boosting the higher yield in the potential zones and also it can reduce the unnecessary application of nitrogenous fertilizer. Nitrogenous fertilizers are being applied in huge amount of huge quantity by the farmers to boost the growth of the crop.

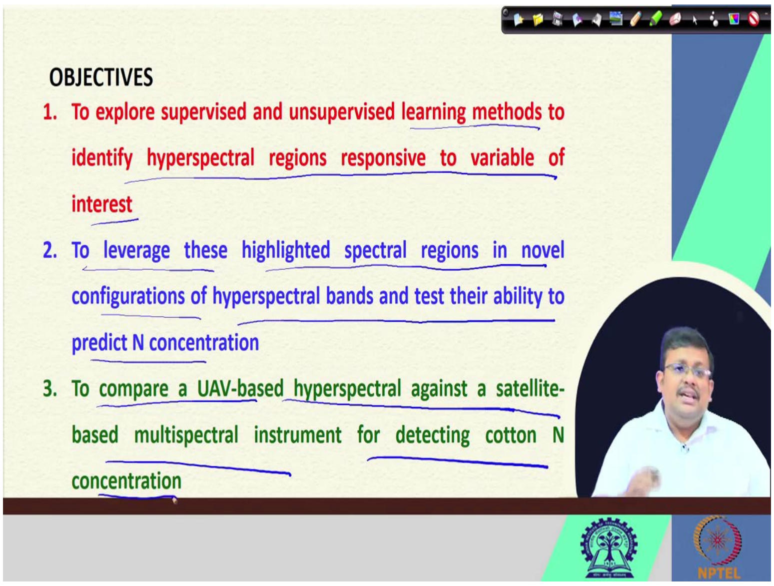
But if we can identify the areas, which require higher nitrogen content, which require higher nitrogen inputs, which can reduce the nitrogen application by adjusting our nitrogen doses for the whole area and by giving less amount of nitrogen for those patches which we do not need high amount of nitrogen application for which we do not need any high amount of nitrogen application.

So, what happens when we go for these judicious plant requirement based nitrogen as application that helps in reducing the environmental impact while increasing the profit because you are now cutting down your nitrogen inputs and cost and simultaneously is the cost involved and also simultaneously that nitrate is a potential environmental pollutant.

And when you apply judicious amount of nitrogenous fertilizer that can reduce the environmental impact because otherwise that will these nitrate will go away through runoff and that will go to the different types of water bodies and it will mix with the groundwater creating environmental hazards.

So, we have seen that with the rise in computer codes and processes speed and also different types of advanced machine learning algorithm there has been a tremendous increase in this hyperspectral remote sensing based crop characterization and specifically agricultural operations since 2010.

(Refer Slide Time: 6:37)

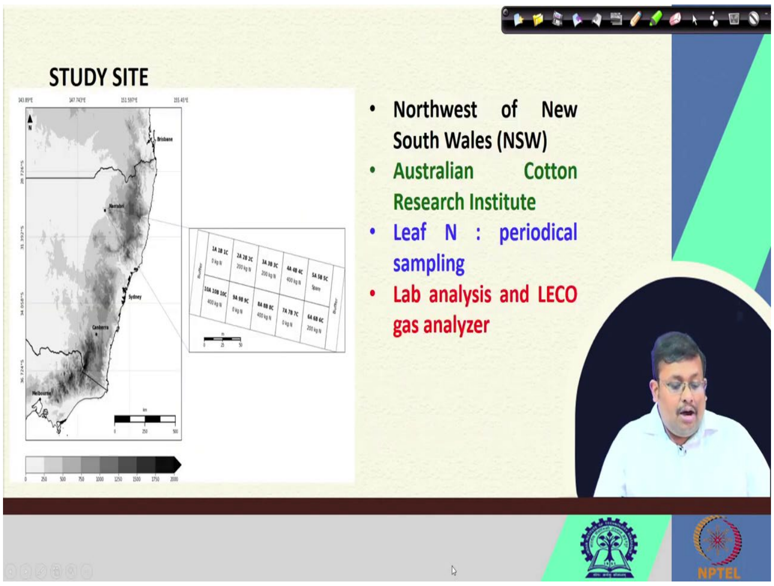


So, in this study, they had three major objective. First of all, they try to explore different supervised and unsupervised learning methods to identify the hyperspectral regions responsive to variable of interest, what is it is crop nitrogen content. So, they try to use some supervised and unsupervised learning methods to identify these hyperspectral regions, which are very much responsive to the plant nitrogen content.

And secondly, they try to leverage these highlighted spectral regions in novel configuration of hyperspectral bands and test their ability to predict the nitrogen concentration. So, using these selected bands, so first of all, they tried different types of supervised and unsupervised method to identify the important bands from the hyperspectral data, which are responsive to nitrogen concentration of the cotton.

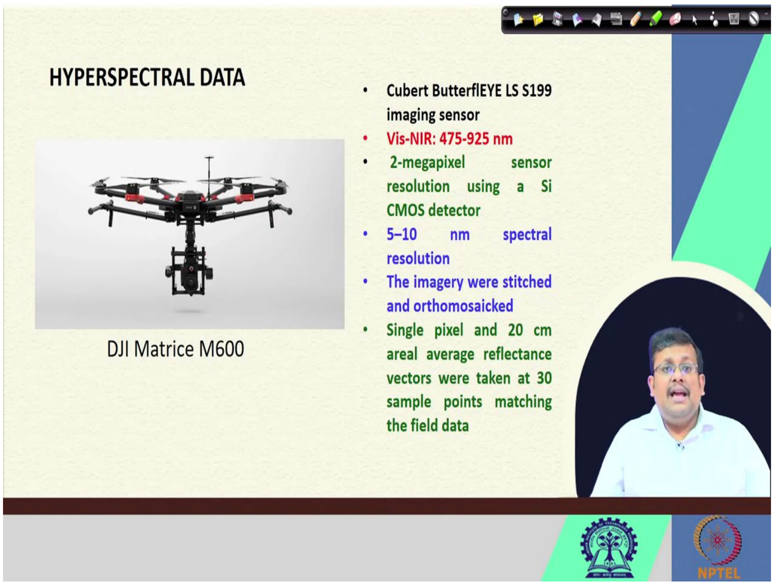
And secondly, they have used those bands to calculate a novel configuration of a novel or I would say it is a vegetation index. And then they try to predict use the nitrogen content of the crop using those vegetation using those vegetation indices. And thirdly, they tried to compare a UAV-based hyperspectral data against a satellite-based multispectral instrument for detecting cotton nitrogen concentration. So, these three were the major objectives for their study, and they have used both hyperspectral as well as multispectral data.

(Refer Slide Time: 8:25)



Now, this was the study site, it is the Northwest this study site is situated in the northwest of New South Wales, and it is basically Australian Cotton Research Institute at Nairobi. And they basically sample the leaves of the cotton periodically and after taking the samples, they analyze the leaf nitrogen content by standard laboratory analysis and LECO gas analyzer.

(Refer Slide Time: 8:59)



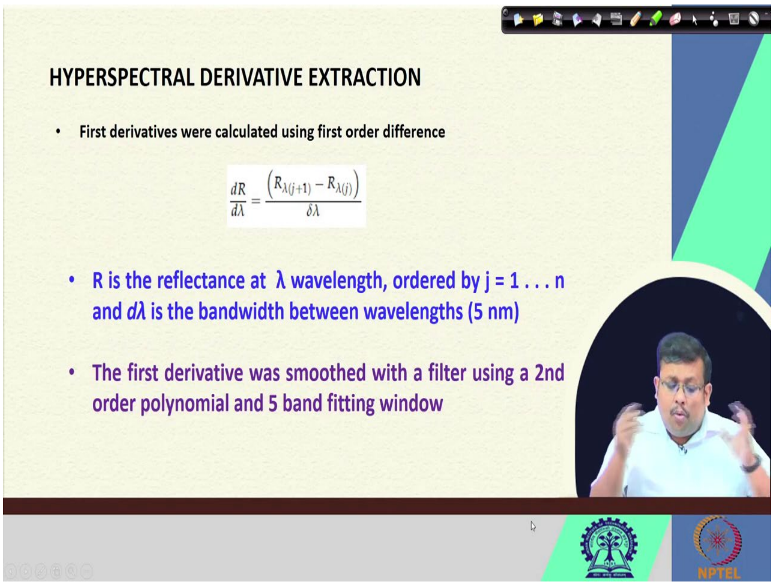
Now, once they did that, at the same time, they are also captured the hyperspectral data using these DJI matrices M600 drone and using these drone they have taken the hyperspectral images and for taking the hyperspectral images they have used these Cubert Butterfly LS S199 imaging sensor and this imaging sensor has a wavelength range of 475 to 925 nanometer with a 2 megapixel sensor resolution using a silicon CMOS detector and they, this sensor had the spectral resolution of 5 to 10 nanometer and after they took the images.

Those images were stitched together and also orthomosaicked, what is orthomosaicing, we have already discussed. Now after Orthomosaicing they have I, they have taken the hyperspectral data into fashion. In one way they have taken the sample they take it they have extracted the pixel spectrum for 30 pixel from which they have taken randomly planned sample.

So, initially they have they took 30 samples from 30 spots and from the hyperspectral imagery they have taken, they have extracted the spectral information from those 30 individual pixels. So, this is one way. In another approach what they have done, they have also taken areal average reflectance vector of for a 20 centimeter aerial average reflectance vectors. So, they have taken an average reflectance also for 20 centimeter aerial coverage.

So, they have compared both these average spectrum as well as they have covered also the individual pixel based extracted spectrum.

(Refer Slide Time: 11:06)

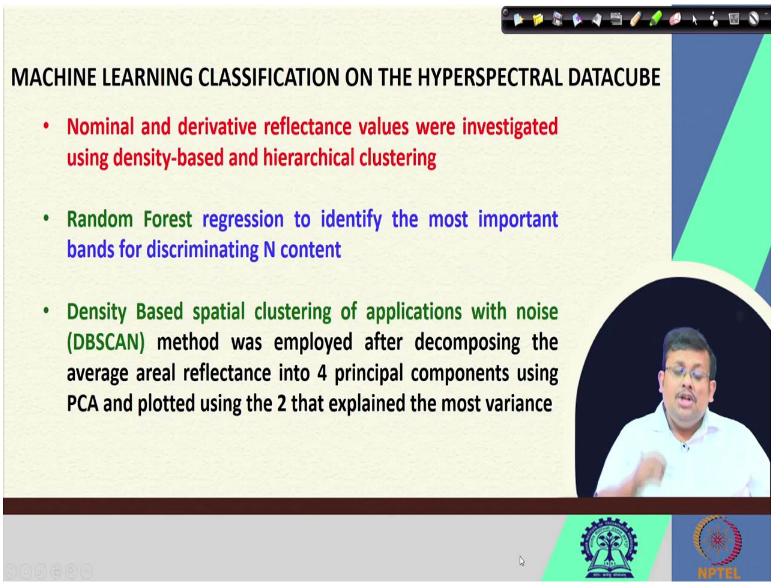


So, once they have extracted the spectrum, they did the first derivative calculation. Now, we have already discussed different types of first, different types of spectral preprocessing in our week 5. Now, remember that this first derivative were calculated using this first order differential equation.

So, here dR by d lambda, R stands for the reflectance and so, basically using this formula they have taken they have calculated the first derivative of the reflectance spectra. So, R is here the reflectance and lambda is the wavelength, which is ordered from j equal to 1 to n and d lambda is a bandwidth between the wavelengths.

Now, in this case the bandwidth was 5 nanometer. Now, the first derivative after taking the first derivative of the reflectance spectrum, the spectra was further smoothed with a filter using a second order polynomial and a 5 band fitting window. We have already discussed these sectors smoothing before.

(Refer Slide Time: 12:11)



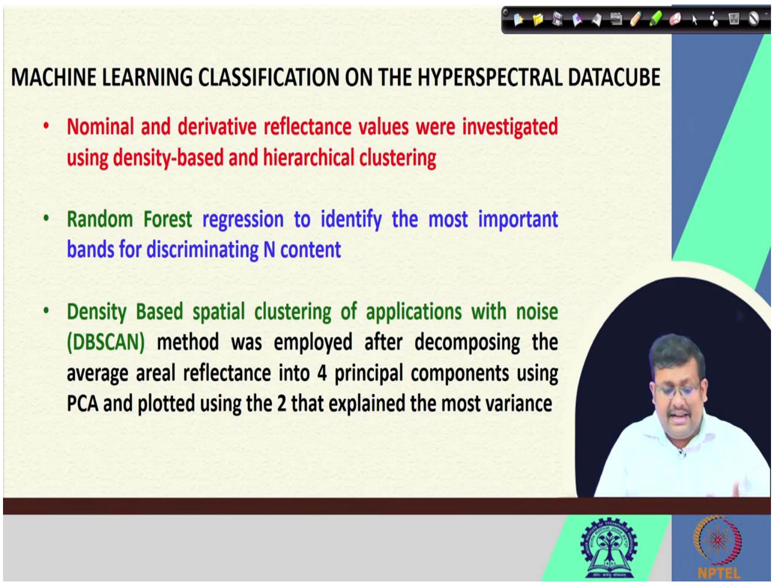
So, they have used both the nominal spectrum that is the original spectrum at the same time they have used both the, also they have also used this derivative spectrum. So, they have compared the performance for both the nominal or original spectra as well as the deviating spectrum. So, nominal and derivative reflectance values were investigated using density based hierarchical clustering.

And at the same time, we are going to discuss these hierarchical clustering, but at the same time, they also did random forest regression to identify the most important band for discriminating the nitrogen content. And, the density based hierarchical clustering which I have told you also they have done the density based spatial clustering application with noise. So, the full name is DBSCAN, so, density based spatial clustering application with noise.

So, this method they have tried after decomposing the average reflectance into 4 principal components using PCA principal component analysis and plotting them using the 2 that explained the most variance. So, what happens using the original reflectance they have done the principal component analysis and after they did the principal component analysis, they selected the first 2 principal components and using this has 2 principal components they plot they plotted the scatterplot or using the first 2 principal components that will the score plot.

And based on the score plot, they try to identify the important spectral regions and that noise in the spectral region. The important spectral region is the spectral region which is more correlated to the nitrogen content or so, and also they have tried to identify the noise in the spectrum. So, these are the different types of machine learning classification on the hyperspectral detective they have used.

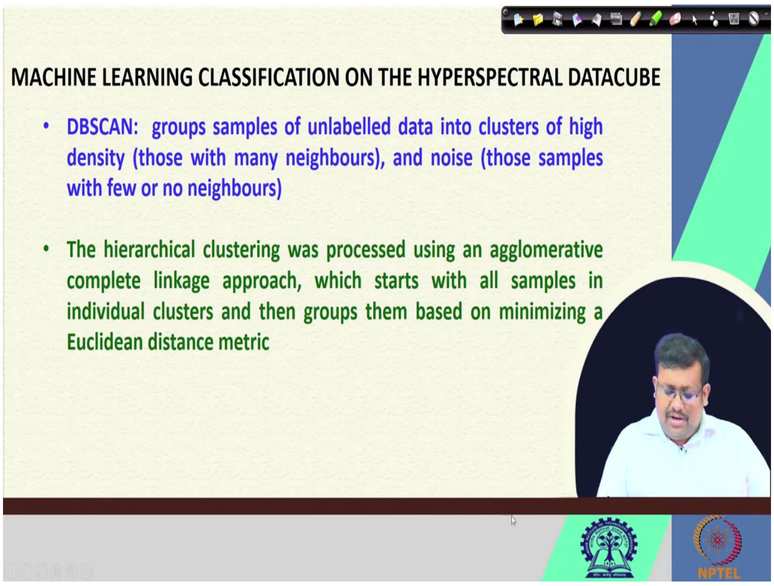
(Refer Slide Time: 14:08)



So, what does this DBSCAN does, so, it basically group the samples of unlabeled data into clusters of high density. So, we have several samples. So, if we can cluster this data through PCA, then these DBSCAN basically try to identify the region of high density that means, those with the many neighbors and also some noise those samples with fewer neighbors. So, from there they try to identify, this region is important and this is that noise.

So, this and also from your previous knowledge that hierarchical clustering is it is a clustering method, and in this research, they use the hierarchical clustering using an agglomerative complete linkage approach. We starts with all samples in individual clusters and then groups based on the minimizing and Euclidean distance metric.

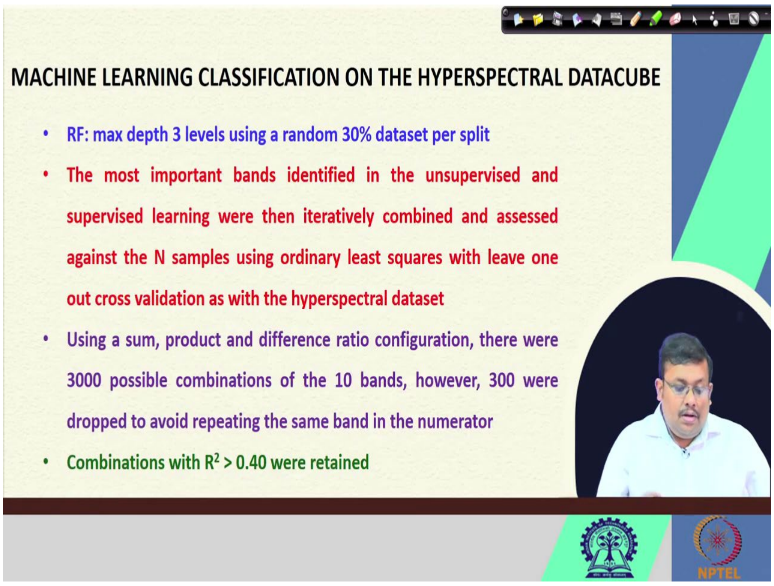
(Refer Slide Time: 15:08)



We have already discussed this, if you can recall from our previous lectures, the clustering methods, these hierarchical clustering method use a set of nested clusters organized as a hierarchical tree. And there are 2 types, one is agglomerative and other is divisive. So, you can see this in hierarchical, a hierarchical is of course, a bottom up approach. The, from the bottom it goes to the upward direction by calculating the Euclidean distance.

So, this is the hierarchical clustering with agglomerative approach or in other words, we also call it agglomerative hierarchical clustering. So, in this method also, they have tried to use these agglomerative hierarchical clustering to cluster the data, to cluster the spectrum or the wavelengths to be precise to cluster the wavelengths which are close together and then they try to correlate them with the nitrogen content.

(Refer Slide Time: 16:19)



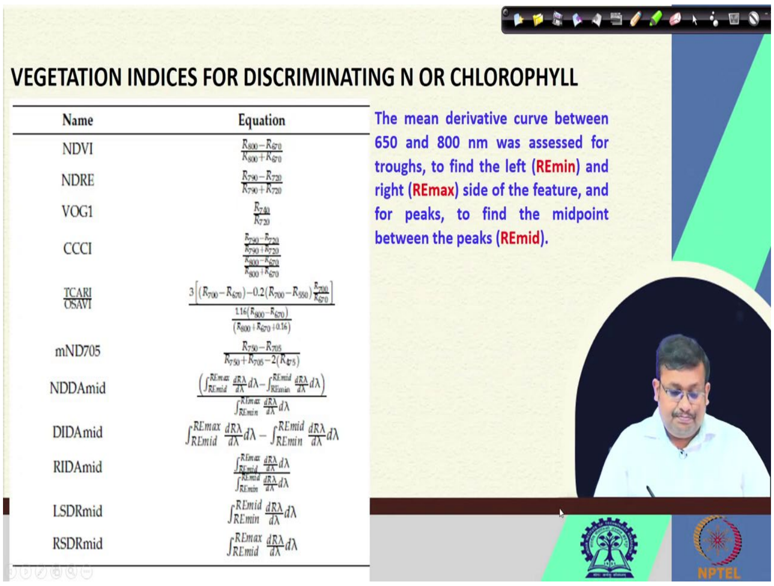
Now, in case of random forests, they have tried maximum three depth levels using a random 30 percent data set per split. And the most important bands identified in the unsupervised and supervised learning were then iteratively combined and assessed against the nitrogen samples using ordinary least squares with leave one out cross validation as with the hyperspectral dataset. So, using a sum product and different ratio configuration, there were 3000 possible combination of 10 band.

So, once I selected the 10 important bands, using these different types of methods, there are different 3000 different types of combination for different types of vegetation indices, they have developed, however, 300 were dropped to avoid repeating the same band in the numerator and only those vegetation indices were kept with R square value of greater than 0.40.

So, this is the methodology they have used, we have already discussed all these important concepts before, we have discussed principal component analysis. So, the principal component analysis, we have discussed the hierarchical clustering also. So, the hierarchical clustering, you also we have also discussed the random forests. So, the Random Forests too.

So, all these which are required to understand the results of the study, we have already discussed. Now, we are focusing on the results only.

(Refer Slide Time: 17:57)

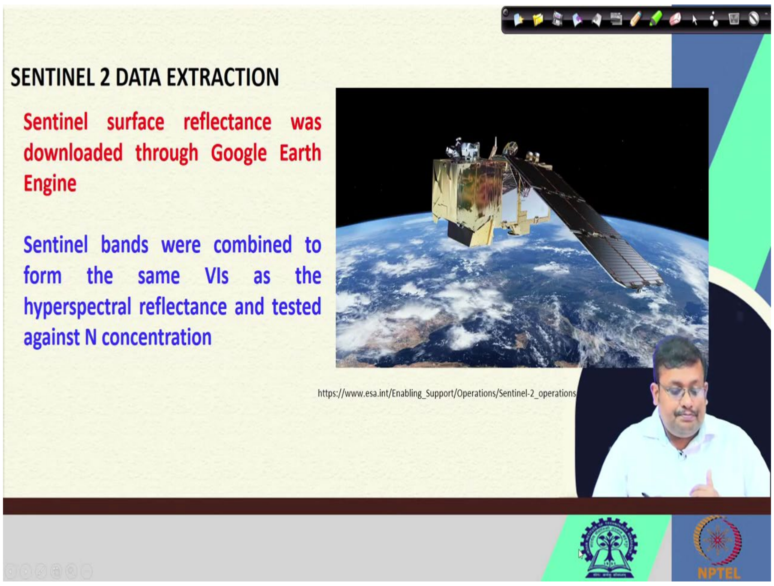


So, these are the vegetation indices, which they have calculated. So, NDVI, NDRE, we have already discussed these things previously, and also some VOG1, CC1 and all these different types of vegetation indices they have calculated and their respective equations. So, here you can see one term that is called REmin, another term is called REmax. So, here you can see REmin, REmax and also REmid.

So, the mean derivative curve between these 650 and 800 nanometer was assessed for troughs and to find the left and right side of the feature and for the peaks to find a midpoint between the peaks. So, while we are talking about these REmax, REmin, REmid, we are trying to identify the leftmost and the rightmost trough like features in the 650 to 800 nanometers zone and in case of REmid we are trying to see the midpoint between the peaks.

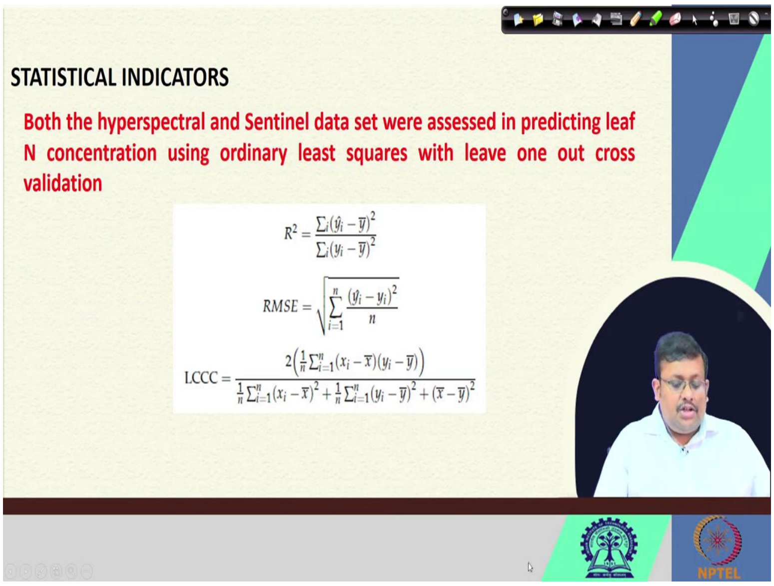
I am going to show you how they have calculated this and why they have calculated this we have also we are also going to discuss.

(Refer Slide Time: 19:18)



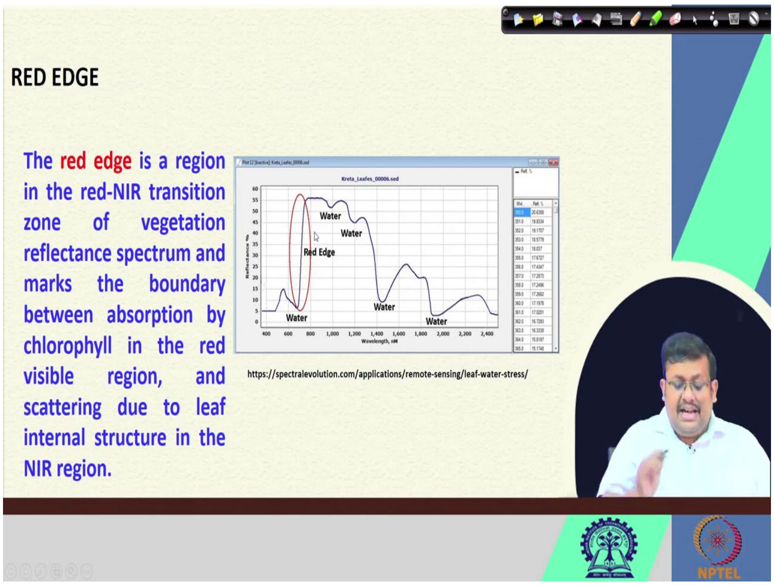
Simultaneously apart from this hyperspectral data, they have also collected the Sentinel 2 data. Now, the Sentinel 2 is a multispectral sensor and the Sentinel surface reflectance was downloaded through Google Earth Engine and the Sentinel bands were combined to form the same vegetation indices as the hyperspectral reflectance and tested against the nitrogen concentration. So, they have compared both the hyperspectral vegetation indices and also the multispectral vegetation indices to correlate with the nitrogen content of the cotton crop.

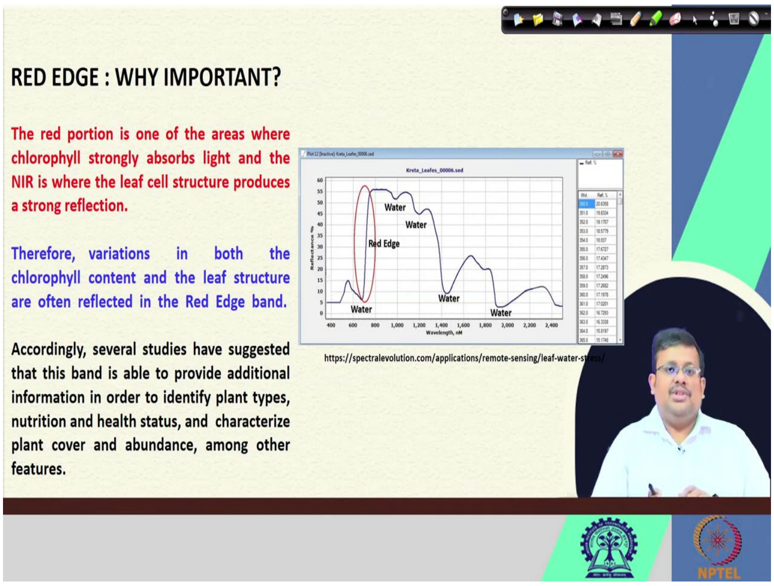
(Refer Slide Time: 20:03)



So, they have used different types of statistical indicators. These are common that is R square value RMSE and lin’s concordance correlation coefficient. So, these R square value they have calculated RMSE value is calculated, lin's concordance correlation coefficient was calculated using this formula. So, both the hyperspectral and the Sentinel data were assessed in predicting the leaf nitrogen content using the ordinary least squares with the leave one out cross validation and these were the statistical indicators.

(Refer Slide Time: 20:33)





Now, if you see the results, in the results it is quite clear that the left plot is showing the hyperspectral reflectance of each geo reference sample points. I told you that there are 30 pixels from where they have extracted the spectral information. So, these are those 30s spectrum for extracted from 30 pixels. And here in the second case, they have also took the average of pixel within 20 centimeter, I told you from the aerial 20 centimeters.

So, they have compared both and you can clearly see that in the first case there is too much noise in this zone. So, to remove the noise or to remove the noise and not to consider these noise in the model they have they did all the subsequent analysis using the average pixel within 20 centimeter spectral, average spectrum which were which are collected or the spectrum which were collected from the average of pixel within 20 centimeter.

So, this is how this is showing the change in reflectance or I would say this basically shows that how the approach of spectrum in extraction from the hyperspectral data can have an important impact on that result. So, here they have taken these average of pixels within 20 centimeter for subsequent analysis.

Now, before we discuss the next step, it is also important to discuss one important concept that is called red edge. Now, red edge is a region so, this is you can see this is a plant reflectance or plant part or leaf reflectance. Now, here you can see that water band and red edge band this portion is known as the Red Edge band and these are the water absorption bands.

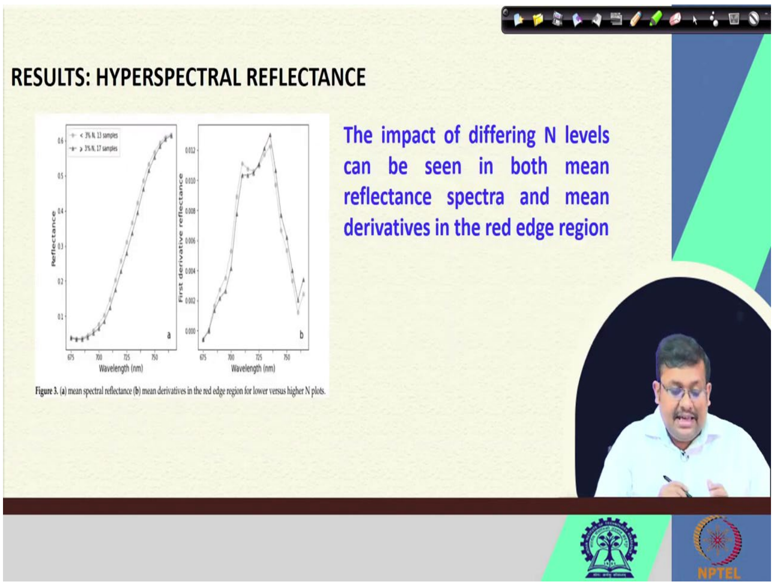
So, that red age is a region in the red and NIR region, so here this is the red region and this is the NIR region and in the transition between red to NIR you will see these red edge band is there. So it is basically these red edge band just generally we can see in the transition zone between the red and NI. So, this red edge is a region and in the red-NIR transition zone of vegetation reflectance spectrum and marks the boundary between absorption by chlorophyll in the red visible region and scattering due to leaf internal structure in the NIR region.

So, two types of events occur. In the red region generally there is an absorption by the chlorophyll and in the NIR region there is reflection by the scattering, but in the NIR region there is a scattering due to leave internal structure. So, you can see there is a clear changes in the reflectance pattern in this transition zone. So, generally these red is utilized in different crops studies, why?

Because the rate portion is one of the areas where chlorophyll strongly absorbs light and NIR is where the leaf cell structure produces a strong reflection. So, the variation in both the chlorophyll content and the leaf structure are often reflected in the red age band. So, here in these red edge band, if you consider this red edge band, it gives a clear indication of the variation of the chlorophyll content at the same time, the leaf structure also.

So, accordingly several studies have suggested that these band is able to provide additional information in order to identify the plant types, nutrition and health status and characterized plant cover and the abundance among other features. So, to identify the healthy plants, to identify the nutrition status, it is important to focus on these red edge absorption features. So, why we are telling these red edge absorption features?

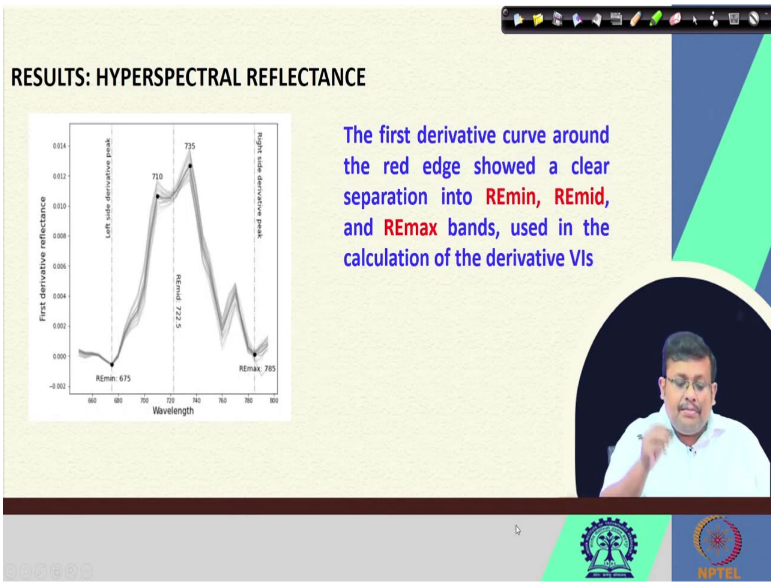
(Refer Slide Time: 25:07)



You will see that in this study also they have utilized the red edge, they have extracted these red edge they have isolated these red edge portion and you can see this is due to the original reflectance or mean spectral reflectance and this is the derivative spectrum. And you can clearly see that the impact of differing nitrogen level can be observed in both mean reflectance and also in the mean derivative in the red edge region.

Of course, you can see that when the nitrogen condenser there are two, we have grouped that they have grouped the samples into 13 samples with less than 3 percent nitrogen and 17 samples with greater than 3 percent nitrogen. So, you can clearly see that, these the reflectance pattern or mean spectral reflectance for these two groups are showing the difference in these zone clearly showing the difference in this zone in both the original reflectance as well as in the NIR region. So, that shows that why these red edge is considered as an important indicator of plant nutrition content.

(Refer Slide Time: 26:30)



Now, I told you about these REmin, REmid and REmax. So, the first derivative curve around these red edge was utilized to identify these REmin, REmid and REmax bands used in the calculation of the derivative vegetation indices. So, here you can see REmin, so, this is one trough and this is another trough, so, REmin, this is REmax and these are the two peaks. So, mid, REmid was calculated at around 722.5 nanometer. So, they use these values to calculate the vegetation indices based on the derivative spectrum.

(Refer Slide Time: 27:16)

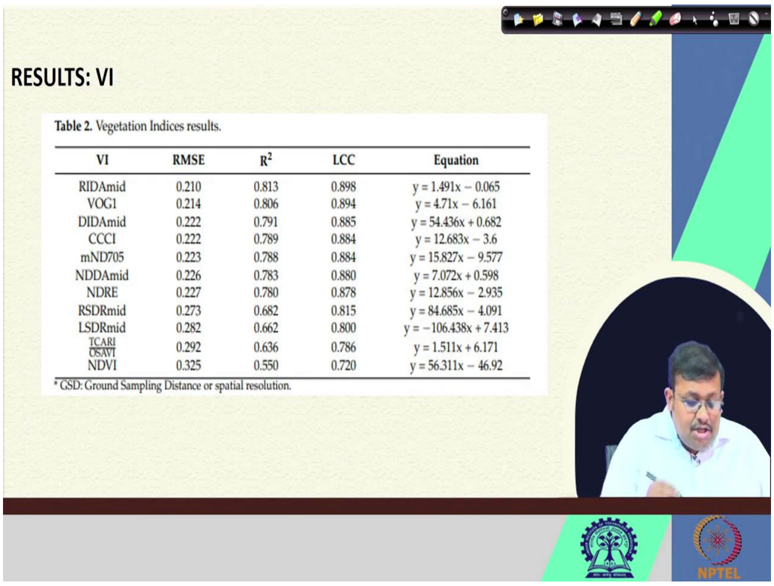
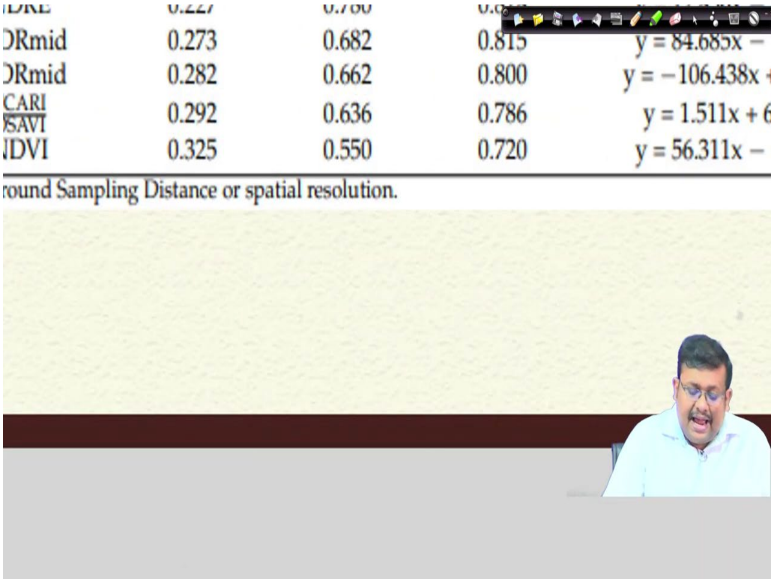


Table 2. Vegetation Indices results.

VI	RMSE	R^2	LCC	Equation
RIDAmid	0.210	0.813	0.898	y = 1.491x − 0.065
VOG1	0.214	0.806	0.894	y = 4.71x − 6.161
DIDAmid	0.222	0.791	0.885	y = 54.436x + 0.682
CCCI	0.222	0.789	0.884	y = 12.683x − 3.6
mND705	0.223	0.788	0.884	y = 15.827x − 9.577
NDDAmid	0.226	0.783	0.880	y = 7.072x + 0.598
NDRE	0.227	0.780	0.878	y = 12.856x − 2.935
RSDRmid	0.273	0.682	0.815	y = 84.685x − 4.091
LSDRmid	0.282	0.662	0.800	y = −106.438x + 7.413
$\frac{TCARI}{OSAVI}$	0.292	0.636	0.786	y = 1.511x + 6.171
NDVI	0.325	0.550	0.720	y = 56.311x − 46.92

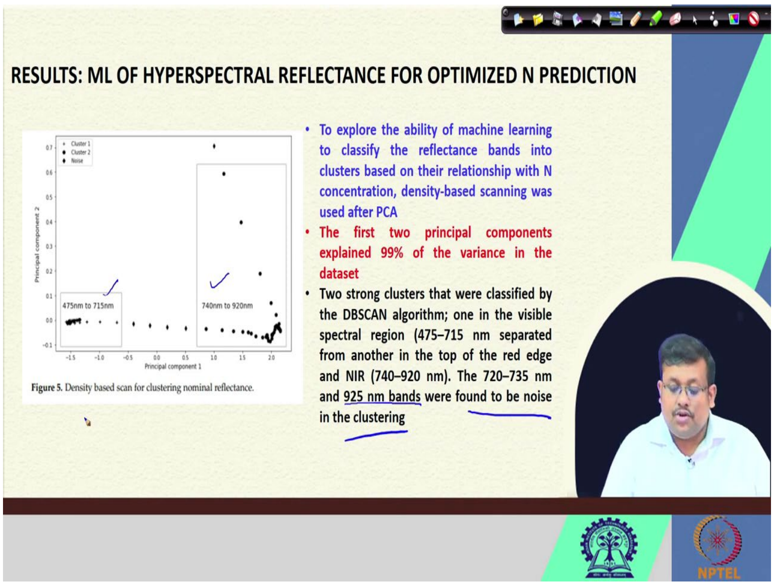
* GSD: Ground Sampling Distance or spatial resolution.



So, if you see the results of nitrogen content estimation using different types of vegetation indices, you can clearly see that using different types of vegetation indices we are we got they have got satisfactory more or less satisfactory prediction performance. Of course, you can see these RIDAmid which give the highest R squared of 0.81 followed by these VOG1 0.80.

Then DIDAmid, so, different types of vegetation indices, they have tried to predict the nitrogen content extracted from the hyperspectral data and then followed by different types of processing. And once they use this, they have seen that these vegetation indices can be used satisfactorily to predict the nitrogen content of the cotton crop.

(Refer Slide Time: 28:19)

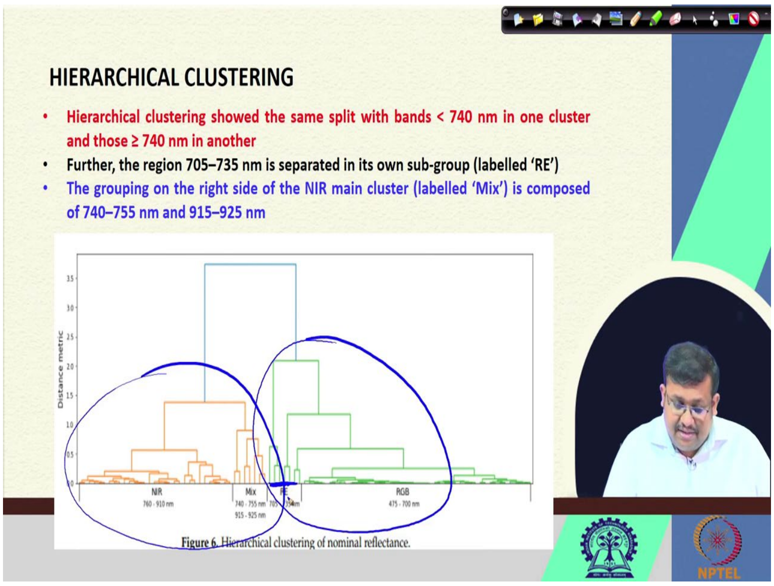


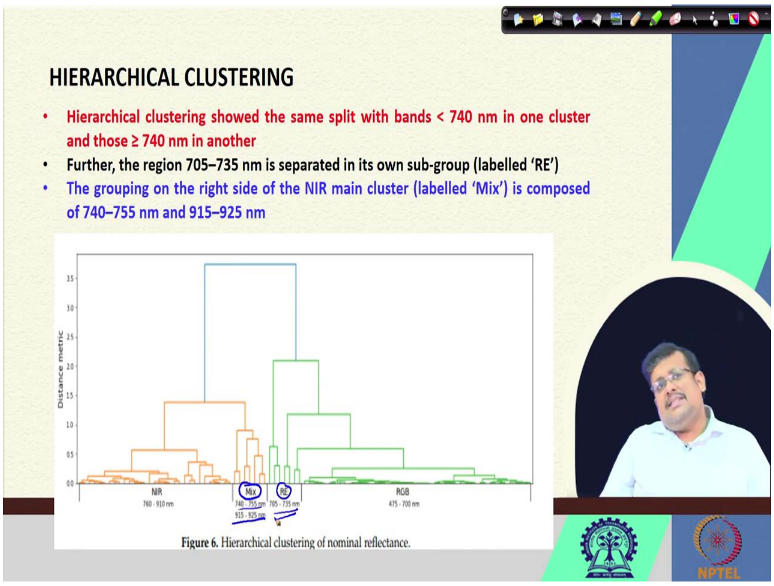
So, they have also used the principal component analysis I told you, so, you can see here this principal component 1 and principal component 2. So, to explore the ability of the machine learning to classify the reflectance bands into clusters based on their relationship with nitrogen concentration, this density based clustering was used followed by after PCA.

Now, you can see this first principal component and the second principal component cumulatively explained 99 percent of the total variance. And you can see here two strong clusters, here the two strong clusters that were identified these DBSCAN algorithm, one is this 475 to 750 nanometer clustered, another one is 740 to 920 meter region and 720, in between the 720 to 735 nanometer and 920 nanometer bands were found to be noise in the clustering which are in between.

So, you can see here based on their closeness, we can see that, their relationship, we can establish their relationship with nitrogen concentration based on this type of clustering, density based clustering, which we do generally by using the principal component analysis.

(Refer Slide Time: 29:44)



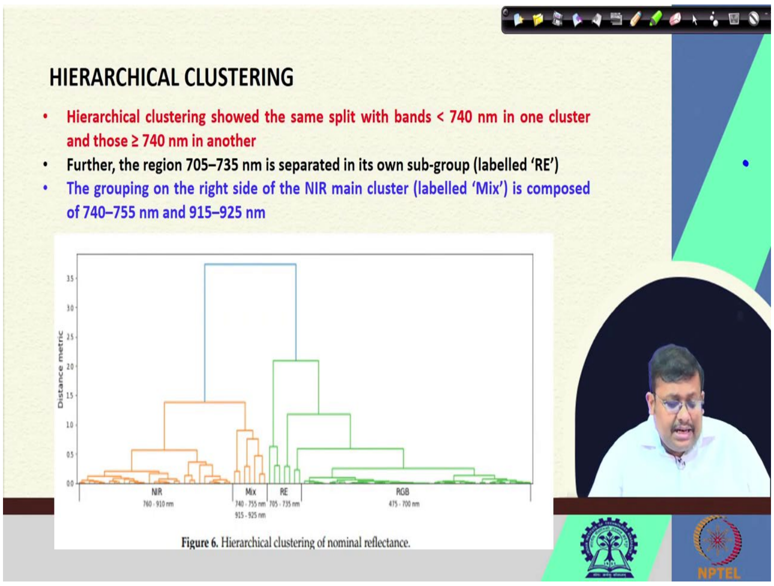


Also, hierarchical clustering, you can see this is a hierarchical clustering Dendrogram and we can see that hierarchical clustering shows the same split with bands less than, 740 nanometer, you can see here 740 nanometer, here one group and one cluster those for who greater than 740 nanometer. So, these cluster is based on the bands which 700 greater than 740 nanometer, these cluster is based on the bands with the greater than 700 less than 740 nanometer.

Further the region between 705 and 735, so, here these zone 705 to 735 nanometer is separated into two into in one subgroups labeled as RE. So, this is RE. So, you can see here these zone is separated in its own group that is called that is denoted by RE and the grouping on the right side of the NIR main cluster which is denoted by these Mix is composed of the 742 to 755 nanometer and also 915 to 925 nanometer.

So, we can clearly see the clustering of the wavelength based on certain characteristics and we can correlate it with the nitrogen content of the plant.

(Refer Slide Time: 31:17)



Now, so, guys, this is the reference which I am which I have used and I hope that you have gathered good knowledge. The discussion is not over yet, we are going to stop here and in the next lecture, we are going to start from here and we are going to see the results from random forest regression and how they were useful for identifying or to for predicting the nitrogen content of the cotton crop.

So, guys, I think that we have gathered some important insights from this lecture, we now know in application in crop where we can extract the hyperspectral data from the average of the pixels and how to extract the information from those spectral data by using different types of supervised and unsupervised clustering methods.

So, we are going to discuss more on this in our comings in our coming lecture. And please stay tuned. And we will finish the discussion. And also in this coming lecture, last lecture of this week, we are going to discuss and some other application of hyperspectral remote sensing for soil

And side by side, I will show you some other novel application of hyperspectral imaging for both crop as well as in soil. So, thank you for joining. And let us meet in our last lecture of week 9 to continue with this and to explore other ideas of hyperspectral remote sensing application in agricultural domain. Thank you guys.