Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 45
Hyperspectral Remote Sensing and ML Applications in Agriculture (Contd.)

(Refer Slide Time: 0:22)

Welcome friends to this fifth lecture of week 9 of NPTEL Online Certification Course of Machine Learning for Soil and Crop Management. In this week, we are discussing about and very important concept that is Hyperspectral Remote Sensing and Machine Learning Applications in Agriculture.

In our previous lectures, we have discussed about what is hyperspectral remote sensing, what is multispectral remote sensing and also we have discussed about the differences between hyperspectral remote sensing and multispectral remote sensing, we have seen the hyperspectral remote sensing spectrometers or different set in the in this satellite platform as well as some airborne hyperspectral sensors.

We have seen the different types of scanner configuration, we have seen the data cube and also we have seen the different types of data sources. In this mission as far as hyperspectral remote sensing is concerned and also we have seen how to handle the data using different types of approaches. In the last lecture, we have started discussing about one case study where hyperspectral remote sensing was applied along with the multispectral remote sensing for predicting the crop nitrogen content in cotton.

And we have seen that how they have extracted the data by two approaches, and then they after extracting the data, they have done some kind of clustering, using supervised and unsupervised method. And then after they selected those also they have calculated some vegetation, vegetation indices.

So, they have tried both density with scanning and density based clustering. And also they have used the principal component analysis. Apart from that they have used the clustering analysis and also they have used random forest regression. So, we will start from there. We will start from where we left in our last lecture. And we will try to complete it in this lecture.

Apart from that, we are also going to discuss some of the other hyperspectral remote sensing application in soil. And I would also like to show you hyperspectral imaging which is another emerging I would say emerging field of have hyperspectral analysis for crop and soil. And also I will show you some applications.

(Refer Slide Time: 3:31)
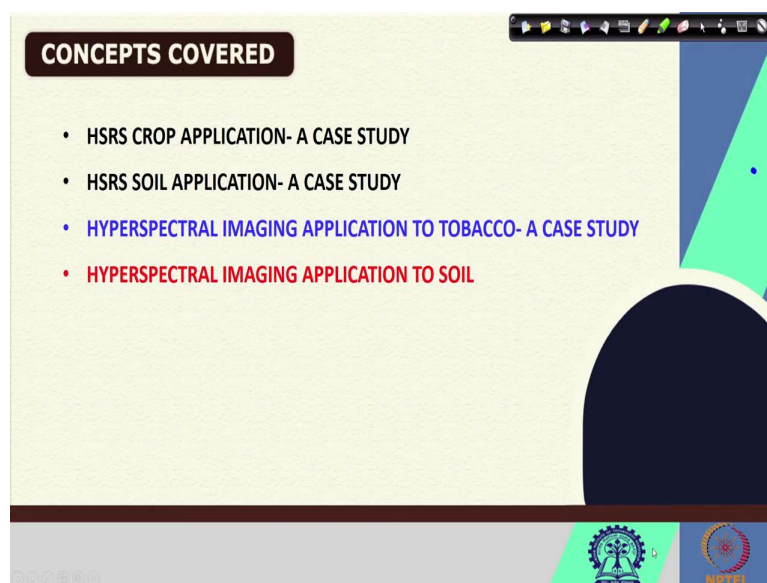
So, these are the main concepts which are we are going to discuss in this lecture. First of all hyperspectral remote sensing crop application, a case study and also the hyperspectral remote sensing soil application another case study and then we are going to also discuss hyperspectral imaging application to tobacco and also hyperspectral imaging application to soil. So, these are the 4 concepts which we are going to cover in this lecture.

(Refer Slide Time: 3:59)

These are the some of the keywords like hyperspectral, random forests, PLSR-VIP, tobacco, and then SVR. So, these are some of the keywords which we are going to discuss.

(Refer Slide Time: 4:11)

So, if you recall in our last lecture, we are discussing about this case study where these machine learning and optimized hyperspectral remote sensing was utilized for retrieving the cotton nitrogen status.

Figure 7. Random forest regression for all reflectance bands.
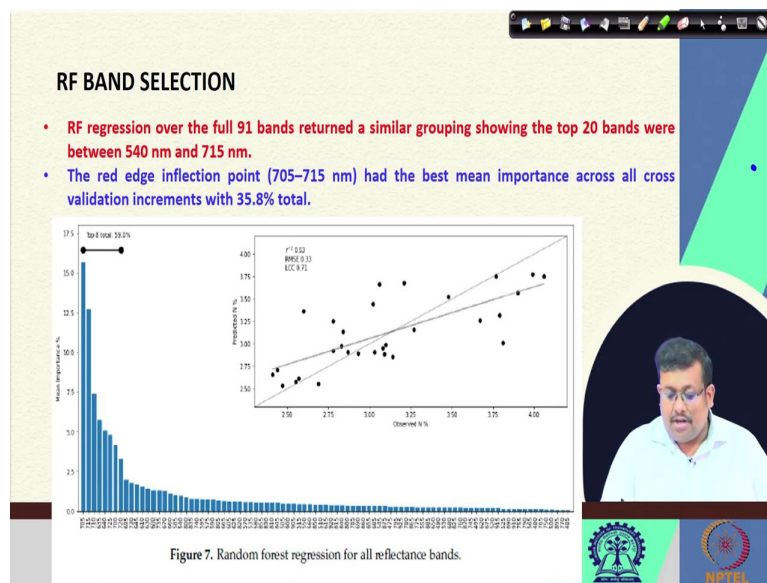
And, we have seen the results of hierarchical clustering as well as density based clustering. Now, while they have performed the random forest regression, what is random forest regression, which is a tree based method, ensemble tree based method.

Our, so these random forests regression they have tried over the full 91 bands, returned a and also returned a similar grouping showing the top 20 bands and showing the top 20 bands are centered on, they have selected these top 20 bands, these random forest regression which were range between 540 to 715 nanometers.

So, you can see here, these are the 91 bands, which were selected by the random forest regression. So, all the reflectance when 91 reflectance band and we can see that the first 20 bands ranges between 514 nanometer and 750 nanometer. Now, the red edge inflection point which also ranges between 705 to 750 nanometer had the based mean reflectance based mean importance across all cross validation increments with 35.8 percent.

So, we can see here that the top 20 bands were between 540 to 750 nanometer region and the red edge inflection point, which is inflection point which is ranging from 705 to 715 nanometer, they had the based mean importance across all cross validation increments with 35.8 percent total.

So, that means, again using these random forest, it this random forest regression was able to identify the importance of the red edge band for predicting the nitrogen concentration, this is the take home message from this graph. So, I told you what is red edge and why it is

important, why it is considered as an important indicator for the plant nutrition status. So, random forests again validated the fact that these red edge inflection point and the wavelengths are important for identifying the nitrogen status of the crop.

(Refer Slide Time: 7:21)

Figure 8. Random forest regression using only the top 10 reflectance bands.

Figure 7. Random forest regression for all reflectance bands.
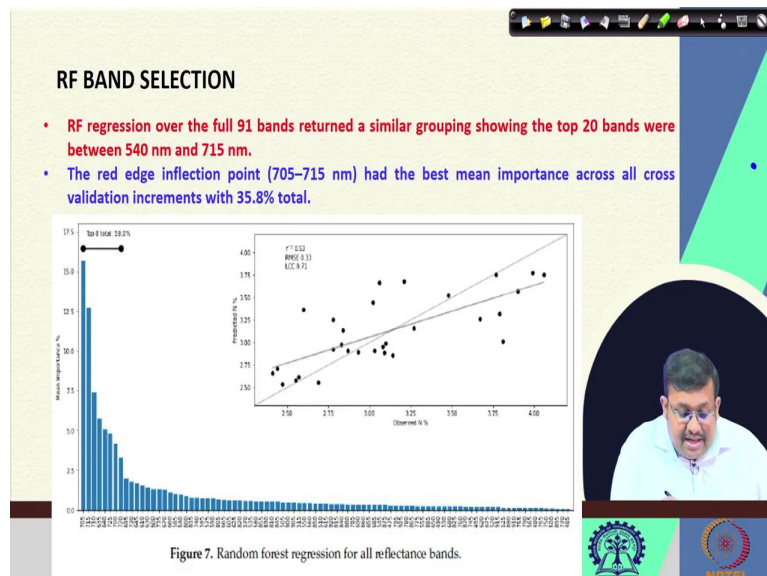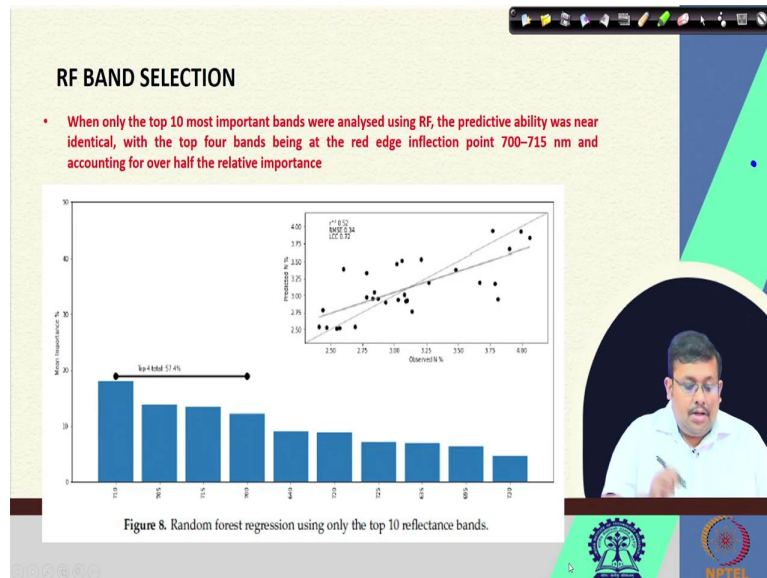
When only top 10 important bands were analyzed using the random forests, the predictive ability was near identical. So, if you go back you will see that R squared value is 0.53 here also, when we are using all the 91 bands, we are getting the R squared value of 0.53. However, when we are using only 10 important band first and important band, then only we also got the R square values of 4.52 which is almost identical with the top 4 bands.

So, these 1, 2, 3, 4 bands being at these red edge inflection point are 700 to 750 nanometers. So, you can see 710, 705, 715, 700. So, all these belong to these red edge inflection points. So, that again gives you an information how important these red edge band is for identifying the nutrient content.

(Refer Slide Time: 8:23)

So, based on all these information, they have calculated two novel vegetation indices. One is for the original reflectance or nominal reflectance and the second one is for that derivative spectrum. So, the first one, so, the novel these vegetation indices focuses on the inflection point from the absorption well in the red region to the reflection peak in the NIR and in case of original reflectance it is named as inflection point ratio vegetation index or IPRVI.

So, they have calculated two vegetation indices, one is for the original reflectance, second is for the derivative reflectance. So, for the first one they have calculated these inflection point ratio vegetation indices using this formula, this is the reflectance for 695 nanometer, reflectance for 700 nanometers by a reflectance of 705 nanometer.

So, this is called IPRVI and for the derivative spectra, they have calculated these derivative inflection point ratio vegetation indices or deep RVI. So, here this is the first derivative of reflectance values at 745 nanometer. This the first derivative of the reflectance values of the 645 nanometer, sorry and this is the first derivative of the reflectance values for 685 nanometer.

And subsequently once they have calculated these vegetation leases they have utilized these for calculating these for producing these map of nitrogen concentration, using these novel hyperspectral vegetation indices. So, what we have seen in this research? We have seen this research that they have taken the hyperspectral images after they have taken the hyperspectral images, they have extracted those images, the spectral information they have extracted after extracting the spectral information, they have calculated, they have done some spectral preprocessing like first derivative and smoothing.

After the spectral preprocessing they have they did some kind of clustering analysis and also density based clustering analysis to identify the regions of interest and they also tried to use the random forest regression to identify the important spectral bands, which are responsible for nitrogen estimation and then they have calculated the novel vegetation indices based on those important points. So, this is how this we can see that, how this hyperspectral data can be utilized in a better way for identification of plant nitrogen.

(Refer Slide Time: 11:42)

SENTINEL PREDICTION RESULTS

Band	GSD (m)	RMSE	R^2	LCC	Equation
Green (B3 543–578 nm)	10	0.187	0.852	0.921	$y = -163.972 x + 13.22$
Green (B3 543–578 nm)	5	0.204	0.823	0.904	$y = -161.298x + 13.001$
CCCI	5	0.208	0.816	0.900	$y = 29.96x - 19.133$
REP1 (B5 698–713 nm)	5	0.225	0.785	0.882	$y = -114.364x + 15.364$
TCARI OSAVI	5	0.236	0.762	0.868	$y = -44.607x + 10.538$
mND705	5	0.242	0.750	0.861	$y = 27.209x - 14.748$
NDRE	5	0.243	0.748	0.860	$y = 23.798x - 12.602$
VOG1	5	0.260	0.713	0.838	$y = 2.516x - 6.481$
REP1 (B5 698–713 nm)	20	0.271	0.688	0.818	$y = -110.096x + 14.984$
NIR (B8 785–899 nm)	5	0.355	0.462	0.642	$y = 19.512 - 7.12$
REP3 (B7 773–793 nm)	5	0.391	0.348	0.536	$y = 22.854 - 8.588$
NDVI	5	0.395	0.337	0.531	$y = 36.428x - 29.282$
Red (B4 650–680 nm)	5	0.442	0.169	0.328	$y = -152.78x + 7.798$
REP2 (B6 733–748 nm)	5	0.486	0.000	0.085	$y = 22.348 - 6.007$
Blue (B2 458–523 nm)	5	0.488	0.000	0.064	$y = -314.119x + 12.034$

* GSD: Ground Sampling Distance or spatial resolution.

Or plant nutritional, status at the same time you can see here this table shows the Sentinel prediction results using the different vegetation indices, which they have calculated. And so, the using different bands and vegetation indices, how the Sentinel was able to predict the nitrogen content you can also see from here. So, this was a very good study which shows the application of both hyperspectral remote sensing and multispectral remote sensing for plant nutritional status identification.
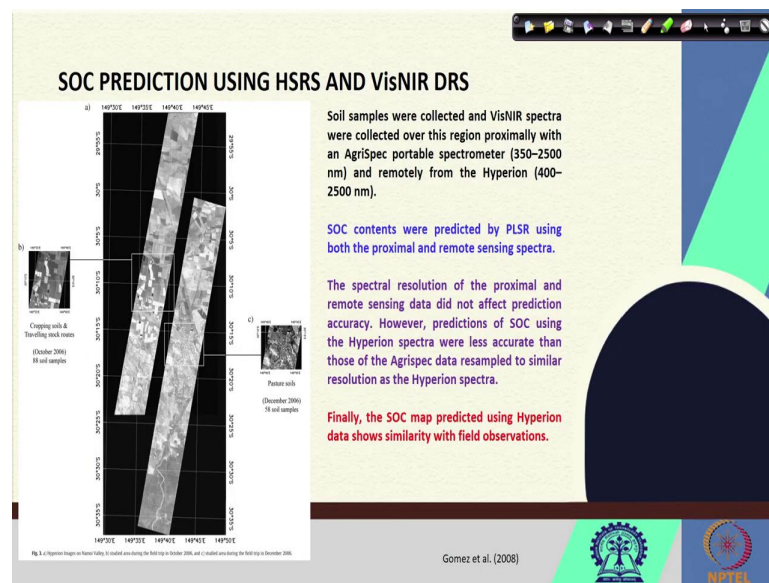
So, guys, let us also discuss another application where these hyperspectral remote sensing and visible near infrared diffuse reflectance spectroscopy. It is a point hyperspectral sensor, we have discussed this in our week 5 in details. So, hyperspectral remote sensing gives us the special information along with the spectral information.

However, the VisNIR DRS is also a hyperspectral instrument, but it is it gives us only the spectral information not the special information. Now, these in this research which is published in 2008 Gomez et al they have used soil samples and they collected the soil samples and they have collected the VisNIR spectra and from 350 to 2500 nanometer and also the Hyperion data was collected.

So, you can see this is in Hyperion data, Hyperion images for the study area. So, they have collected the soil spectra using a point portable spectrophotometer that is AgriSpec and the same time they have gathered they have collected the Hyperion data for that particular region also and they predicted the soil organic carbon by partial least squares regression using both the proximal and remote sensing spectra.

So, both of them are hyperspectral data, Hyperion is a remote sensing platform however, these VisNIR spectroscopy is a proximal sensor approach. So, using the both the spectra, they have tried to predict the soil organic carbon and the spectral resolution.
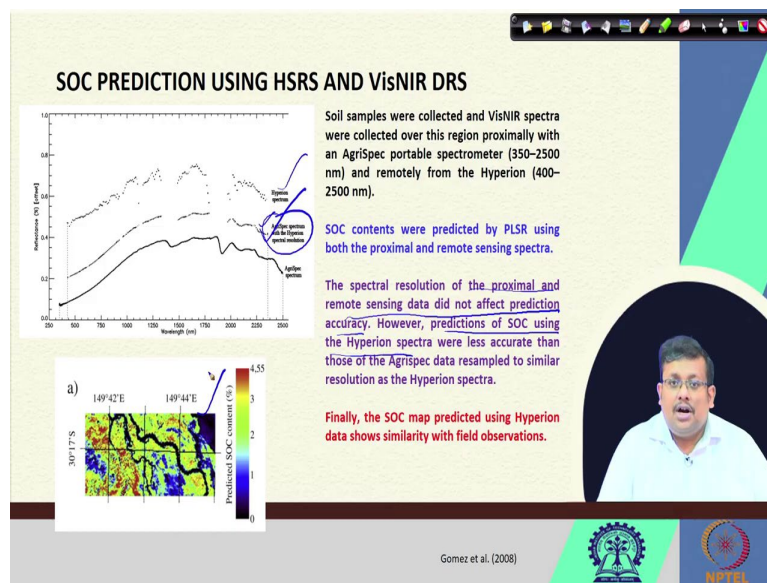
So, you can see in this picture, this is the original VisNIR the spectrometer spectroradiometer spectra and this is the Hyperion spectrum. And when the AgriSpec spectrum that is VisNIR DRS Spectrum was resampled as per the Hyperion spectral resolution, so hyper based on the Hyperion spectral resolution, if we resample the data of these AgriSpec spectrum we will get this type of curve.

So, in this three condition they have tried to predict the model using these three condition, they have utilized this AgriSpec spectrum, they have utilized this Hyperion spectrum and also they have utilized this AgriSpec spectrum with the Hyperion spectral resolution. So, in these three condition they tried to predict the soil organic carbon using the partial least squares regression.

And they have seen that the spectral resolution of the proximal and remote sensing data did not affect prediction accuracy. However, prediction of SOC using the Hyperion spectra was were less accurate than those of AgriSpec data that is sampled to similar resolution as the Hyperion spectra.

So, if we compare the Hyperion spectrum and if we compare the AgriSpec spectrum with the Hyperion spectral resolution will get better accuracy in this case, where these AgriSpec spectroradiometer spectra was resampled to with the with the Hyperion spectral resolution. And finally, they have predicted the soil organic carbon and they have mapped the soil

organic carbon using these Hyperion data and which shows the enough similarity with the field observation.

So, this also shows guys, how we can utilize these both satellite as well as air bone as well as point hyperspectral data using this point spectroradiometers to predict important parameters of the soil and ultimately we can upscale it to create the map of that particular property or the special variability map of that particular property for that region. So, now, you have a clear idea, how we can utilize this hyperspectral remote sensing data for characterizing the crop as well as soil, okay guys.

(Refer Slide Time: 17:16)

Now, we are going to discuss another emerging and important application of hyperspectral data that is hyperspectral imaging. Now, here we will see another case study of hyperspectral imaging for tobacco nicotine estimation, that tobacco is a very important crop and tobacco nicotine estimation is traditionally done in the laboratories using the cumbersome chemical method, time consuming chemical methods. So, here, we have tried to use this hyperspectral imaging for tobacco nicotine estimation.
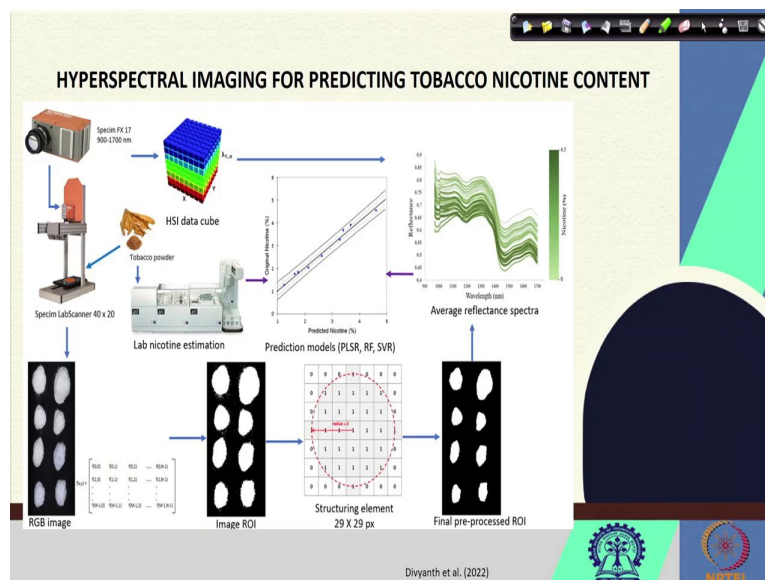
(Refer Slide Time: 18:05)

It involves different types of chemical reaction in the process, color development and all these things. So, our group tried to develop an alternative way to measure the nicotine content of the tobacco and which will have a huge application in the tobacco industry.

(Refer Slide Time: 18:29)

So, this is the flowchart of the, how we took the hyperspectral images and how we process the images for tobacco nicotine content. So initially, the tobacco was powdered and sieved. And we have used 57 different types of tobacco samples, we scan them using this spacing FX 17 hyperspectral camera, which is a spectral range from 900 to 700 nanometer and we took

the scan using the spacing lab scanner 40 by 20 centimeter dimension, and we took the scan from upside. And after taking the scan, we did some kind of spectral preprocessing.

At the same time, we measured the nicotine contained of those samples using the standard laboratory estimate. And after we did some kind of spectral processing or image processing, we extracted the average reflectance spectra and then we tried to predict the original nicotine content or the laboratory nicotine content using those spectral information to develop a prediction model.

So first, we have developed, we collected this hyperspectral data cube and from this data cube, we have isolated the images and then from images, we have selected the region of interest or ROI. From these ROI, we did some processing to get the final image and then we have got the average reflectance spectra, and then we use to use this spectrum to predict the nicotine content.

(Refer Slide Time: 20:25)

So, the first what we did actually, so, first after taking the image using these hyperspectral sensor, these RGB image representation of the hyperspectral image was used to derive these region of interest. So, you can see here this is the region of interest, these RGB digital images was converted to greyscale image and subsequently binarized using the im2bw function to remove these background.

So, here, you can see here, these white patches are basically showing the these white patches and these black patches are showing the region of interest and that discarded background. So, these white patches are of importance to us, but these black patches are not interest are not our target.

So, after we selected these ROI or region of interest, these ROI was eroded with a disk shaped structuring element with a neighborhood of 29 by 29 pixels to restrict the contribution of the spectral information of a neighboring background pixel to this process. So, after we selected these ROI, we did the final preprocessing using the structuring element to get the final processed image with a well-defined ROI.

(Refer Slide Time: 21:56)

And this these ROI selection was facilitate the elimination of the interference of non-representative pixel in the average spectrum. And the mean spectrum data was collected for each sample and calculated by considering the spectra of the corresponding ROI. So, from these each ROI, we have calculated their mean reflectance spectrum and we utilize these other means reflectance spectrum or mean reflectance spectra for all the 557 samples and they were used for predicting the nicotine content.

So, what we did again? We took the scan, we develop these data cube from this data cube we took the RGB representation of the HSI, then we did the binary segmentation to get the ROI and then from these ROI, we use the structuring and structuring element to get the final preprocess ROI.

(Refer Slide Time: 22:52)
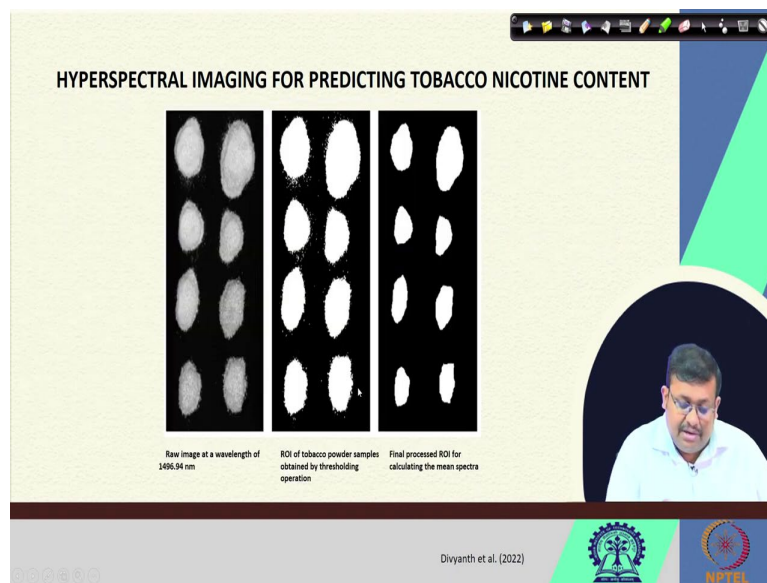
So, you can see here this is the raw image at a wavelength of 1496 nanometer. This is the ROI of tobacco powder samples obtained by thresholding operation and then we have done this is the final process ROI for calculating the mean spectra. So, this from this final process ROI we have calculated the mean spectrum to further for further machine learning based prediction model.

(Refer Slide Time: 23:19)
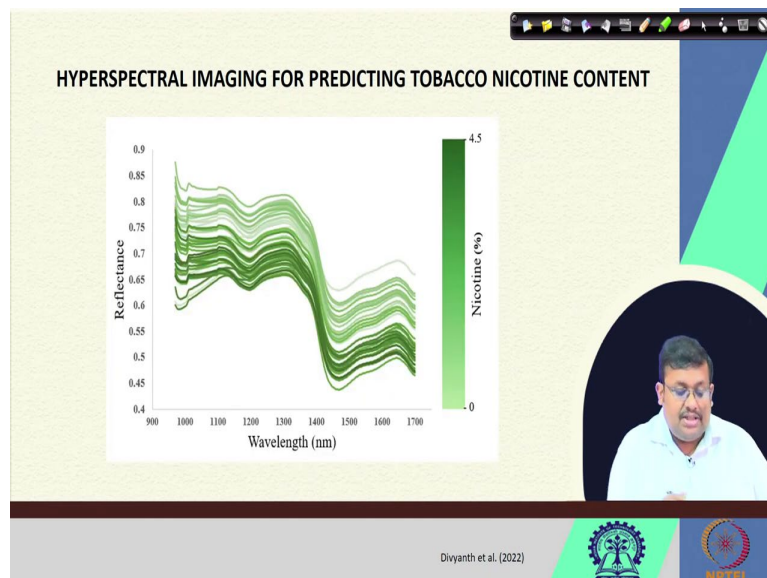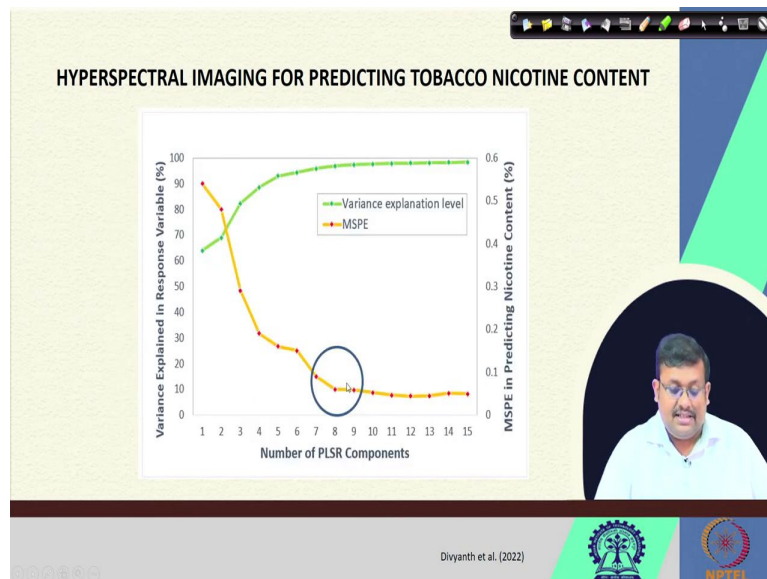
So, this is the hyperspectral these are main spectra from each of these ROI or each of these samples, one thing is clear that with the increase in nicotine content, we can see that reflectance is getting decreased. So, this is one pattern we have seen with the increase in nutrient content, we are getting reduction in their reflectance pattern, but more or less their reflectance shapes are same.

And then we have tried 3 different regression model, one is actually 4 different regression model, one is partially squares regression, second is support vector regression, third is random forest regression. And the fourth one is partial least squares regression with VIP selected wavelengths, we are going to discuss this later.

So, once we deal, we have seen that the for the partial least squares regression we have selected, we have plotted the 50 Latin factors and we have selected we have identified that the optimum number of Latin factors first will discuss Latin factor was 8. So, 8 Latin factor was used for partially squares regression model.

(Refer Slide Time: 24:27)
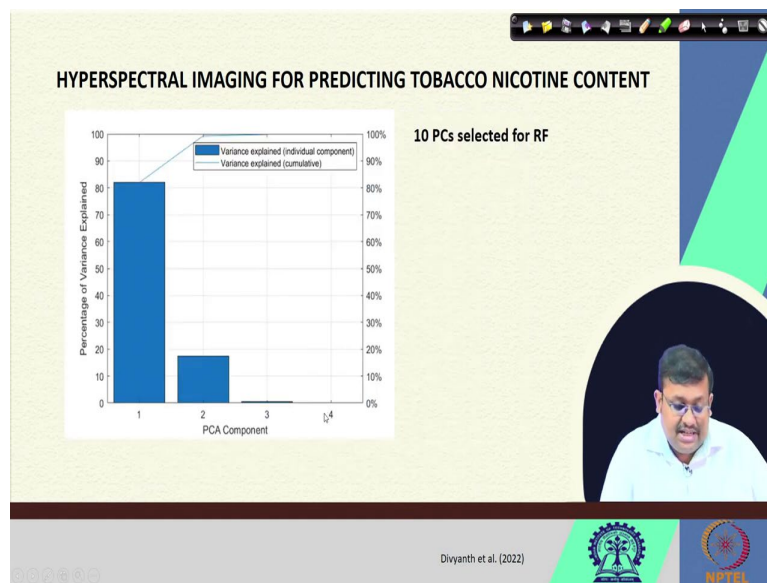
Then we did the principal component analysis for selecting the important principal components and then we included that for random forest regression. So, we have, using the principal component analysis you can see this a PCs scree plot and also we are selected 10 principal component analysis for subsequent random forest regression.

(Refer Slide Time: 24:49)
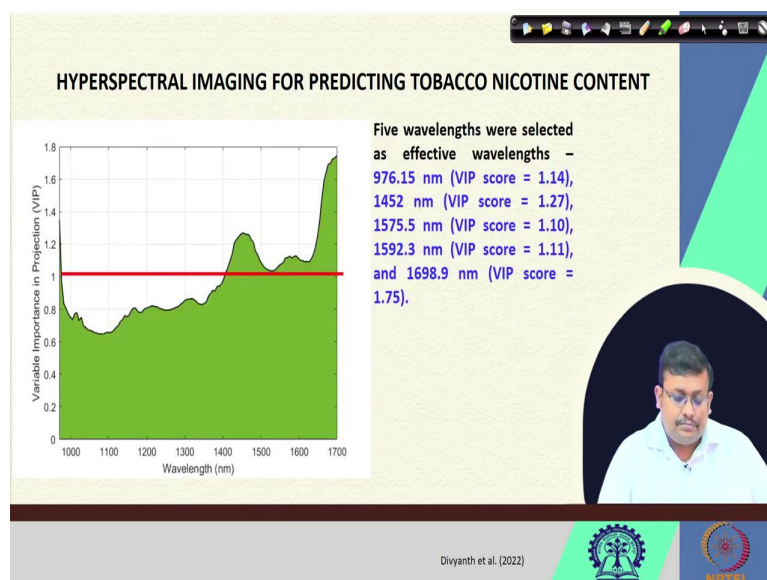
And also, we have also identified the important as effective wavelengths by based on their VIP score, VIP stands for variable importance in projection and these variable importance in projection was selected based on the VIP value of greater than 1.

So, after we select this VIP value of greater than 1, we have selected 5 important wavelengths that is 976 nanometer, 1452 nanometer, 1575 nanometer, 1592 nanometer, 1698 nanometer. So, why we go for the VIP method? There is when we take when you consider all the wavelengths, of course, our model becomes complex.

So, it is always desirable from the model parsimony point of view that, if you have two models side by side, you will always go and both these models are giving you the similar performance, you will always select the one which is giving you the which are having the lower complexity that means, if you have two models and both of them are giving you the similar accuracy, you will select the model with the less number of predictors that means, you always try to simplify your model for better interpretation.

So, here also instead of using all the wavelengths in the traditional PLSR model, we have tried to select those 5 wavelengths to the VIP method and then we combine them in the PLSR and try to compare their accuracy.

(Refer Slide Time: 26:31)

So, this is the model accuracy comparison, we can clearly see that the partial squares regression model, random forest regression models, support vector regression all of them are more or less satisfactory is in comparison. If we compare the test results of course, PLSR gives us the best result with R squared values of 0.93 in the test set. And in case of lever or not cross validation also we are getting the highest R squared for PLSR. But if we consider the PLSR-VIP also it is also giving us very close R square value.

So, given the less number of bands, that is only 5 bands in case of this VIP method and in case of traditional PLSR methods, we are using around 130 bands. So, compare one model with 130 bands and another model with only 5 bands and we are getting almost similar accuracy. Then based on this discussion, we have seen that we this we will select this PLSR-VIP method for future characterization of nicotine content using the partial least squares regression model and hyperspectral imaging.

So, this is one very good example this result is recently published in Journal of bio-systems engineering and it shows a novel application of hyperspectral imaging for predicting that tobacco nicotine content.

(Refer Slide Time: 28:05)
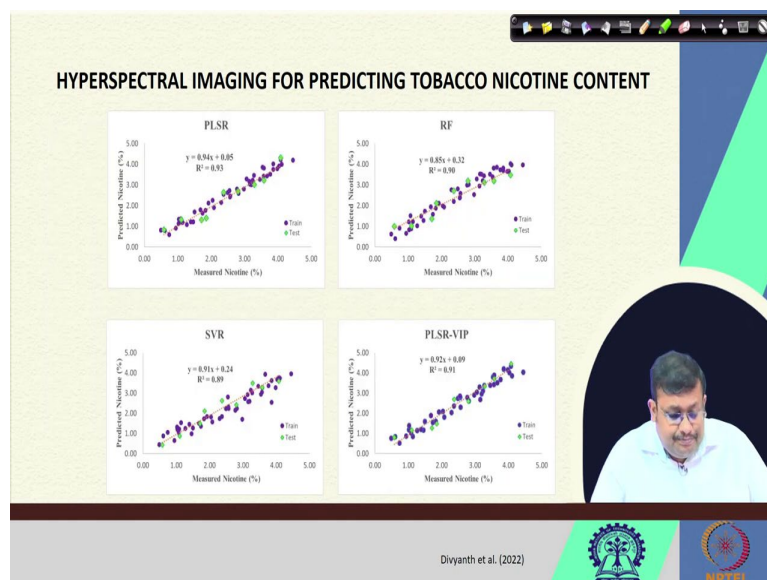
So, these are the predicted model predicted values, you can see that PLSR-VIP method and full PLSR method, we are getting almost close accuracy in terms of FR square value. There are certain other approaches also where hyperspectral imaging was used for mapping the soil nitrogen, one of these research was published by Xu et al in 2021 where they have used they have collected some soil profiles.

And they have used these hyperspectral imager scanning system that is Imspector V 10 E to collect these HSI data cube. After they collect these HSI data cube and the simultaneously they made some laboratory analysis. After they did the laboratory analyses of different parameter like total nitrogen, available nitrogen, ammonium nitrogen, nitrate nitrogen and a microbial biomass nitrogen.

So, they got these reference results and after extracting the spectra from the region of interest, they combine these result y and x spectra by these target parameters and these ROI spectra by using different types of machine Latin machine learning techniques like partially squares regression, artificial neural regression, Cubist, extreme learning machine, KNN, SVMR, XGBoost.

So, using different types of calibration validation and then the identified the best model and then they have produced the nitrogen maps. So, you can see this is another novel application where hyperspectral imaging has been utilized for mapping from the soil nitrogen content.

(Refer Slide Time: 29:57)

And other research was published Jia et al in 2017, where they have used these hyperspectral images of the soil samples and from three different types of soil and after they collected the soil they did some kind of image processing like once they collected the image they have done such image correction and then identifying the region of interest, then they have done some successive projection algorithm and after that they have classified the images using the support vector machine and then they have finally predict develop the prediction model using the partial least squares regression.

So, using the partial least squares regression they try to predict the different types of soil properties and also you can see here these are the reflectance average spectrum for each of the soil type. There are three different soil type, paddy soil, red soil and seashore saline soil for each of the soil type, they have got these average spectrum ranging between 975 to 1645 nanometer, and you can clearly see the difference of their spectral pattern and also they have done some grouping of 183 soil sample based on the Fisher LDA we have already discussed what is linear discriminant analysis using the first 4 principal components of full spectrum.

And they have utilizing those utilizing these different types of model, they have classified the different types of soil using they do classified different types of soil and you can see for three different types of soil and you can see their classification accuracy, which is quite high. Using different types of input variables like full spectrum, effective wavelengths, textural features and effective wavelengths and textual features.

So, they combined different types of input variables in different fashions, and then they used the support vector machine model to classify the soil types, three different soil types, red soil, seashore saline soil, as well as paddy soil and they got the high classification accuracy and using the partially squares regression they tried to predict different soil properties also.

So, this is how the guys, these are some of the examples where hyperspectral imaging could be an important tool for the upcoming scientist in this domain of crop as well as soil who can explore these ideas for further cost effective as well as rapid identification of both plant and crop as well as soil properties.

You can utilize these features, which you can extract from the hyperspectral images for predicting different parameters of crop as well as soil and also you can do some classification for different types of crop properties as well as you can also classify different types of crop based on certain properties.

So, guys, I hope that you have you have gathered some knowledge's important knowledge's or insights of hyperspectral remote sensing and how nowadays people are using these four different types of soil and crop application. Of course, our time is limited, and we cannot go beyond that, but there are plenty of information of hyperspectral remote sensing which are available in the internet.

So, please go ahead and explore those ideas and see how you can utilize those novel ideas into your own research. These are the references which I have used, please explore more papers there are numerous papers on application of hyperspectral remote sensing in agriculture, both soil and crop, please go ahead and get more and more confidence because the tools which are needed for understanding these we have already more or less discussed. And so, you will get more and more confidence while exploring these new new researches, research ideas and you can develop your own research ideas also.

(Refer Slide Time: 34:43)
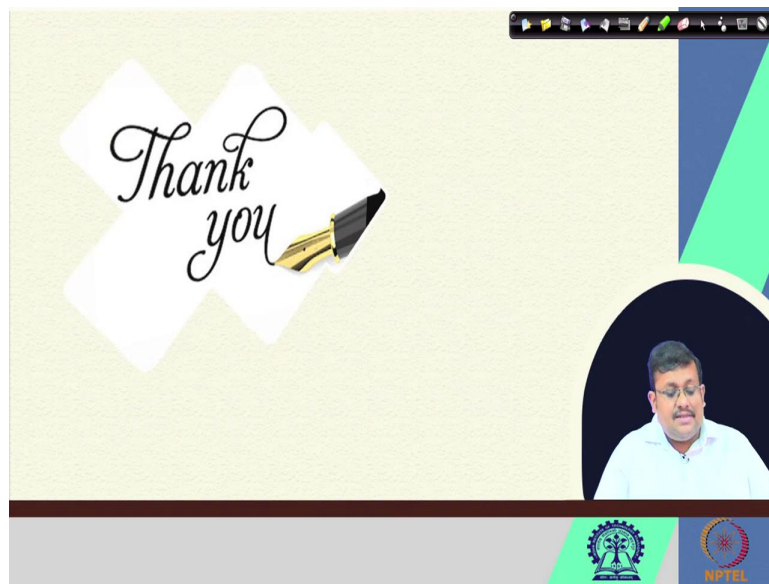
So, thank you guys for first for being here. And let us wrap up this week 9 and from next week onwards we will be starting the discussion on digital soil mapping. Thank you.