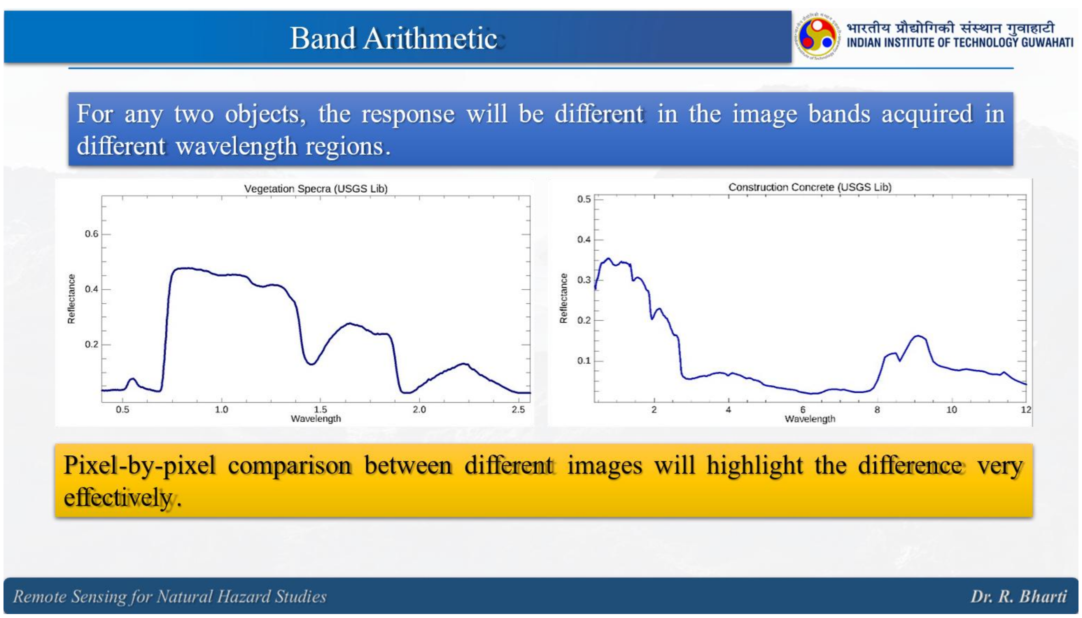
REMOTE SENSING FOR NATURAL HAZARD STUDIES

Course Instructor: Dr. Rishikesh Bharti
Associate Professor
Department of Civil Engineering
Indian Institute of Technology Guwahati
North Guwahati, Guwahati, Assam 781 039, India
e-mail: rbharti@iitg.ac.in
Website: https://fac.iitg.ac.in/rbharti/

Lec 9a: Remote Sensing Data Analysis-II Part A

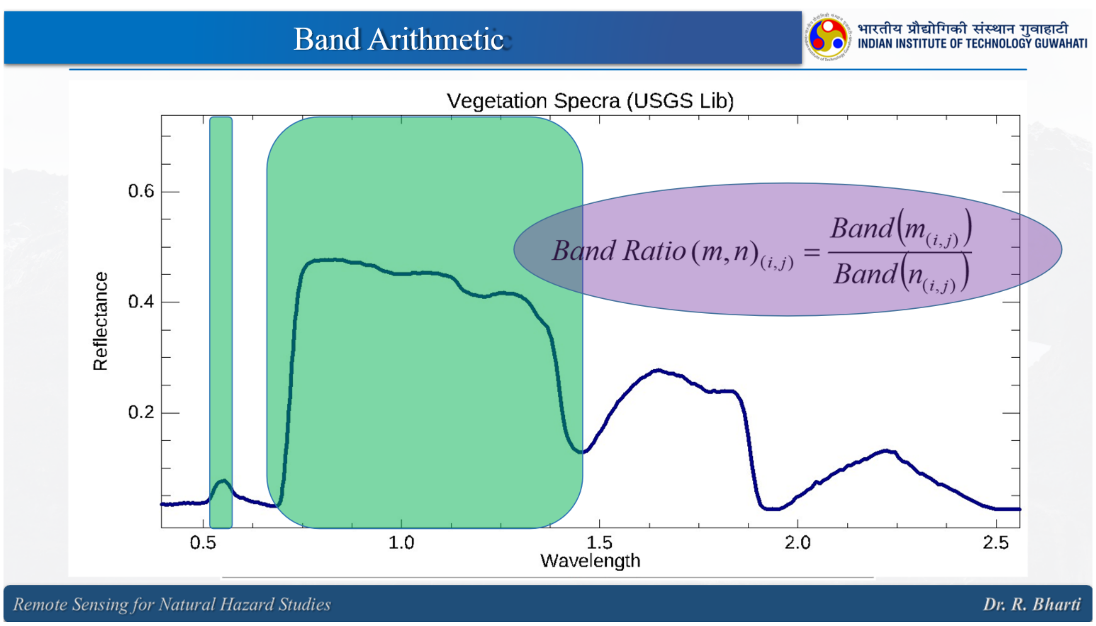
Hello everyone, welcome to Lecture 9, which is on Remote Sensing Data Analysis. So we will continue this. So, digital image processing means the images that we acquire from remote sensing sensors, which are basically reflected/emitted values, So, those are associated with the pixels you see here. So, these pixels are nothing but an array of digital numbers. So, when we look in the background, it is actually the DN values that are further processed into radiance, reflectance, or emittance. So, these digital numbers, how do we process them? How do we use them in our objective? So, we have to learn some basics, So, these are nothing but the processing of digital images.



So, when we talk about the band arithmetic, which here refers to the modification of an image for extracting particular features by transforming the values of each pixel using multiples of the band. So, here we are using the multiples of band and this information which is captured by your sensor, they are used together to enhance one particular object or target. So, the band arithmetic process results in a new image from two or more images to highlight a particular object or target. It operates on individual pixels, not on the image average, because if the image is, let us say, 1 kilometer by 1 kilometer and if we use the global operator. So, then what will happen to this entire 1 kilometer by 1 kilometer area if you have only 1 object of this size, this is not feasible or practical, So, that is why when we perform this band arithmetic, it is operated on individual pixels, not on the whole image.

So, there are a few questions that we need to answer; then your doubts will be clear, and you will understand more about digital image processing, especially when we talk about band arithmetic. So, what could be the possible benefits of band arithmetic in remote sensing? Why do we do that? So, here the answer is to highlight a particular object or target, what is a band ratio? So, the band ratio is like other operations such as difference and addition. Ratio is also one of the options that we can perform, So, if you have two bands, Of the same area captured by the same sensor in different wavelengths, So, if this is the target and you know the spectral response of this target, then you can use the lambda 1 and lambda 2 wavelength images together; if this is A and this is B, it can be A = B, A - B, or A / B, So, anything is possible.

So, this is simple operation, but very powerful; it will give you more information about the target. Then the next question is how we can perform a ratio over a remotely sensitive image. So, pixel by pixel, because here we are not considering the whole image, So, it has to be pixel by pixel, and how many bands will be required for the band ratio? So, here is a minimum of 2, because you need something to do. A+B, A-B, or A/B. So, if you have more, you can have A+B+C, or you can have A+B-C, or A-B+C. This type of arithmetic is also possible. You can also have (a plus b) divided by (a minus b) to normalize it. Or you can have A+C/A-C. This kind of combination is allowed in band arithmetic when we use the remotely sensed images. So, here this will give you a better idea about why we actually go for the band arithmetic and how it is enhancing the object.



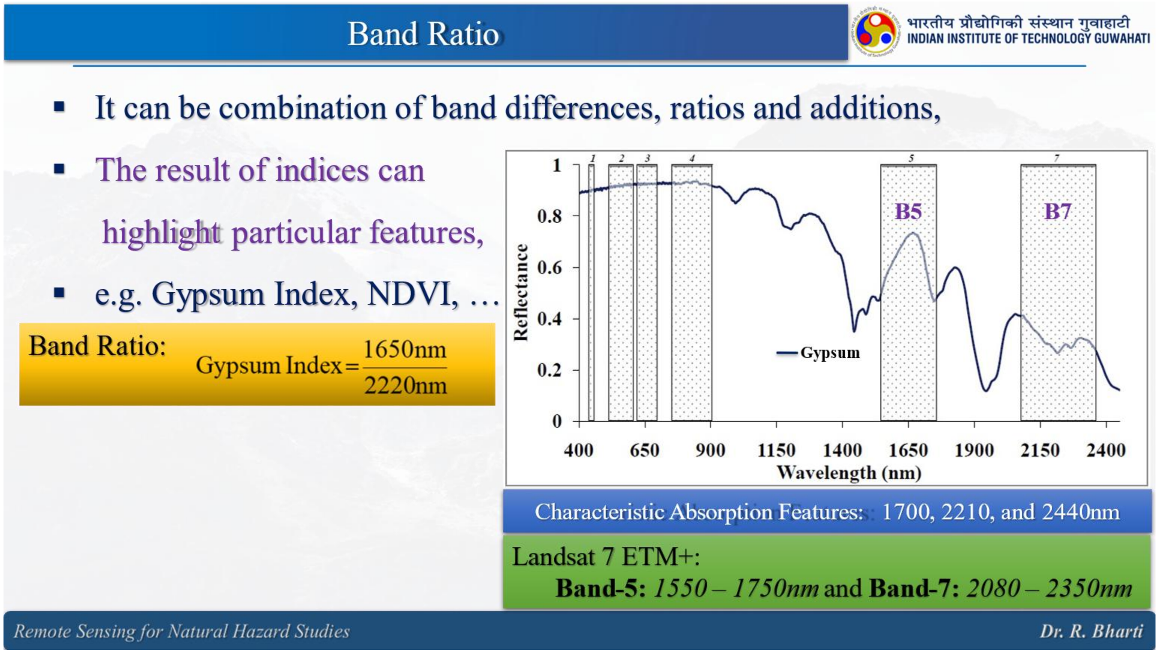
So, let us say, when we talk about two objects and essentially, they will have different spectral responses, you can see here that only then are we identifying them based on their characteristic absorption features, So, here this one is vegetation, and this one is concrete. , and here you can see the response of vegetation between 0.4 and 2.5 micrometers. Here, if you consider the same wavelength, it will be from here to here, So, this is for the concrete.

Now, if you compare these two from here—this much, compare only this much, then you will be able to easily notice that both have different spectral responses. So, here one absorption feature is coming; one is here, one is here, another one is here. Here you have 1, 2, and 3, which are very characteristic. Now, when we talk about the band arithmetic, we also try to keep the highly reflected areas in mind. So, if you consider this vegetation spectrum, this is one absorption feature; this is one peak, So, if we calculate this ratio, we will gain more insight into this vegetation because all other objects will have different spectral responses, if you compare them. So, if you do this ratio, only vegetation will be highlighted. So, that is the basis for this band ratio. So, a pixel-by-pixel comparison between different images will highlight the differences very effectively. So, that will highlight the difference between concrete and vegetation. Similarly, any other object or target can be considered in this context.

So, this is one way of extracting information from remotely sensed images that is band arithmetic. Now, arithmetic band operations like difference, ratio, and combinations are commonly used. I told you A+B, A-B, or A+B+C, or A+C/A-C, anything is possible here, depending upon the spectral response of the target. So, how do we design? A very common operation is the band ratio. So, here are band m_{ij} and band n_{ij}. So, here is the band number and this is the pixel location because it has to be pixel by pixel,

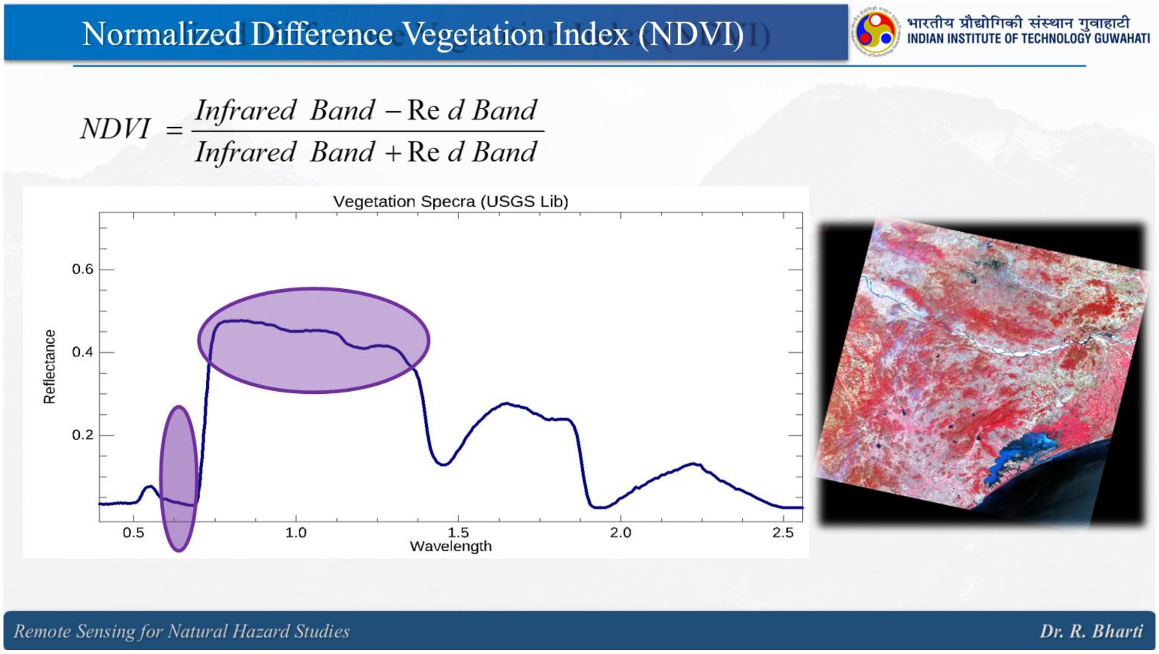
So, when we talk about this band ratio, we should have a minimum of 2 bands, band M and band N; that is the minimum requirement, and it should be pixel by pixel. Now, this graph can explain to you or give you more ideas about how we design particular indices. This graph will provide you with more information about how we design a particular band ratio, which is also known as indices. So, this is one absorption feature, this is another one, this is one, this is here, here. So, as I mentioned earlier, the peak will also be considered when designing the band ratios.

So, both minimum and maximum values will be considered to design or highlight a particular object. So, when we talk about this, let us see a vegetation spectrum. Because the absorption features are very characteristic, it will be easy to understand. Now here, when we talk about vegetation, one peak is here, another peak is here and the absorption features are here, here and here. So, if I have to design, I have to consider this location and that location and if I do this ratio, what will happen? We will get clear information about this vegetation. So, that is the basis of vegetation. band ratio design. So, the band ratio can be a combination of band differences, ratios, and additions, I told you A+B/A-B; that is also possible. So, here you will normalize the ratios.



When we talk about the gypsum spectra, the characteristic absorption features are located at 1700, 2210, and 2440 nm. So, if you see these absorption features, we can have the band ratio using these locations and one of the peaks reflected areas, So, when we consider this Landsat 7 ETM Plus data, band 5 has this wavelength range, while band 7 has this. So, band 7 is capturing this particular absorption feature, and band 5 is capturing this peak. So, if we consider these two, we can design this gypsum index. So, here it is: 1650 divided by 2220 nanometer.

So, if you use this, all the gypsum pixels will be highlighted, and the resultant image will have a ratio input of 2 bands. But the output will be a 1:1 ratio image, and in that image, the brighter pixels will highlight the presence of gypsum. I hope this is clear. We will take another example, which is a very famous index called NDVI, the normalized difference vegetation index. The formula is infrared band minus red band, infrared band plus red band, So, here these two are divided. So, this is a normalized index; this is a normalized ratio. So, here when we see this particular region, this is the highly reflective region. So, this is infrared. And when we talk about red, red has this absorption feature. So, when we have these combinations, we are basically trying to capture the minimum and maximum values produced by the target, which is vegetation.



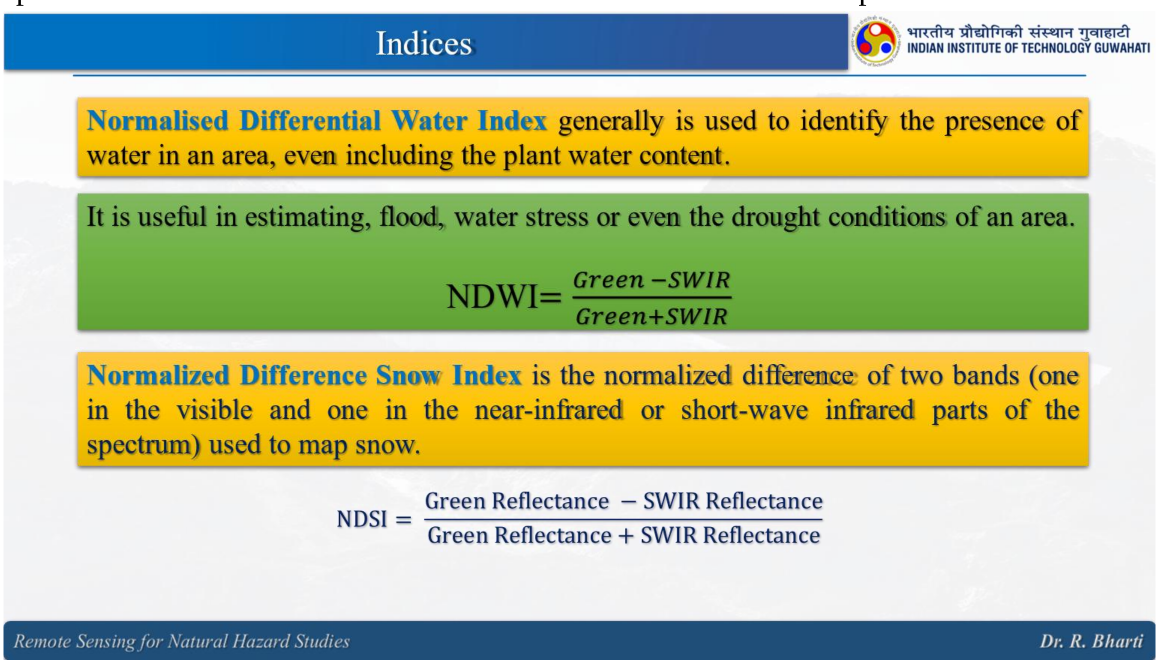
So, in this case, what will happen is that you will have the ratio, and then the ratio again; it will be a single band image, and in the ratio image, the brighter pixel will indicate vegetation, So, the image that results from this NDVI is a ratio image, and it will be grayscale because it is a single-band image. So, the brighter the white pixels, the brighter they are. So, those are vegetation pixels, we have one example, which is Landsat 7 data only. So, here you have the standard false color composite. So, you can easily see these are the vegetations, these are the vegetation pixels where you have the red color or dark red color, and those light red ones are sparse vegetation, but they are still vegetation, when we consider the inputs, the infrared and the red band, this is the infrared and this is the red band.

So, in the red band, basically, if you remember the vegetation spectra in the red band, we have the absorption feature, So, the darker pixel here will represent the vegetation, but some other targets may also have the absorption feature here in the red, which represents your water. When we consider the NIR, we expect the vegetation to have very high reflectance, So, in this case, the brighter pixels are basically representing your vegetation. So, we are going to use them in this equation, and then we will have the ratio image. Now, remember this infrared band can be band 3 of Landsat, band 3 of some other sensor, or band 2 of some other sensor. So, when you are performing this NDVI calculation, you should be very careful while assigning band numbers 2 or 3 as the infrared and red bands, as only the wavelength is mentioned here.

So, the wavelength can be mapped to your sensor's band designation and the wavelength it is using to capture the images. So, once we have this, we will be able to produce this NDVI image, and here, as I said, the brighter pixel will represent your vegetation. So, the NDVI results in high values where IR dominates over the red wavelength, which is the normalized difference vegetation index. Strongly highlights the green vegetation portion of the image. This happens because green vegetation has a very high reflectance value in the infrared region.

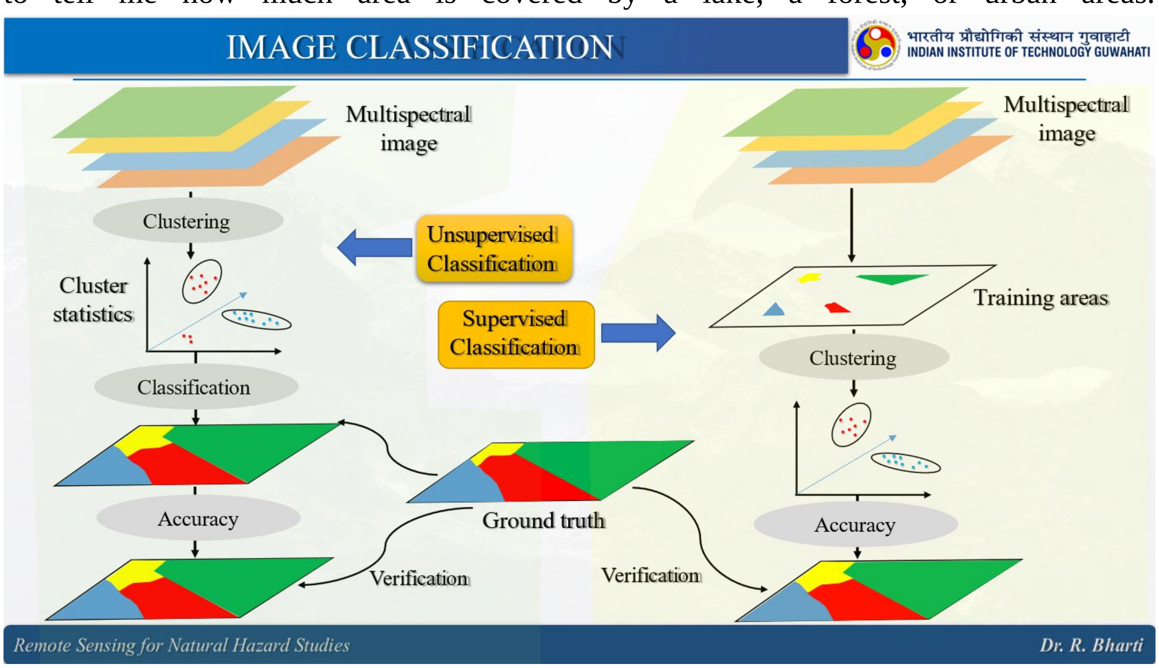
The value of NDVI ranges from -1 to +1. So, this is a ratio that ranges between -1 and +1. This is the correct value range. Widely used in agriculture, forestry, and biomass estimation applications, it can also be used to study crop health and growth. Why? Because this characteristic absorption feature, which we are observing in this red region, is present, and in the IR, we have this high reflectance. So, these represent the health of the vegetation. So, what happens when we have this kind of spectrum? This is also vegetation; I hope you are able to see this. This is also vegetation. So, shift in this slope from here to here indicates the health of the crop or the vegetation. So, this is widely used in crop health and growth.

This is called the red edge. So, you can refer to some of the standard literature, and you can use it in vegetation-related studies. Another example is the normalized difference water index, generally used to identify the presence of water in an area, including the plant water content. It is very useful in estimating flood, water stress, or even drought conditions of an area because this will capture or highlight the presence of water on the surface. Another example is the normalized difference snow index. It is the normalized difference of two bands, one in the visible and one in the near-infrared or shortwave-infrared parts of the spectrum used to map snow.



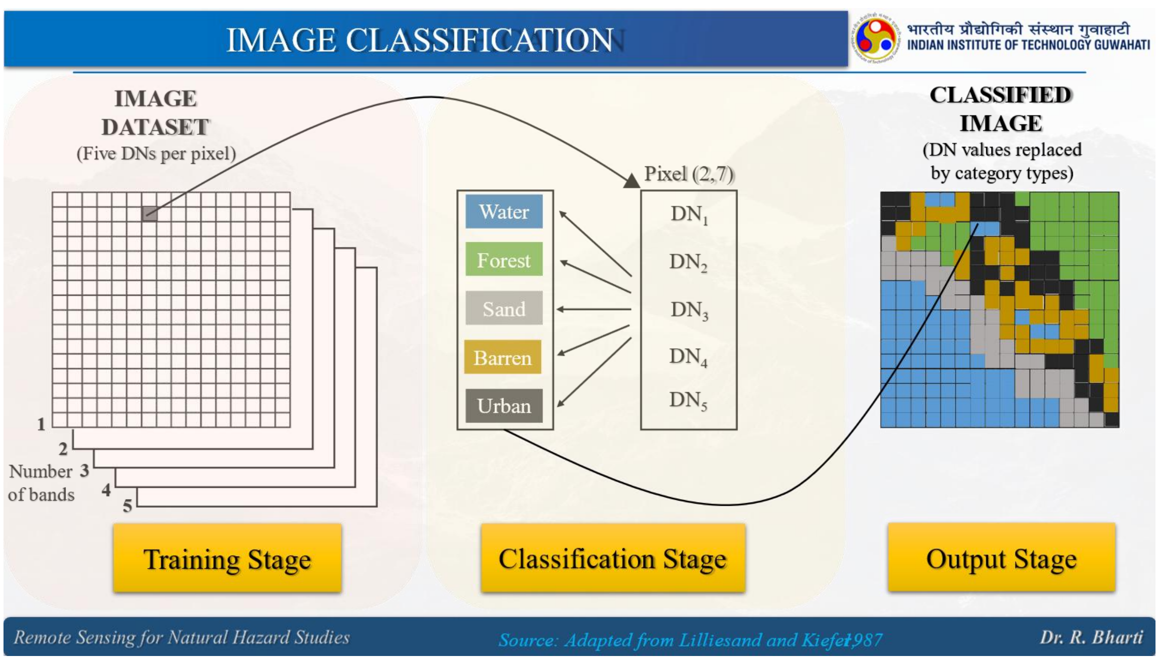
Here we are using the green and SWIR bands to calculate the NDSI. This is the normalized difference snow index. Now, one thing I would like to mention here is that when you are trying to capture images, we are also influenced by the atmosphere, So, when we are doing this band ratio, Band ratio: this is band A divided by band B, So, here when we consider this band A, it also has the atmospheric effects. This band B is also having an atmospheric effect because it was captured at the same time by the same sensor. So, when we are dividing them by each other, what is happening here is that we are removing the effects of the atmosphere. So, the band ratio of the resultant image will not have the impact of the atmosphere; it will get cancelled. , because both bands are captured by the same sensor at the same time. So, that is one of the advantages of band ratio that when we are trying to identify any object or material, we are also removing the effects of the atmosphere or any other errors that are present in or introduced into your data while acquiring it. Now we will move to image classification. This is one of the very important concepts which is used to generate the inputs for our analysis, So, when we talk about remote sensing, we try to generate the inputs for our decision support system; or, suppose we are going for landslide-related studies, then we need several parameters.

So, remote sensing can provide you with parameters that are captured from space, which were not possible or easy to capture in the field, So, these are the processes, or image classification is a process by which you can generate your inputs for your analysis, So, we will try to understand what image classification is. So, image classification is a process by which we assign pixels to a particular class or category, So, remember we have this image, and initially, when we captured it, we had this DN, then we converted it to radiance, and then we converted it to reflectance or emittance. So, let us say now we have the reflectance value, but this reflectance value is arranged in such a way that you are able to see the image. You can see the image here. But you will not be able to quantify immediately if I ask you to tell me how much area is covered by a lake, a forest, or urban areas.



So, what we do in image classification is try to assign a class or category to each pixel of my image. So, I should be able to tell what the area is immediately. , what is the area that belongs to the lake, forest, urban area, or water body that can be easily identified, and then we will be using that information in our further analysis, To assign each pixel of the image to a particular class or category, a relationship must be established between the data and the category. Now, let us say this is a particular image, From the field, we know that this particular pixel belongs to the forest, this belongs to water, and how do we know that? Because each of the pixels has x, y, and z values.

So, x and y are your latitude and longitude. And z is basically your DN value, reflectance, or radiance, whatever it is, So, when we have the location value, we can use the GPS to track that particular pixel where it is on the ground, and then we will see whether it is water, forest, or some other class or category. Once we have that, then we can collect such information in the field that all these pixels belong to water. So, if we know the latitude and longitude for water, these are the latitude and longitude that belong to water from the field, and from the image, we will also have the DN values or the reflectance values. So, let us say R1, R2, and R3 are all different reflectance values from the image, now we have a set of values representing water. Similarly, I will have a set of values for the forest.



Similarly, I will have a set of values for urban and any other class that I want to classify my image correctly. So, likewise, you have to collect, or you should have, the information from the field that these are the classes or categories that are available in my image, and then you should have the information about the corresponding reflectance or DN values from your image, So, this classification is divided into two major classes. So, the first one is the object-based; here it is based on the geometry, So, when we talk about geometry, that means your spatial resolution should be very high. In other terms, we call it fine resolution.

So, when we talk about the high spatial resolution, that means the pixel size is very, very small.

It should be in centimeters, millimeters, or maybe a few meters. So then we can do it with the geometry. So, we can use the geometry as one of the training parameters to classify my image. But most of the time, we have moderate or coarse resolution images, So, in that case, this pixel-based classification is done, and here we are using the DN values, not the geometry. Geometry is used in the case of high spatial resolution. So, pixel-based is good for medium to coarse resolution datasets. Now we will try to understand image classification. So, when we have the remotely sensed image, these images are captured in several bands, and they use different wavelengths to generate these images. So, each image has a spectral response of material or target at that particular wavelength. When we try to extract this information, we have the spectral curve if you remember the spectra we are able to generate.

So, when we talk about the panchromatic, we have single values; when we talk about the multispectral, we have a few bands; a minimum of 3 is required, but we can have up to tens of bands. So, those 10 values will be used to plot our graph for a particular pixel. When we have hyperspectral data, we have high spectral resolution. So, a greater number of points is there, and those are used to plot the graph, and they will have more characteristic absorption features of the target. So, utilizing this spectral information on how the material is behaving in different wavelengths, we will be able to classify our image.

So, this is pixel-based, where we are going to utilize the DN values; right now, we will see them one by one. There are two major classes: the first one is unsupervised, and the other one is supervised. So, let us understand this unsupervised first right. So, this multispectral image has 1, 2, 3, and 4 bands, and these are nothing but the array of digital numbers. So, we are having 4 data matrices, So, these 4 data matrices will be used to plot into N-dimensional space, and this ND is equal to your number of bands. So, when we try to plot this band 1 versus band 2, we will have a few clusters because similar objects will have similar digital numbers or reflectance or radiance or maybe emittance, and that will be used to generate this kind of cluster. So, when we have these clusters. When we have plotted band 1 versus band 2, band 2 versus band 3, band 3 versus band 4, and also in the N dimension. So, they are coming together.

So, all other bands are in the background. So, when we talk about this first cluster, it represents a particular class or category. So, this is, let us say, class 1; this is class 2; this could be class 3. So, all the pixels that are in this particular cluster will be named class 2. This is how we are going to provide a class or category to each of the pixels of this image. So, this is called classification. After classification, we also need to validate whether our classifiers have worked efficiently and produced the correct classes. So, what will we do? We will go to the field, collect the different class or category information in terms of their latitude and longitude, and then we will try to identify those pixels into which my classifier has actually classified them. So, if class 1 is mistakenly classified as class 2. So, according to this algorithm, it is giving class 2, but actually, in the field, it is class 1.

So, that is the error associated with my classification. So, what we do is use the ground information, and then we try to estimate the accuracy of this particular process or the image classification, and then we will say that my classification is 60% correct, 70% correct, or maybe 90% correct. Now, here was the case when we were talking about the unsupervised. When we did not have any input while classifying them. So, in unsupervised classification, we do not supervise this particular classifier. So, the classes were given by the system itself or by the method itself. But when we talk about the supervised classification, here "supervised" means supervision is required. You will train the algorithm based on the ground truth, and that ground truth will help the method or technique to train, and then it will give you the class or category. How? So, the training areas. So, the pixel information which is collected here while doing the ground truth. So, here it was used for the accuracy assessment, but in the supervised learning, you will be providing them to train the algorithm, and this particular set of values, let's say this is forest. So, this will overlap here, and then it will extract the corresponding DN values of all the bands, and then it will make a cluster and say this is forest. Similarly, for all other classes, a similar relationship will be developed between the field and the DN value or reflectance value. So, once you have this kind of clustering, you will have the forest, and this is your water body, and then you will have to assess the accuracy of this again. So, remember that the same points you have used here for training the data should not be used for the accuracy assessment that we will discuss in the further slides.

So, we have learned that we have unsupervised and then supervised classification techniques. The same thing is explained here. So, here we have this pixel, and for this pixel, we have all these values from different bands, and these pixel values will be used to classify this image.

I hope you understand this. So, here we have 5 bands. So, these five values are extracted, and this location was identified based on the field information, whether it is forest, water, or maybe an urban area. So, all these classes will have a different set of values from this pixel and from the ground information. So, this is training; here we will do the mapping. So, this is classification, this is the output, and then we will go for the accuracy estimation. So, we will continue this in the next part of this lecture. So, thank you very much for listening, and we will have Part 2 very soon.

Thank you.