Learning Analytics Tools

Professor Ramkumar Rajendran

Educational Technology

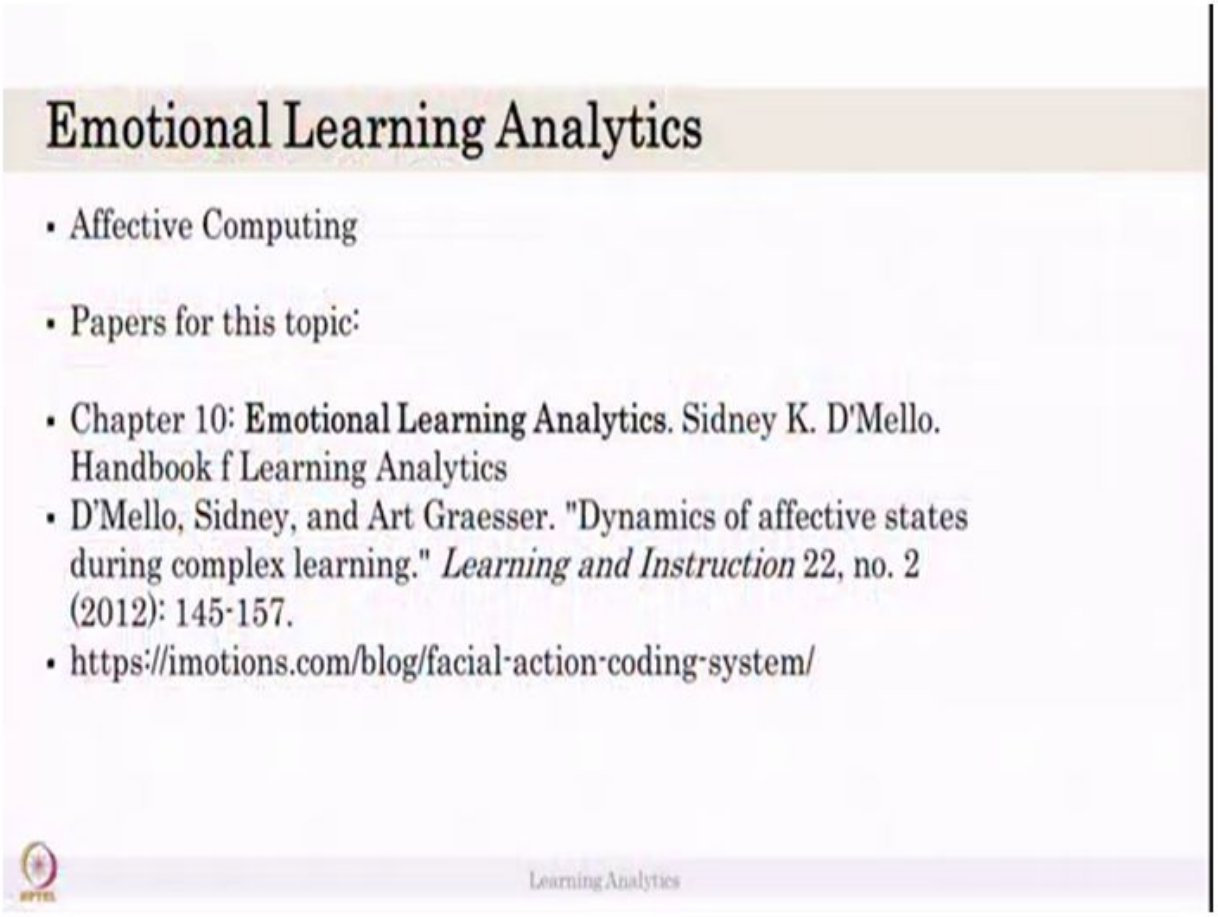
Indian Institute of Technology, Bombay

Lecture 11.3

Affective Computing - 2

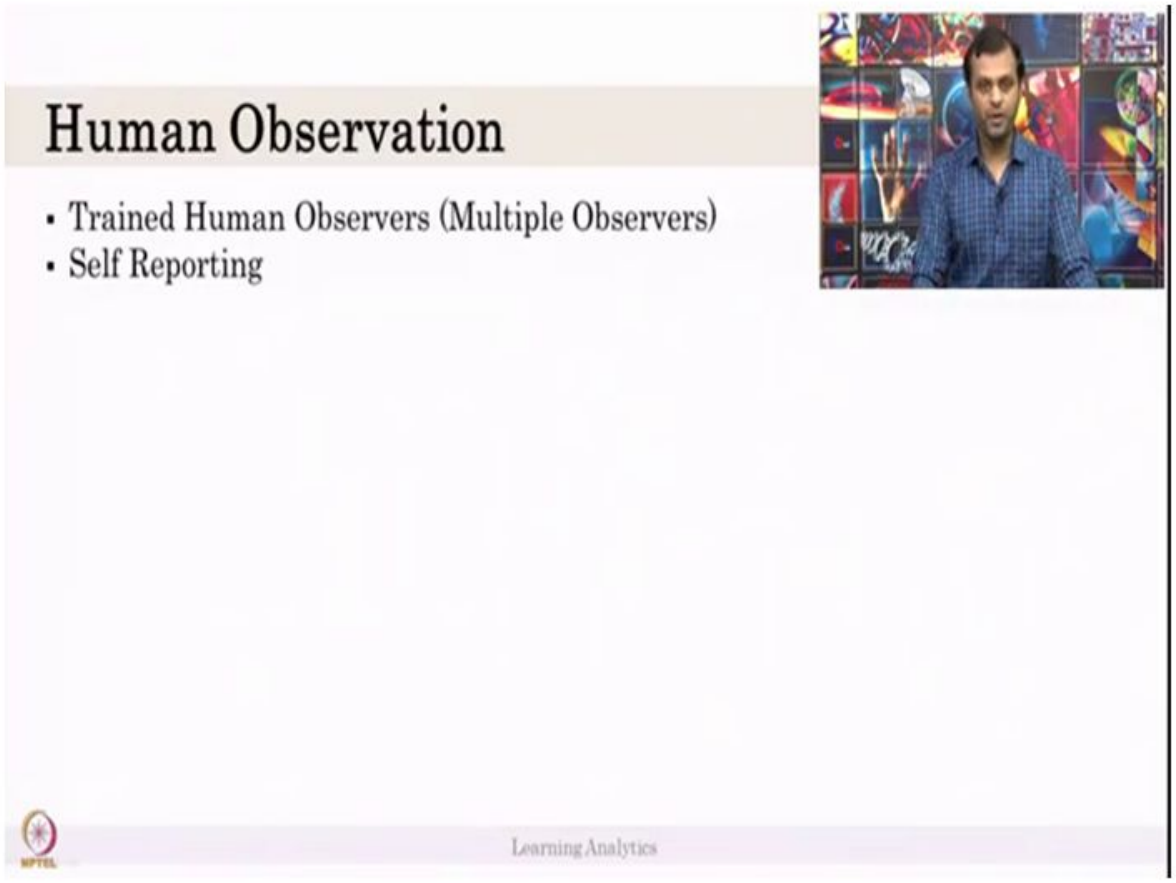
In this video, we will continue on affective computing and we will see how to detect affective states from facial expressions.

(Refer Slide Time: 00:31)



So again, the papers which were listed in these slides will be useful for you, go and check these papers to understand it even more detail.

(Refer Slide Time: 00:44)

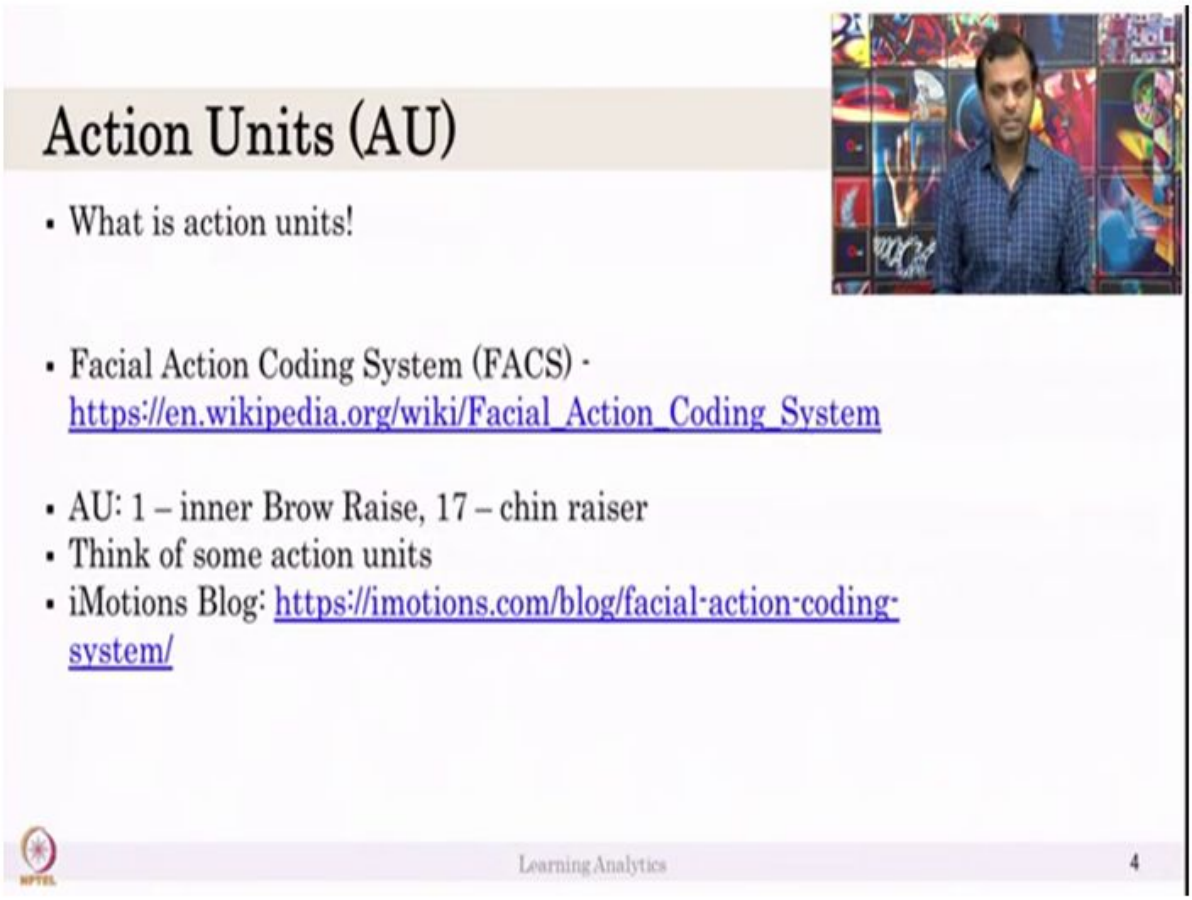


So, the one good way or the one method for detecting an emotion is - "human observations". Basically, you get trained on how to detect emotions and there are ways to do that. And it is better always you have multiple observers instead of you alone detecting students emotions. So, we talked about this in a Kappa score prediction for inter-rater reliability.

So, if you are doing the human observation you have to establish the inter-rater reliability using Kappa score Cohen's kappa. The other way is self-reporting, you know the person is self-reporting their emotions by observing their own facial expressions, how to do that we will

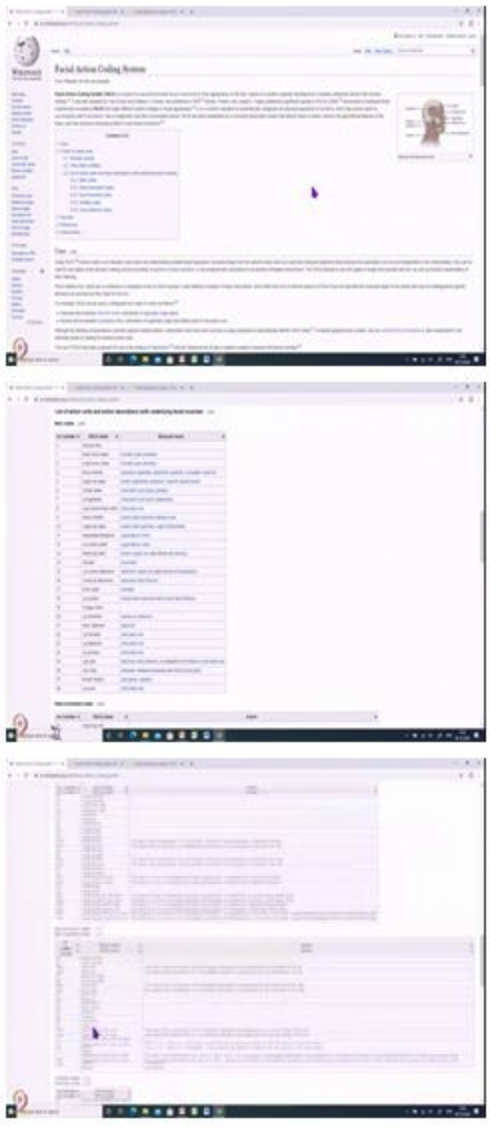
see. Or you can have some questionnaire coming into some, how are you feeling now, you know I am feeling bored or confused they can select their own options that are a way you can think of.

(Refer Slide Time: 01:54)



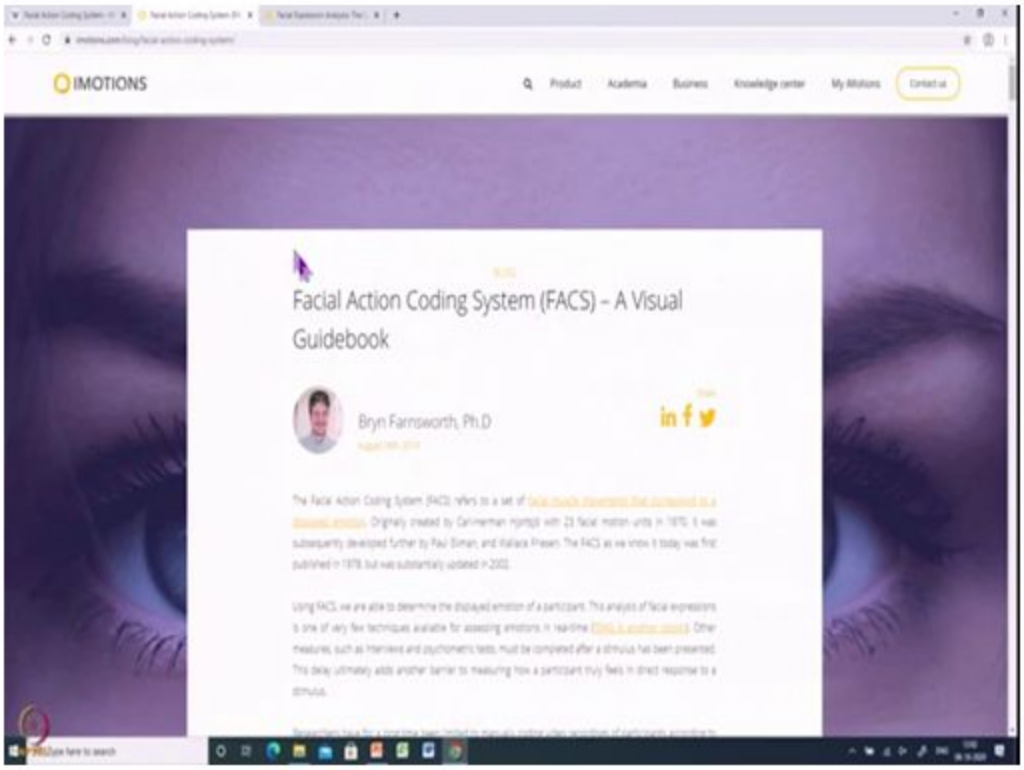
So, in order to detect facial expressions automatically. Paul Ekman's research suggested that their action you need to see in their face, there is a lot of action you need to see in the face such as inner brow raise, chin, lips rising, nose wrinkle, all these things are actions. So, the Paul Ekman listed down all this action units in a facial action coding system FACS, let us see that in detail.

(Refer Slide Time: 02:30)



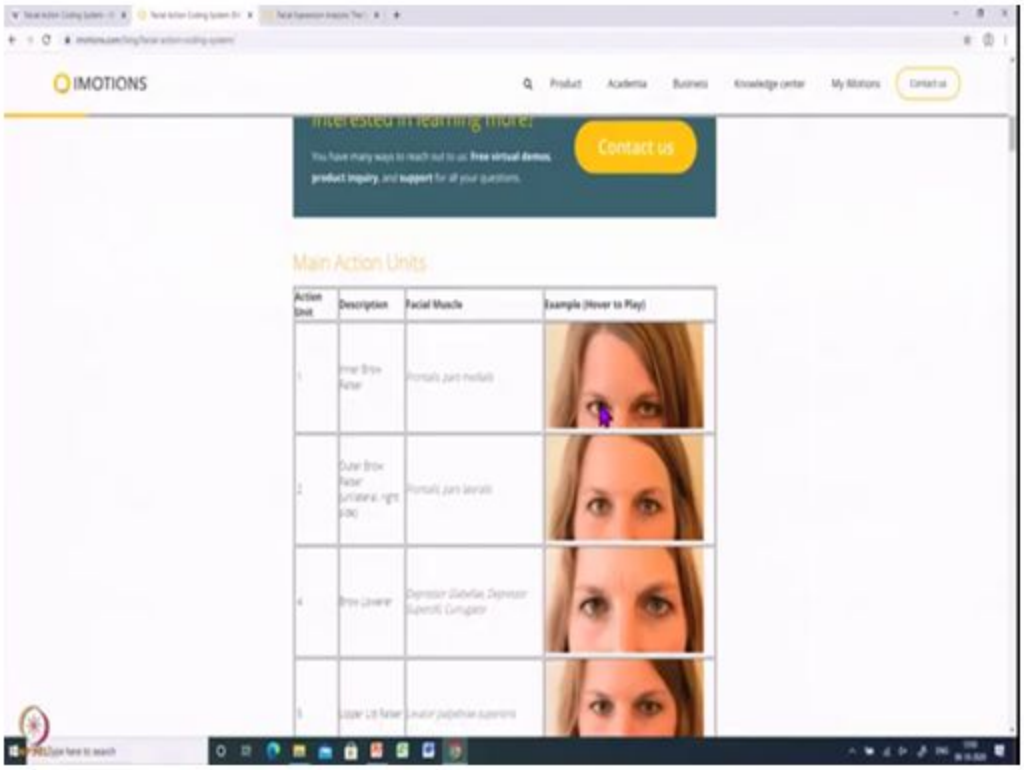
If you check that link in a facial action coding system, it is Wikipedia link you can go and check it. And you will see there are 28 AUMN codes and based on head movements, there are many codes and based on high moments there are many more codes in MSBT codes. So, what is this code, for example, inner brow raise or outer brow raise and this each and every action unit in your facial muscle is actually mapped.

(Refer Slide Time: 03:08)



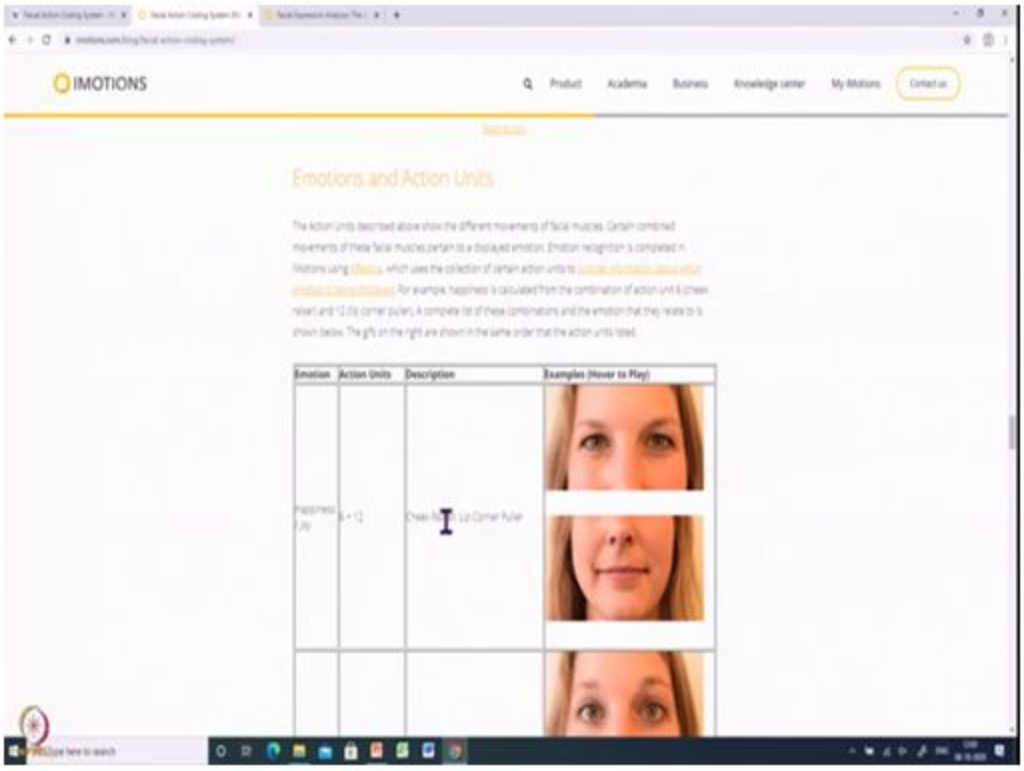
To know more about this you know these 28 codes, go to the link about facial action coding system, a visual guide book. A blog by a company called imotions, which is the pioneer in collecting data from different sensors like eye tracker, facial expressions, GSR, EG, collect all the data and the sink deploy a complete picture, so that is one thing so let us see.

(Refer Slide Time: 3:28)



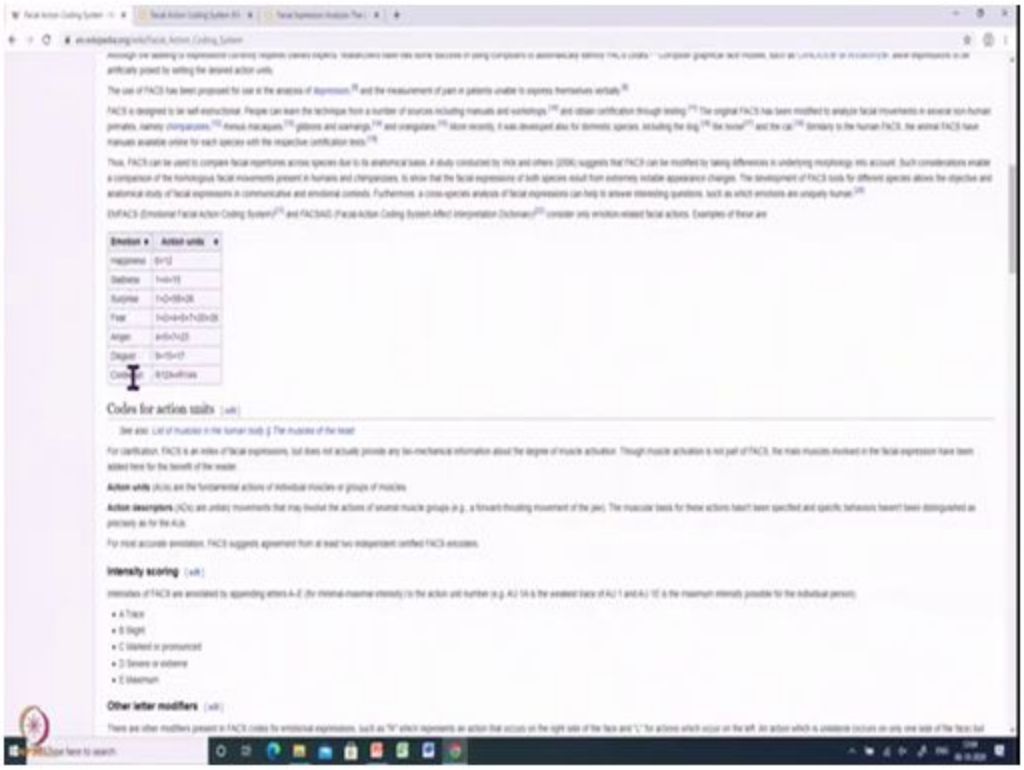
What is inner brow raise, and what is outer brow raise. So you get trained on this coding, once you know getting trained of these coding.

(Refer Slide Time: 3:47)



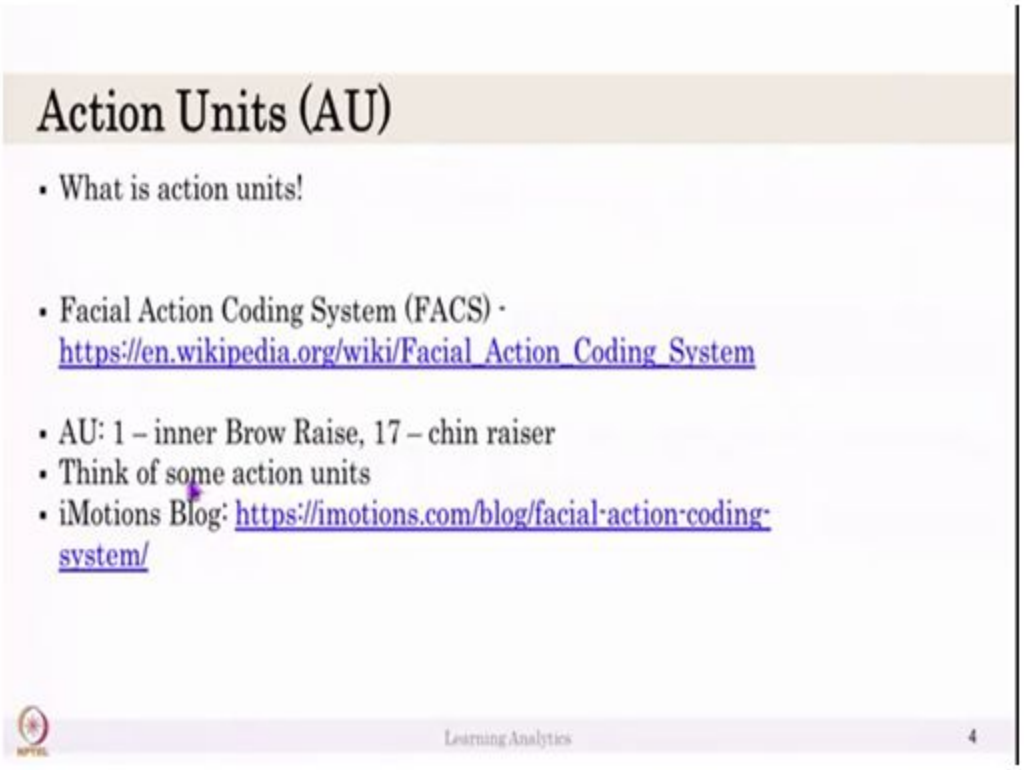
Now you can say happiness or joy is 6 +12 that is cheek raise plus lip corner puller which we have both, then it is a kind of happiness, you can say it is happiness. And sadness is kind of that. So you can combine these action units and that can produce an emotion. The first step is identifying what are the action units and you get yourself trying these 3 combinations are sad, this thing combination is happiness or something like that.

(Refer Slide Time: 4:32)



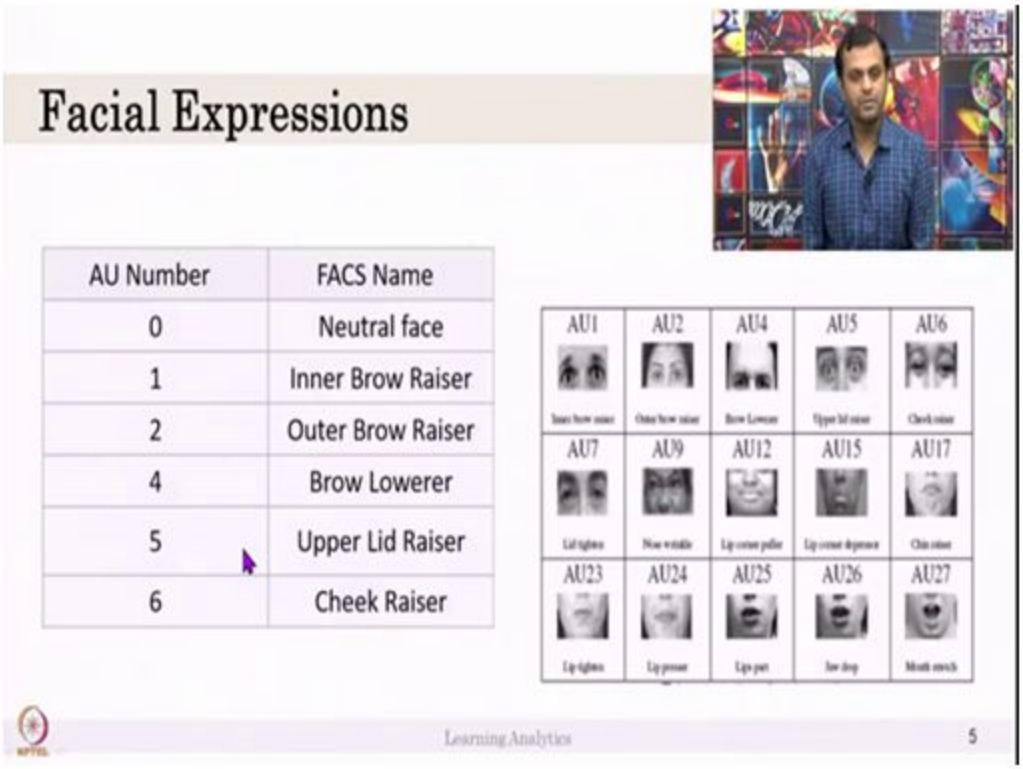
Then that is exactly your happiness, sad, surprise, fear, anger, disgust and contempt. So this is seventh one contempt is recently added. So there are six basic emotions, which Paul Ekman suggested in 1960s paper. So this is Paul Ekman's work and using this you can actually detect the human six emotions. Let us go back to the slide.

(Refer Slide Time: 4:49)



So some of the action units we thought about it, so we check the blog again. And how do we combine how to do it in automatically is we will see.

(Refer Slide Time: 5:00)

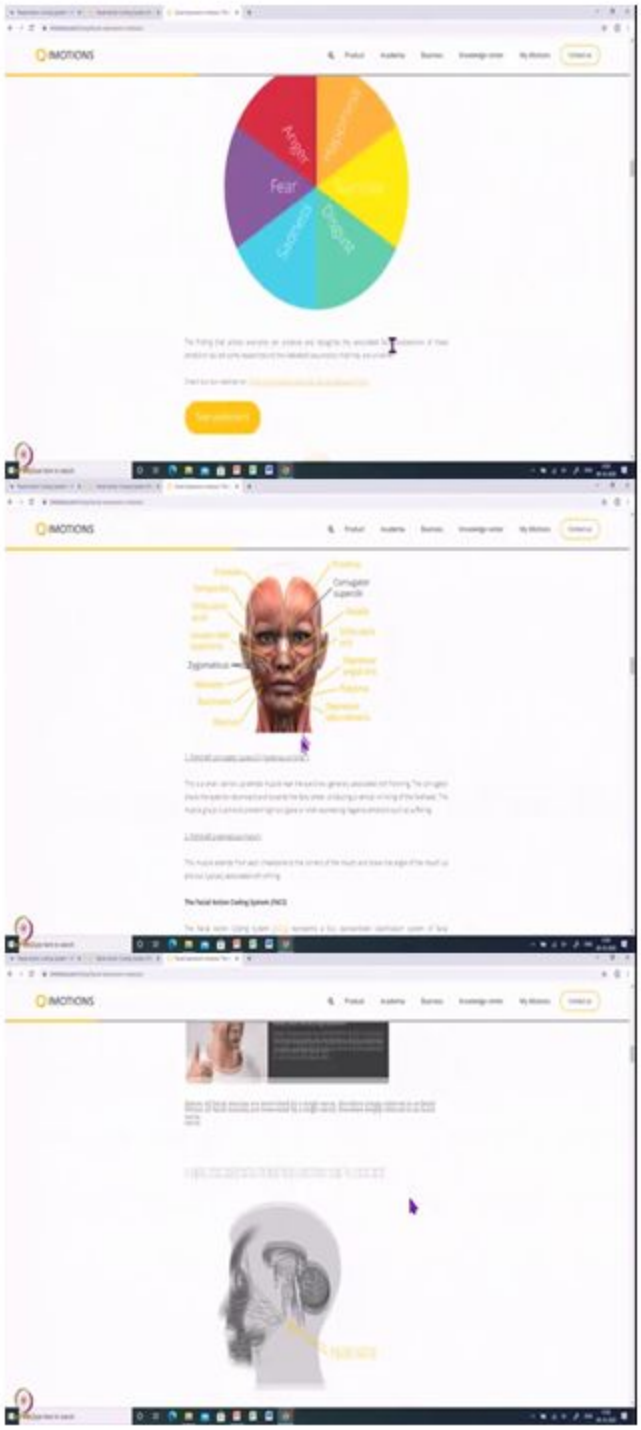


So there are more examples like AU1, AU2 with the figures and pictures, you can check those details here. And all the blog which I showed in imotions is really helpful to understand that.

(Refer Slide Time: 5:13)

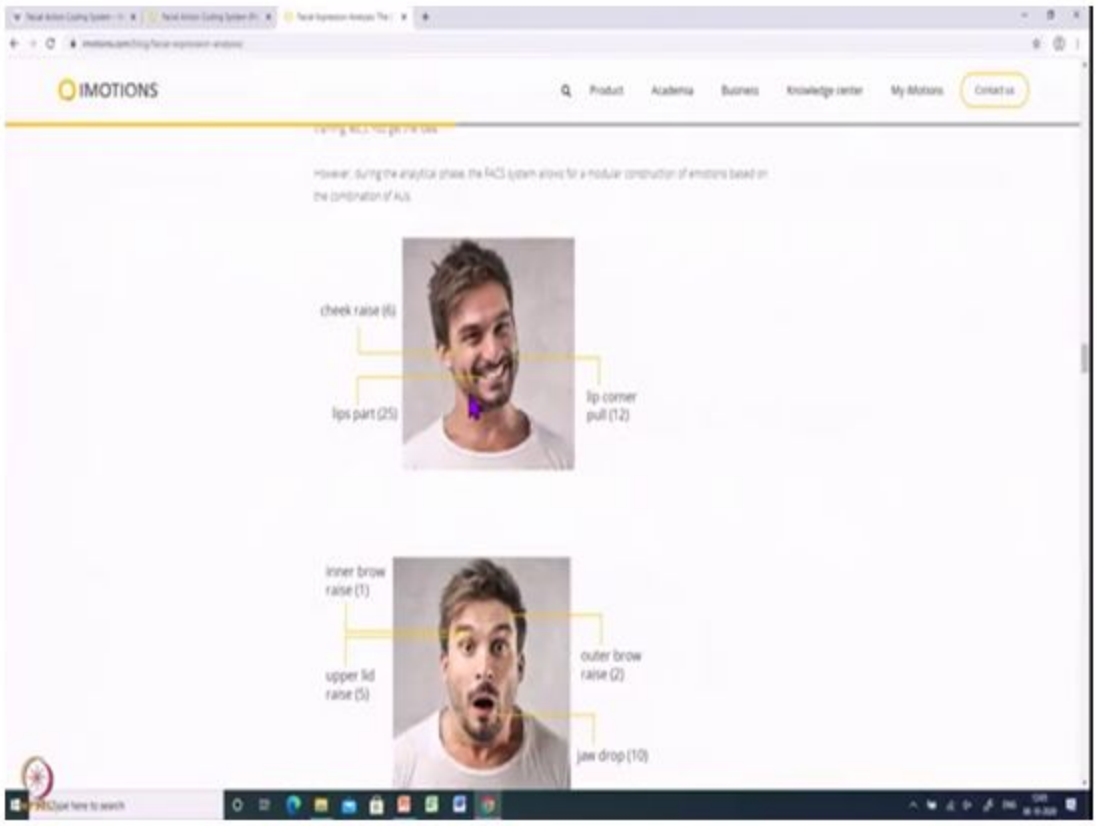
So what is Affect? It is a combination of action units like 6 basic emotions and how to detect these affective states. Let us look at another blog by imotions.

(Refer Slide Time: 5:26)



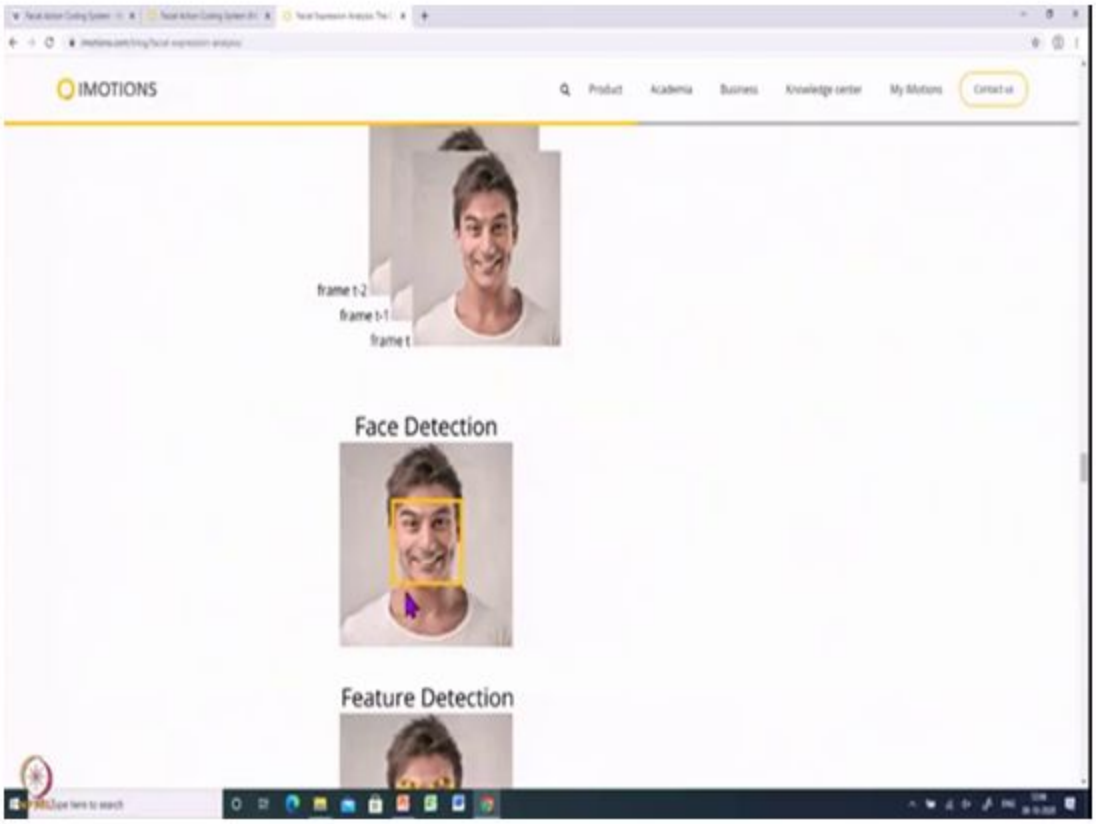
So, what is a facial expression? So, while it is happening, the details are given here, you can go more detail and the basic emotion is what detected here and you know, there are a lot of details on why this expression is happening, why the 24 action units, the details are given here.

(Refer Slide Time: 5:46)



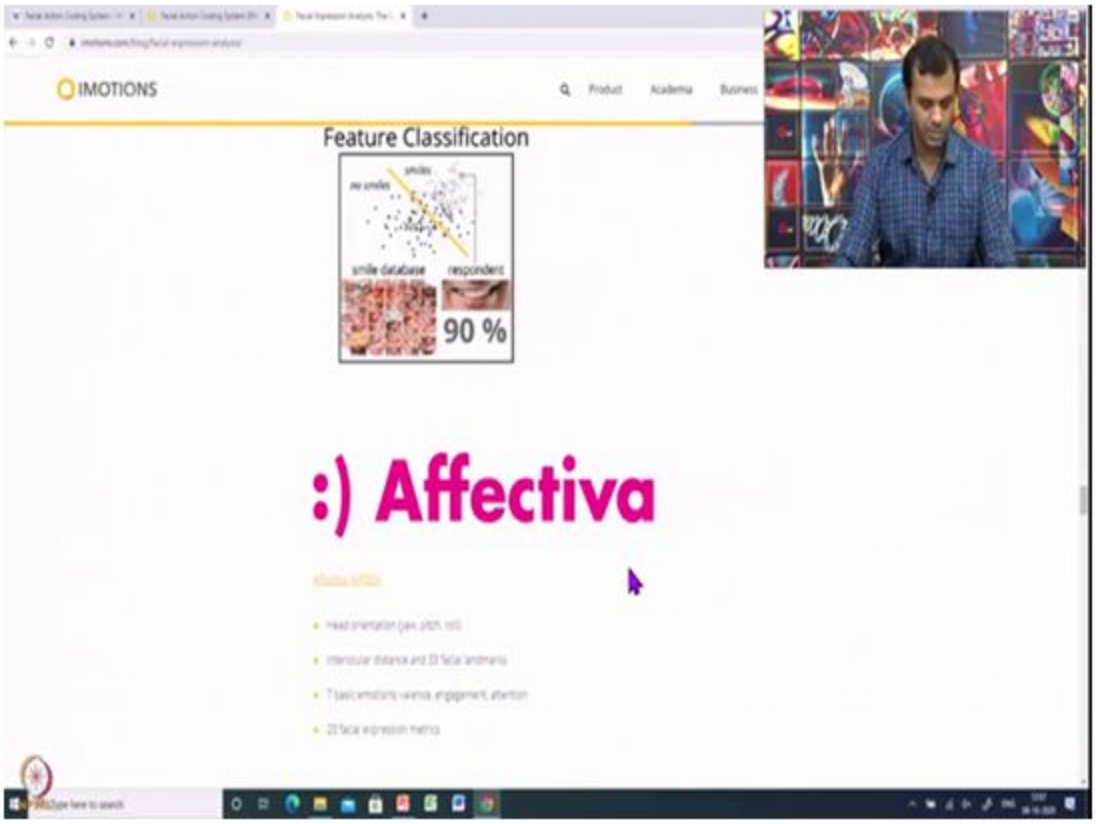
So, for example, the 3 action units indicate the person is kind of happy or surprised or something like that. So, or also these are the combination of the expression that indicates the person's emotions. So, how do we detect that automatically by using a web camera or something like that?

(Refer Slide Time: 6:12)



So, what happens really is on a web camera, we capture the video, actually, video is 25 frames per second, 25 pictures in a second. And a second, you will have 25 frames you know, the frames time is there, what happens here, the first step is using the AI, we detect only the face of the human. So, in the whole picture, the face is detected. After detecting face, face detection is easier compared to detecting the macro expressions that are “AU” because a lot of work has been done in this field of detecting a face or detecting facial expressions or facial detections nowadays, so face detection is easy.

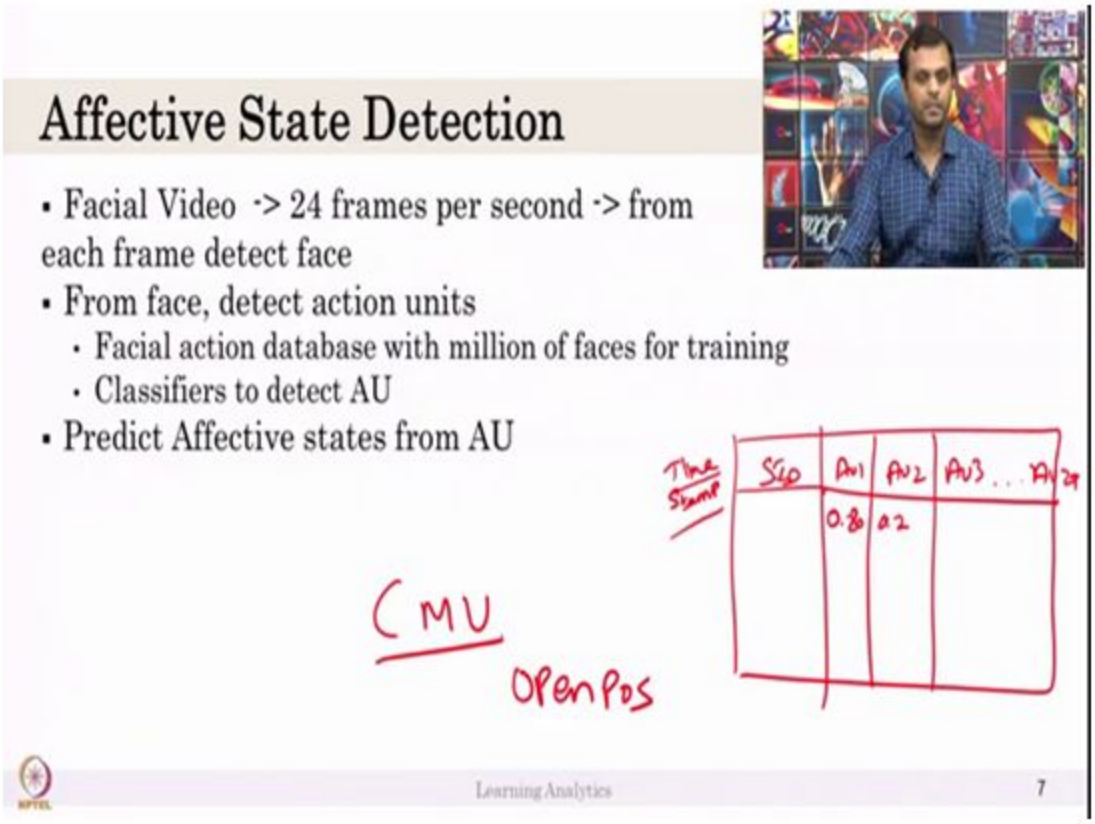
(Refer Slide Time: 8:00)



After detecting face, the part comes here is like detecting the action units, like marking them which part is eyebrow or what is the cheek, what is the lips, nose, you marking them and classifying them as action units, au1 or au2. So, to do that, there are billion faces database and they have trained all the face has been labelled properly and they trained it.

So they compare with that, and they classify the given image as smiling or not, or lip corner pulling or not, those kinds of classification is happening. Once you classify, if you identify the au you know, that is the company called Affectiva was doing it, the Imotions is actually a company which integrates products from different companies. This particular model we are talking about is by a company called Affectiva. And after you collect these facial expressions, after you have these action units, again give those to a new classifier the classifier will help you to detect the emotions that is the whole idea of how to use it.

(Refer Slide Time: 8:13)



Let us see that again in briefly. So, what happens here is your facial expression in a video is captured. After that, it has 24 frames per second and from each frame it detects the face, after detecting face it detects the action units that is the action is detected by comparing that particular area with the training database.

And from that information, it classifies the action unit is happening or not. There are classifiers are trained to detect action units like a simple machine learning class, where most of them it supports data machine or advance neural network source. So, what happens here is from the action we have to predict affective states. So what actually happens is, for example, the data will be like this.

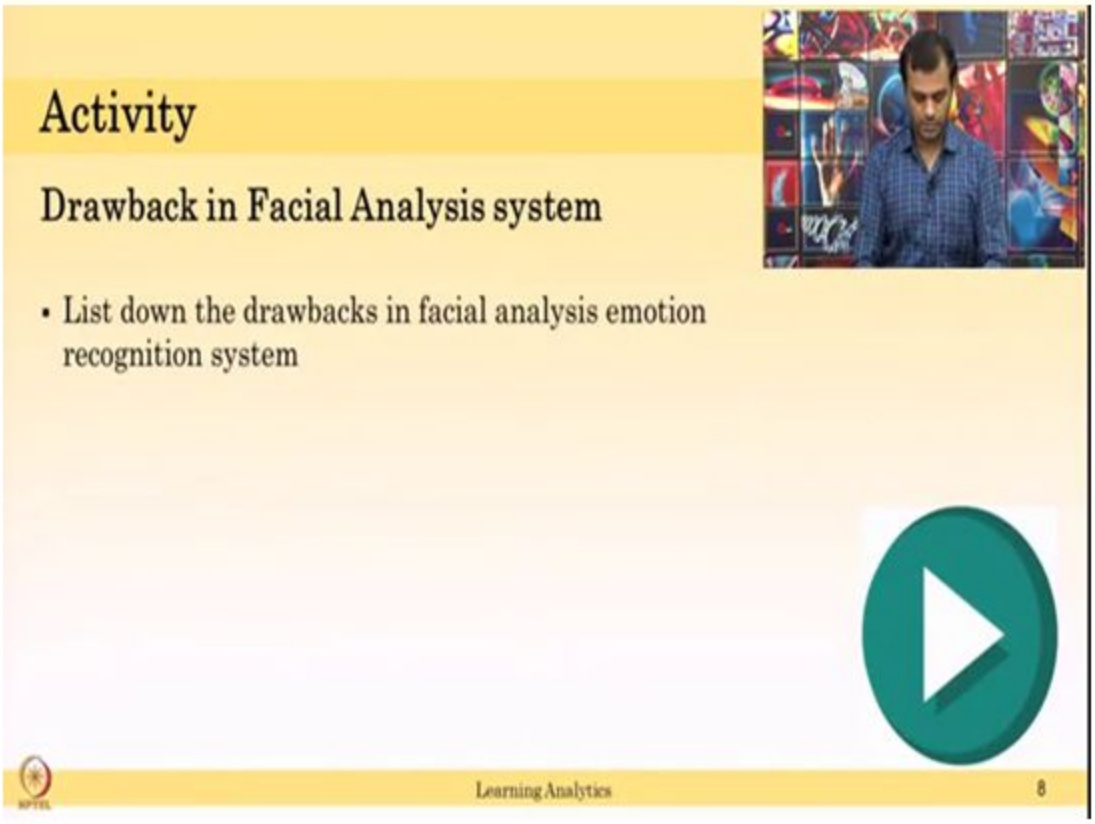
Say suppose the student ID action unit 1, action unit 2, au 3 and au 29 or 30 then timestamp. So, what happens is basically the action unit is not giving a binary classification, instead of binary

classification, they are given the probability value saying that what probability that this action unit might be occurring so say 0.8 some values, this is a 0.2 or something.

So now you have a lot of values coming or you know au 1 is happening, it might be au 1, also au 7 or au 8 we do not know. So again, they trained a big classifier and to combine this action units to detect and give it the effective states i.e. emotions - boredom, confusion or anger, fear kind of emotions. So, that is what is actually happening in the backend while detecting the student's facial expressions.

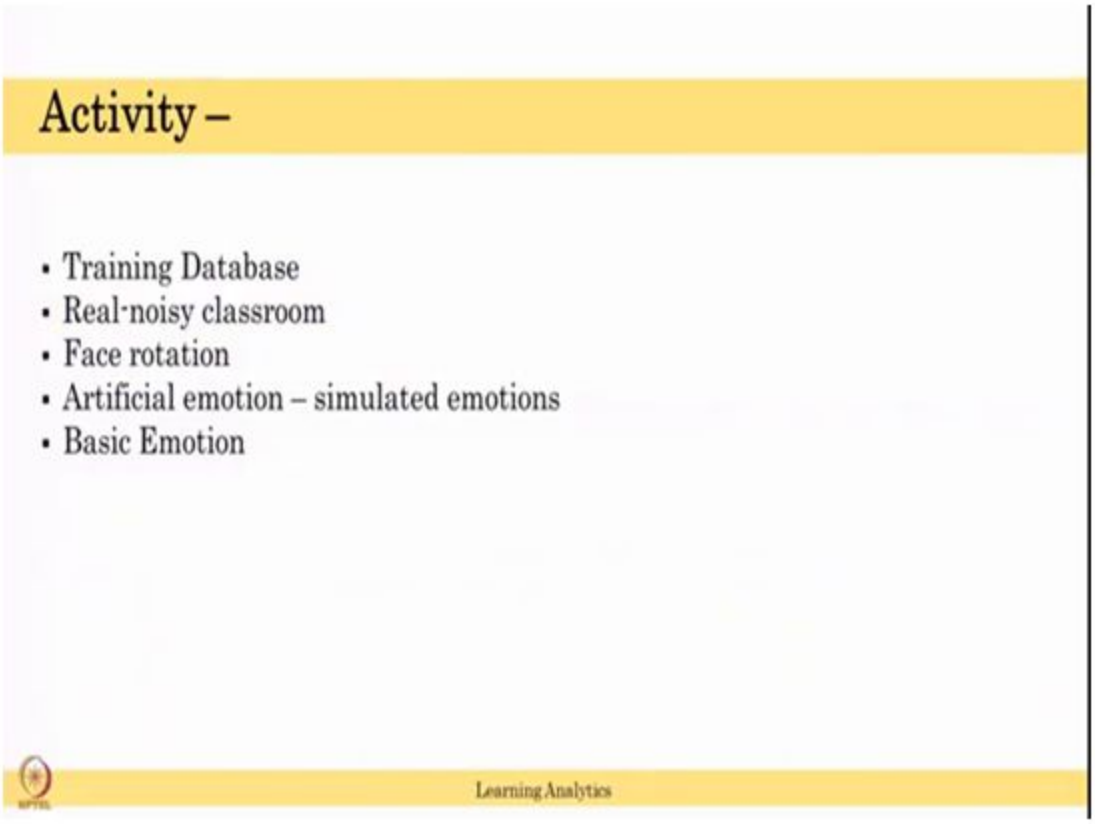
So, Imotions is a commercial company and the cost in India, it cost is really high, it could be around 5000 to 6000 euros. There are some open-source things available from Carnegie Mellon University, it is called Open Pos. So, from CMU Open Pos another library available in GitHub for free to use it, but CMU is not giving you the emotions detection accurately or final emotions maybe, but what you have to do is check those open posture and they detect not just facial expressions, also the posture of your body. And using that data, you might have to do the labelling by human observation or self-reporting, then you have to train your own classifier, we will see how to do that now.

(Refer Slide Time: 11:23)



So, before going into detail about the papers on facial expressions, how the interaction between facial expression has happened, can you list down the drawbacks in facial analysis emotion recognition system, like the one we saw in a previous slide like Imotions, where they use student's facial expressions from the video and they classify the face and classify the action units and they detect emotions. Can you list down the drawback? Yeah, please list down the drawback then resume the video to continue.

(Refer Slide Time: 11:58)



So the drawback the key is training database, you need a lot of training database for this. And which means millions of data has been labelled to create the data and to classify action units correctly and provide accurate emotions. And in real places, like noisy classrooms, it will not work, a student has to interact alone, or if 2 students are working you have to map it 1 or you have to show anyone's student's facial expressions, detect emotions and do that. And facial rotations are not captured by Affetiva, but "appdecks" do that you know the movement of the head. That data has to be if I just changed my face and I just show my different emotion what happens, the system may not able to detect that.

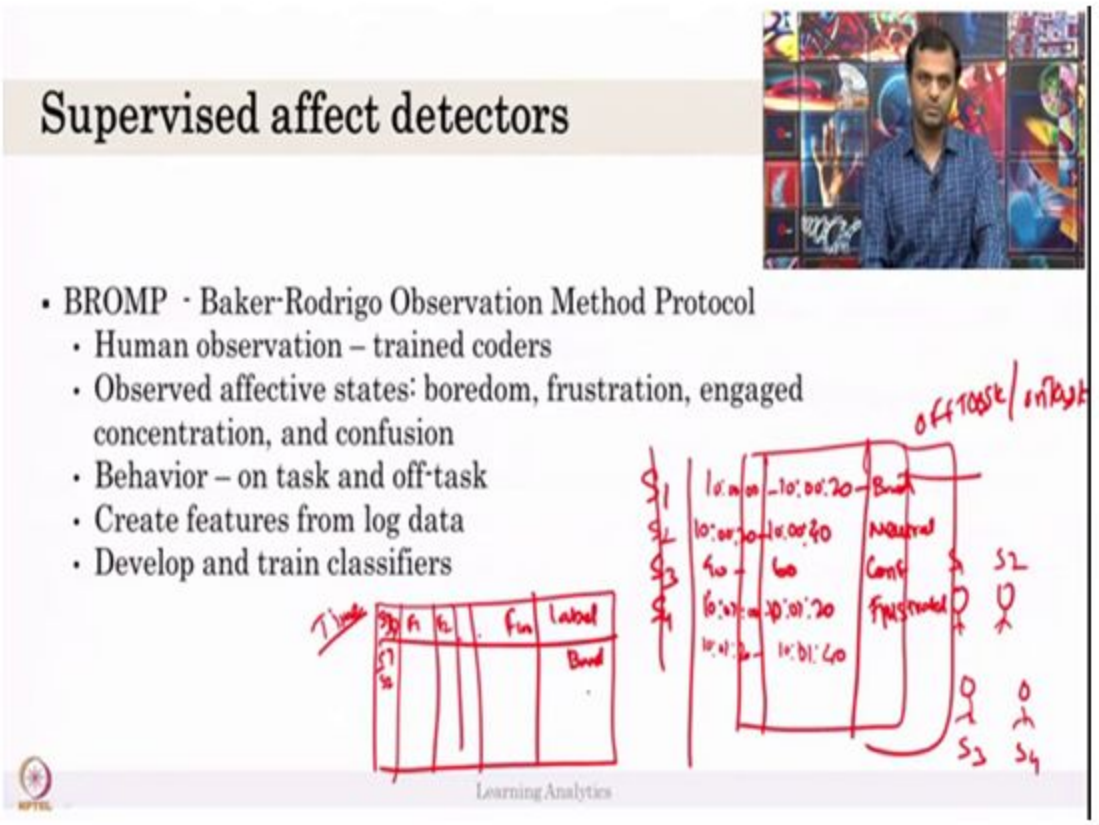
And artificial emotions i.e. simulated emotions is used to claim these kind of classifiers, how they got labelled because they asked the participants to simulate the emotions like you show happiness or you show anger then they try to code and that is the labelling is happening. So that may not be correct to detect the real natural emotions.

And the most important problem in Imotions or the facial detection system software is, they detect only basic emotions, because Paul Ekman studied basic emotions, and he listed down all

the action units. And you know which kind of action is combined together will give you the basic emotions, but when we are interacting with the learning environment, anger, fear, you know, that is not going to happen anger.

Instead, that we move learner-centric emotions as I was telling, boredom, confusion, frustrated, engaged, delight, those kinds of thing. So in order to detect those emotions, we cannot use imotions kind of software to detect emotions automatically.

(Refer Slide Time: 14:09)



So I will show a couple of methods about how they do it. 1 is a human observation. In this human observation, there will be coders trying to decode and there is a tool called BROMP, it is created by Ryan Baker, Jacqueline, and Rodrigo. So Baker Rodrigo observation method protocol and the prompt tool is a simple tool.

It is an automatic system, do not get that. It is an automatic system to detect emotions. It is a simple mobile app, It is the app, what happens is if you are observing, say 10 students in the class, you are the human observer.

How do you observe basically, so we want to observe. So first you will observe in a round-robin method, first, you will observe student 1 for 20 seconds, say 0 to 20, you observe the student 1, and this is student 1, and you will note down the emotions, maybe bored, then you go to s2, then you note down you observe for 20 seconds, then you will note down the student is neutral, there are no emotions here, he is actually focusing engaged, then 40 to 60, you might observe student 3 may be confused.

And you will be observing the student's emotion as say s4 maybe confused, bored or frustrated. So what is happening here is, as a human observer, you are actually having a pen and paper. And s1 you are looking at them and making sure there is a time, we might be looking at the time, then you know. So instead of that, the prompt protocol tool actually helps you to do that. You just have the student name, you click the student name and the time is automatically recorded from your computer and there will be like 6 emotions, you have to click it. So your time of recording is saved that is the simple observation method protocol, the tool which tells how to do it. But here, they observe the humans not just a facial expression, the holistic approach, you know, the student's interaction with the system, the gesture, the screen, what are they working on? Are they talking to a peer, all this information is observed then this emotion is noted down.

Why 20 seconds? That is a very good question. They want to observe at least 20 seconds and make emotions. And if you talk to a psychologist working in HSS department, they say emotion will not last for 20 seconds, there is no possibility that you get confused for 20 seconds, the emotion surprise comes for 3 seconds, 2 seconds that is it. But if you want to go to 2 seconds or 3 seconds, such a fine-grained kind of emotions.

It is not possible by the human observers to observe those every time unless you are observing, recording the student's action in a video then observing, coding every second by second. The problem is, whenever video to code it might take more than 2 or 3 days, such a huge data such a fine-grained data you will end up with if you observe the videos file every second. So 20 seconds is the logical approach they found out, it may not be true, you might have the new ones.

So BROMP is a protocol, it is a kind of app which helps you to map it. If you are happy, if you have your own notebook, and you have your own timing schema to do that, it is good. You always go s1, s2 in the round-robin method, there is no need to record it. And you have a timer above it every 20 seconds you press the timer and you mark it, then no detail timers needed, just you have to save, bored, confused, neutral, then you are good. There is no need for the app or tool that is it.

So, basically what happens is, it is a human observation the trained coders, the coders got trained in the sense so 1 recorder couple of coders code the student's emotions and they make sure that interpretability is high, the Cohen's Kappa is you know 0.8, then they move on to the next, then observers got trained and they can go independently start collecting data from the students.

So the observed affective states can be different, whatever emotion you want to do. So now we can use the learner-centric emotions like boredom, frustration, engaged, concentration, confusion, all these things. You can also detect another behaviour, the off-task or on task. It is very easy to detect, this is one of the easiest to detect, are they talking about the task they are doing or they talking about something topic and still engaged, that is the whole thing you can also observe. And after you observe this data, which means student 1 you observe at 10 minutes 20 seconds, again, you will be observing the student 1 at 11 minutes 40 seconds. So you will observe a student at every 80 seconds.

Once you have data like that, then they create a lot of features from the log data. So what happens here is, again, you go back to your features creations using the machine learning classifier, for example. So, what happens is the timestamp, so there is a timestamp and student ID - s1, from the students logged data that is the interaction of the system you compute feature 1, feature 2, so 100 features, then there is a label.

The label is the emotion you detect that is bold. Maybe this is student ID, s1, s2, if you did that, you know, if you create that, so which means you need to have at least thousands of observation happened in the real classroom, then you will have this table, after having this table, what happens is you use machine learning classifiers to predict the label from the features and that is it, that is the whole idea.

And they create different ML models for different emotions, we can train only for boredom, you can train only for confusion, so different ML algorithm might give a different result. So use one which gives a better result. This is the whole idea of creating automatic emotion detection using log data. Here the labelled information is from human observation.

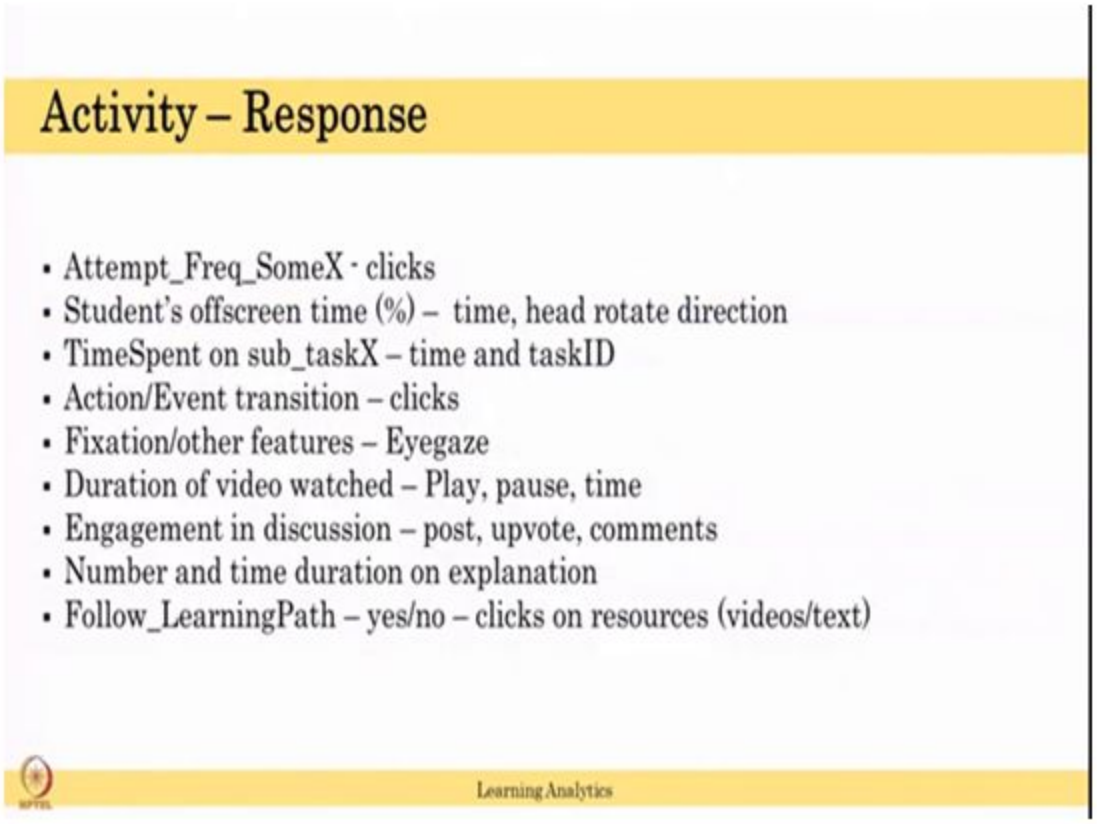
(Refer Slide Time: 21:59)

So, I said that in a TELE you try to come up with a lot of features to detect emotions. Now pause this video and list down what are the five features you think is needed to detect emotion from a TELE? We know, what are the interactions people do with TELE, you know that they interact with a problem map or they create a concept map, they go to simulator, and imagine the TELE we discussed in our class "MettLE", and try to think of five emotions.

Think of any technology-enhanced learning environment, anything you have used and think of what are five features, you think it is important to detect emotions. So the 5 is not number, just

giving you the idea to think about features to extract from the log file, pause this video write down the answer, after writing it down assume to continue.

(Refer Slide Time: 22:51)



So, there are a lot of features used in emotion detection; item frequency, clicks, how many clicks have happened or other interactions happened? What is students off-screen time? Are they looking at it or not? If you have facial expressions camera or a webcam? Or are you spending the time on this particular task?

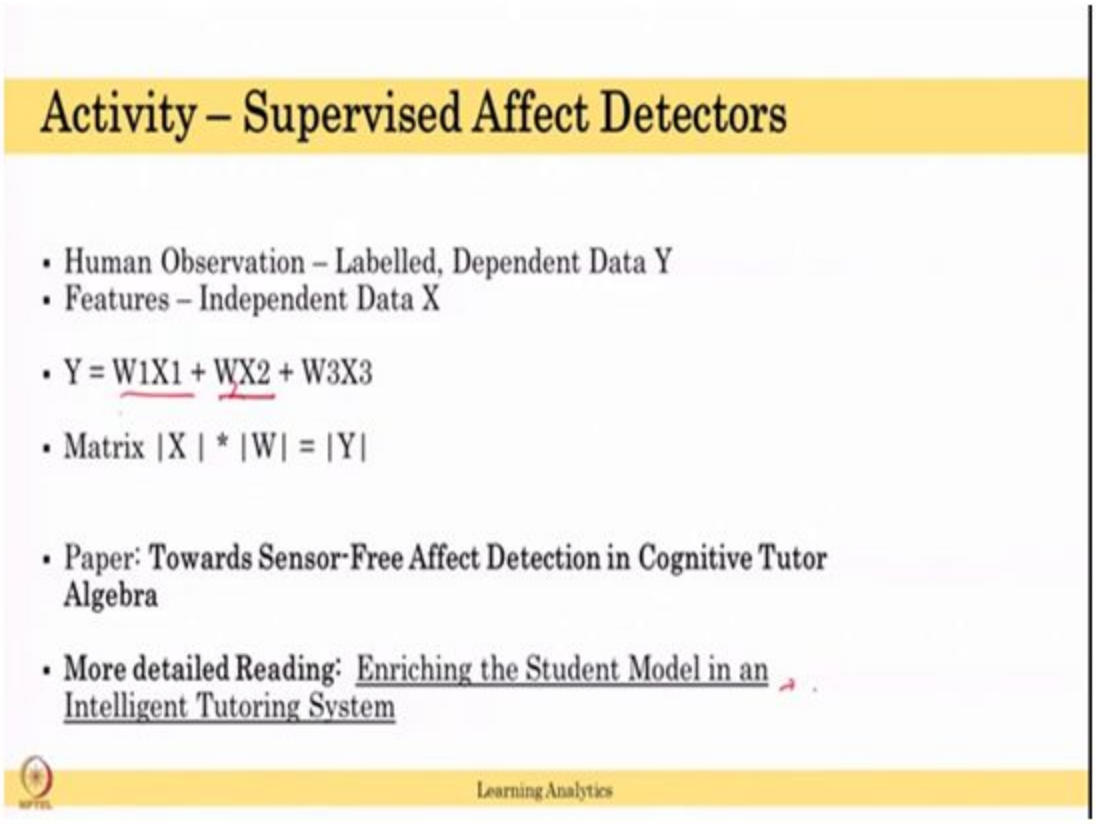
And are they spending time on this particular transactions from this page to this page? If you have eye gaze data, you can talk about fixations or if you are watching video how they playing, pause time, list down all those features.

And make more number of features, for example, how many clicks in the last 3 minutes, how many clicks in the last 5 minutes? How many things in the last 10 minutes? So the frequency for

last few instances, all the time displayed, all my test time spent on this particular video in last 1 minute, how much time they spent on reading resource for last 5 minutes such features, we expand it more.

This features generation comes from domain expertise. Professor Ryan Baker, when he was telling that this knowledge he gained from creating detectors for multiple systems over a decade and he is able to create the feature engineering and features come up with. So it is not an easy task, you will not get it all the features in the first time, you will not get a better classifier in the first time, so it starts with that and collect more features, create more features or read papers, published papers from Ryan Baker on detecting emotions using different systems then you get the idea how to create features.

(Refer Slide Time: 24:52)

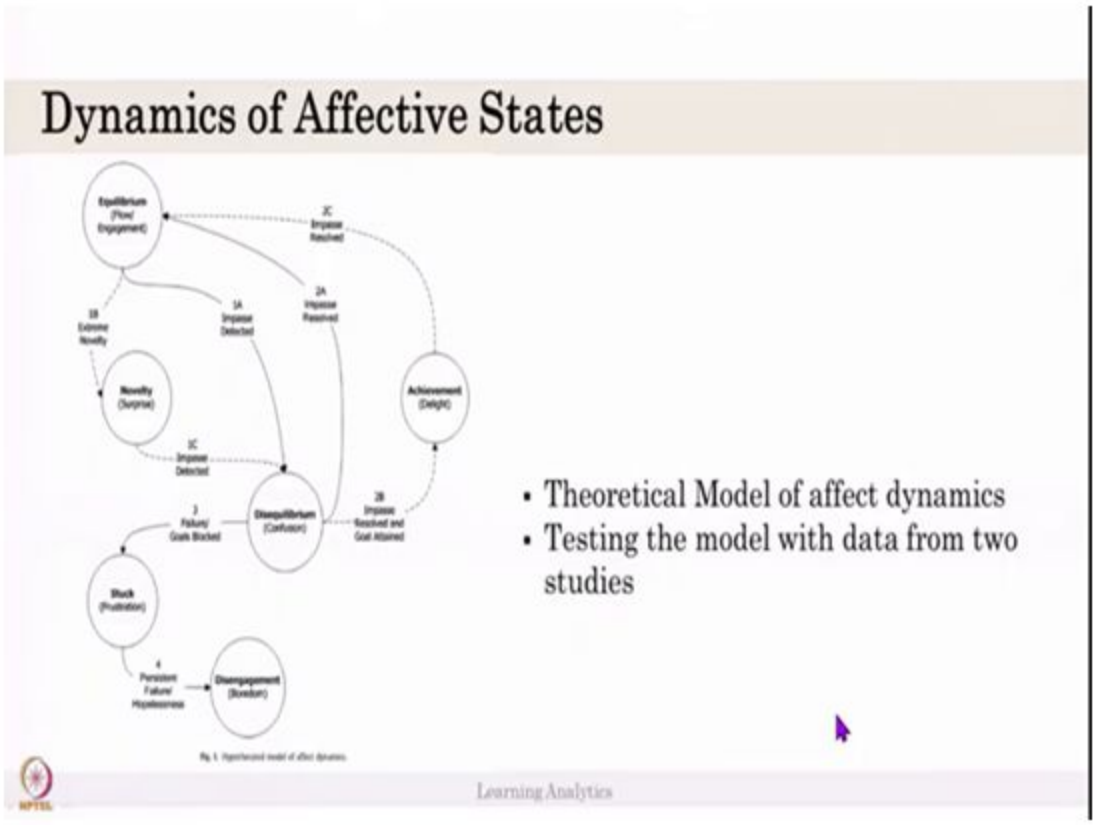


So after you create features, human observation using independent data, features is you know, the x is the features and let us take a simple linear equation kind of approach,

$$w_1x_1 + w_2x_2 + w_3x_3$$

kind of weight, which is what you are trying to estimate from the training data. It is kind of a matrix, once you have a matrix you can apply to multiple ml methods. So, check this paper, it is an interesting paper, check it or if you want to more detail on detecting only how to detect frustration from a log data, check this thesis, I worked on this to detect student's frustration from log data and here also I used human observations a lot.

(Refer Slide Time: 25:45)

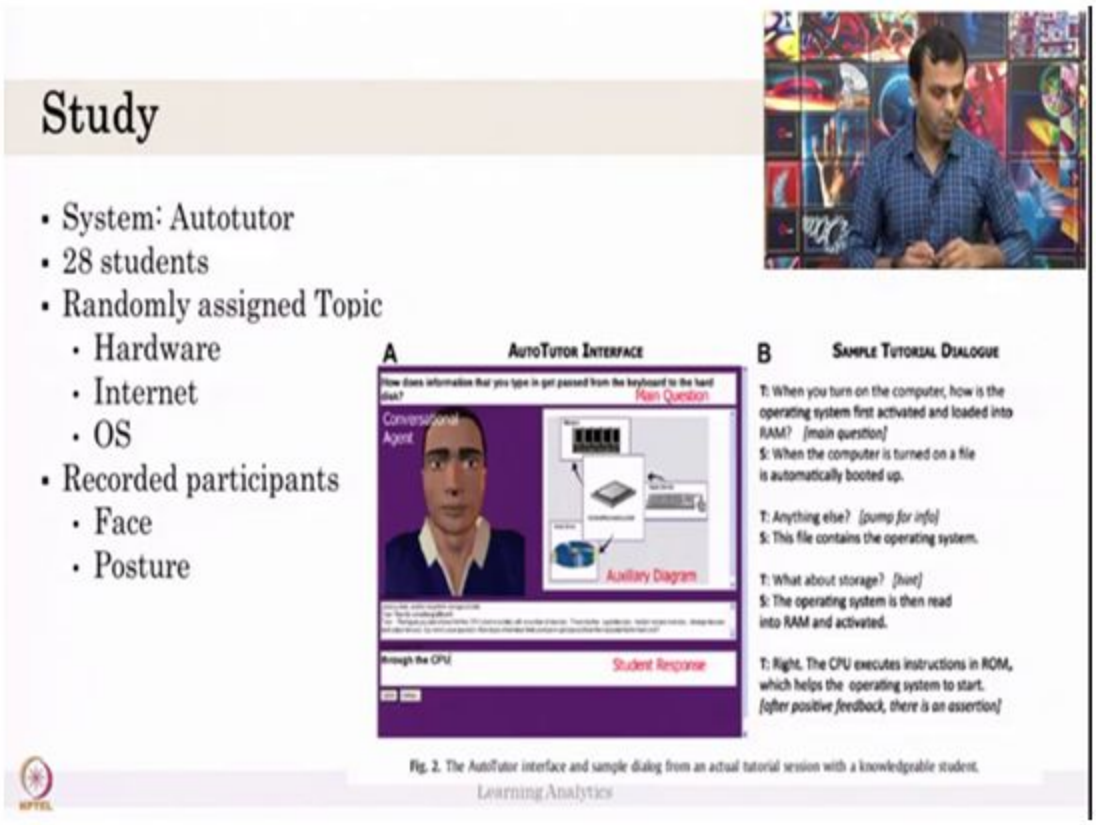


So let us move on to other approaches. Here, Professor D'Mello tried to understand the dynamics of Affective States.

So this affective state is from, you know, psychology. So from the theoretical knowledge of affective states, this is what is kind of coming up, like there is a flow, flow leads to confusion, if there is some impasse detected, something is stopped. And if the confusion leads and your goals are blocked, then you might get a frustration. And if the frustration continues, you will get bored and you might drop.

And if suppose if the confusion, it gets resolved, there are some issue, you are reading it, you are stuck in it, you are not understanding, you might get confused. If that stock part you know what has stopped you, which has got resolved, then you might go back to flow, flow is like engaged. Again, you continue to engage in particular learning content, this is the dynamics of affective states from the theoretical approach, let us see, is it holding true in a data we collect and predict.

(Refer Slide Time: 27:10)



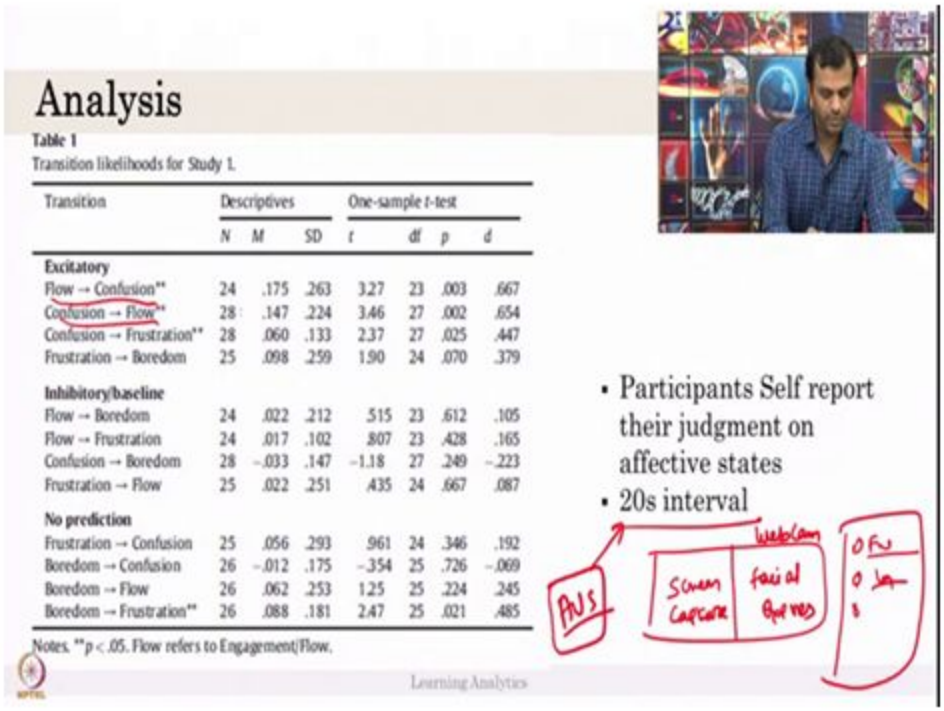
So what D'Mello and his colleagues did this, they used a system called Autotutor, it is voice-based interaction system where you have to speak to the system and based on what you

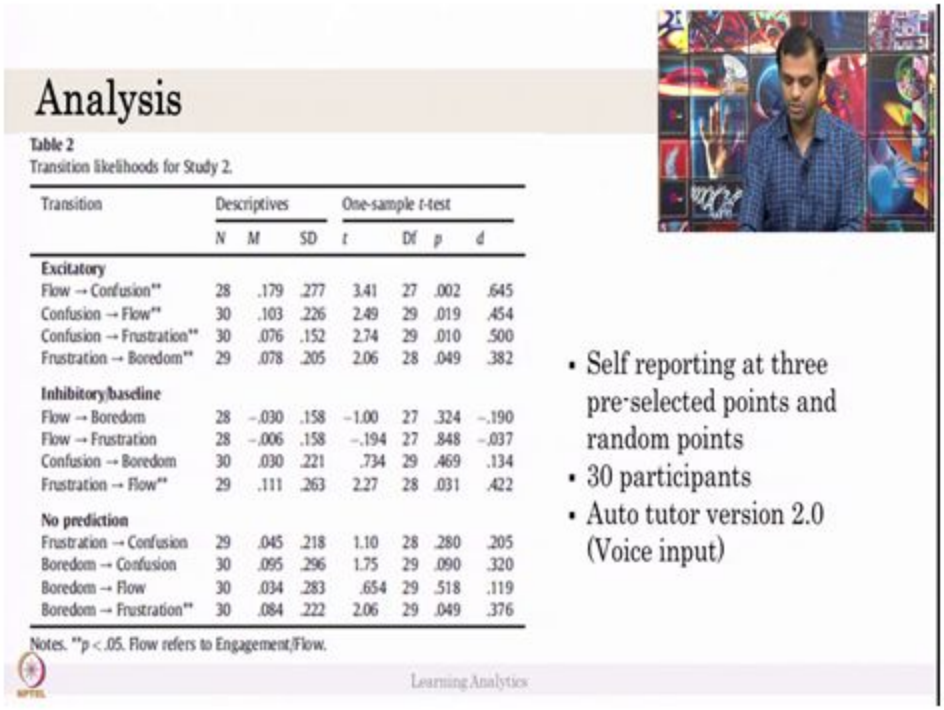
speak, the conversational agent answers your questions. Here, you have to learn topics like hardware, internet, OS or something like that.

For example, the system asks you the question, when you turn on the computer, how was the operating system first activated loaded into RAM. So, it is actually a question, the agent actually a text to speech converter speaks. Then the student types answer instead of speaking because voice recognition was not so great at that time when this was created.

So, when a computer is turned, a file is automatically booted up, then it asks you for more information like what happens. So, then if you are stuck it might provide a hint all these things happening in this auto tutor. So, imagine a student is working on this kind of environment like Autotutor and what you do is, you record student's facial expressions and posture, also you record the interaction screencast, screen capture, you observe both.

(Refer Slide Time: 28:31)





So, after observing both what you do is you ask the students to self-report, what I mean is, so after the student has interacted with the system, show 2 screens to the system, 2 screens, one is a screen capture, one is students facial expressions, the webcam image, this is from the webcam.

Now, if a student interacted with the system for say 30 minutes, you pick 20 intervals of time note or 30 minutes if you want to say I want to detect 20 instances of emotions.

So, at every 1 and a half minute or every 90 seconds, you pass the video and you show in five seconds of screen capture and the facial expressions, ask the students to self-report, there will be a menu brought, ask the students to report the emotion like a frustrated, they just select their option.

You show what was students doing a context to information also the facial expression to the student and ask them to report their own emotions at a particular time so self-report. You can have a human observation also to check is this happening or not or you can ask a peer to report the peer's emotions, then you can compare whether two people are reporting it correctly or not. So, randomly it picked 20 times and detect the emotions or you know that emotions can be used as a label and you can use data from Imotions action units, I was telling therefrom Imotions, you can detect the action units.

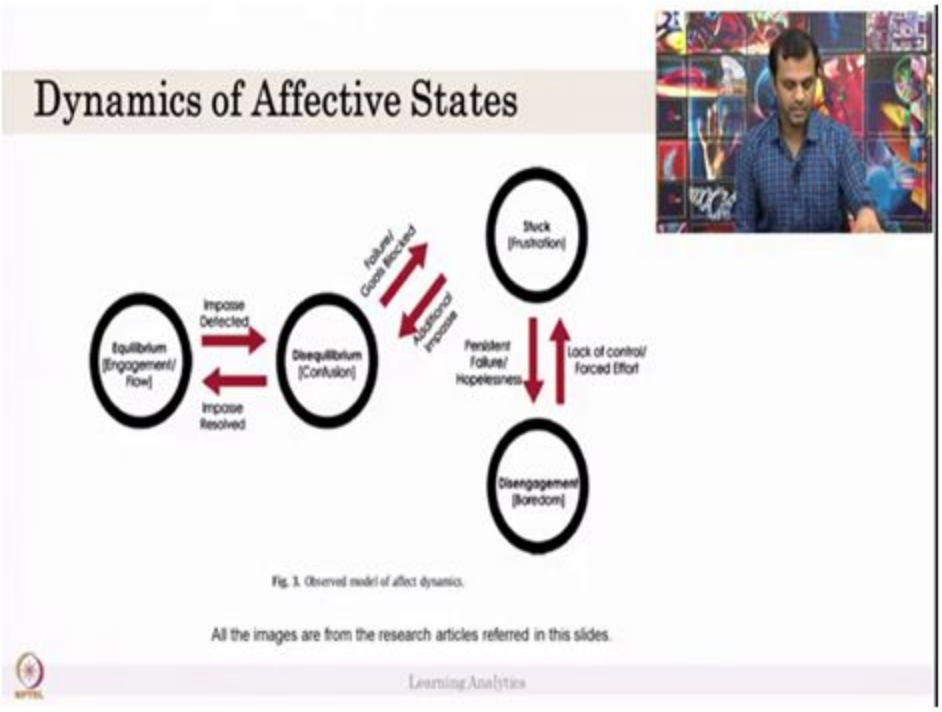
Now using the action units from the Imotions and then label data from the self-reporting, you create a machine learning classifier, again the classifier will be using 20 seconds interval, or you can use the log data 20 seconds interval also because the action units 20 seconds interval will be used to detect a motion that is also 20 seconds interval.

So how do you combine the action units to 20 seconds, get the dominant action units there and just do it or read the paper to know the exact approach. So from the analysing that what actually happened is this table has come out. Here, the double star indicates there is a significant flow. There is a transition from engagement to confusion, there is a transition from confusion to engagement, and there is a transition from confusion to frustration, these 3 are significant, and they did a one-sample T-test instead to check whether it is reliable or not and the significance is good and the number of students samples used is also given here.

Similarly, boredom to frustration, also significant transition is happening, for other things significant transition is not happening. They repeated the study again, with the new system and again, they found out, not all of them except frustration to flow is also happening, and frustration to boredom also is happening within the students.

Now in this report, they randomly selected 20 preselected points, also 3 random points have been given to students to self-report and 30 participants. And now with these 2 data from the empirical study, they go on to see whether we can plot the theoretical model of affected dynamics so that is what they did here.

(Refer Slide Time: 32:09)

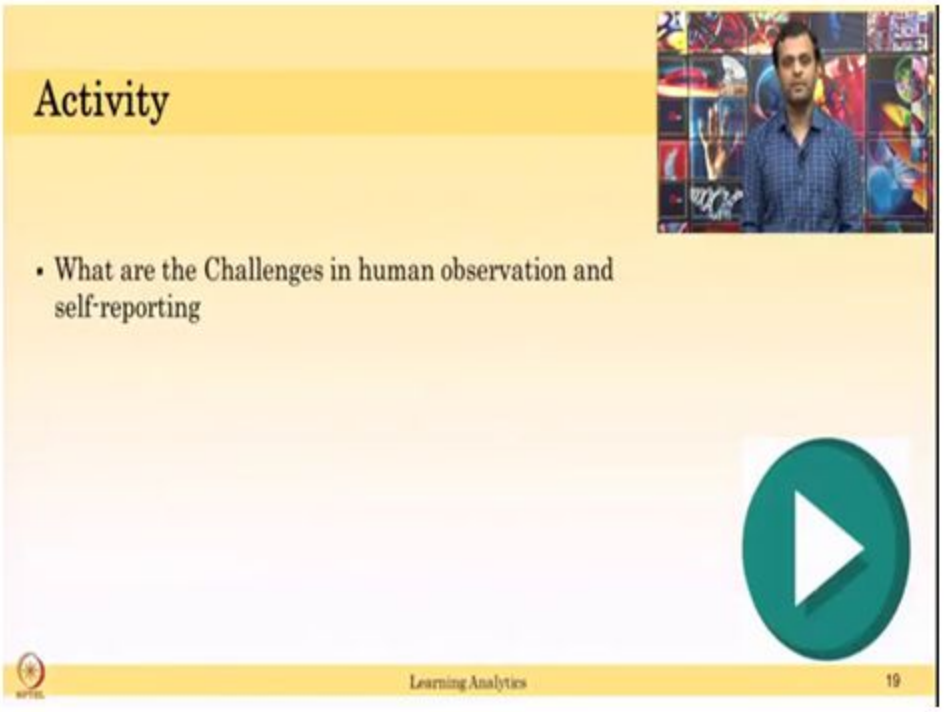


So this is the dynamics of affective states obtained from the self-reporting and predicting students emotions based on the action units. So here equilibrium, the flow is actually there is a significant transition in both, you can go to confusion, or they might come back to engagement, also a significant transition by both and significant transition between both. There might be a transition from boredom to confusion or boredom to engagement, but it is not significant that is what this indication says that. This particular picture says that wherever the significant transition observed from all the students is shown here, but it is not that there will not be any transition from boredom to content there might be a transition.

So this is how the facial expressions are used to detect learner-centric emotions that is, by human observation, and collect log data and predict or you can ask the students to self-report at that particular 20 or 30 instances of the time they are interacting with the system and predict the student's emotions.

And that is the 2 methods we saw and the dynamics of this affective state gives us more important information that if the student is in confusion state, you need to understand why that student went to confusion, what actually impasse is created. If he solves that, he will go back to engagement mode, otherwise, he will get stuck in the frustration mode. So yes, so we have to think of what kind of feedback can be given at this particular moment. So that is why this particular model is very, very important.

(Refer Slide Time: 33:51)



So we saw human observations, and the self-reporting has been used to detect emotions. What are the challenges in human reporting? Please think about the challenges. Resume to continue.

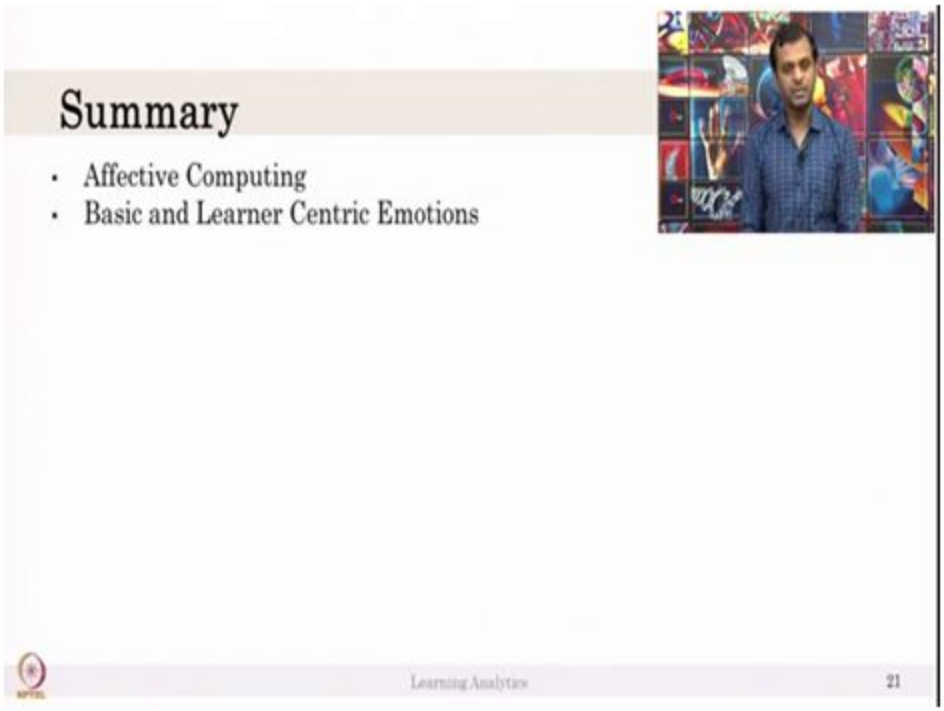
(Refer Slide Time: 34:07)

Human observation and self-reporting are time-consuming. Firstly, if you do human observation, I said that it takes several hours to code 5 minutes of video, and self-reporting itself is a problem in the sense students are self-reporting, it may not be true. So you have to confirm whether the self-reporting is right or wrong.

And so if the human observation is happening once in 20 seconds, and self-reporting also happens in the interval, the accuracy of this particular system is not so great.

So our case is not really great, it is really bad and we cannot go and make a decision by using the report coming from these indicators or detectors, so that is the issue. And when you do the human observation, the main thing is please do the inter relater or inter-observer reliability that is Cohen's Kappa, and make sure we get more than 0.8. And that is not easy to train the human observation that needs more time and more training for the human observers, that is the idea.

(Refer Slide Time: 35:25)



So, in this slide, we saw what is affective computing and what are basic and learner-centric emotions. And how to detect those emotions from the basic emotions from imotions kind of webcam-based systems, automatic systems of open posture, and learner-centric emotions can be detected from log data or other data if you have a human observation or self-reporting as a label data. Thank you.