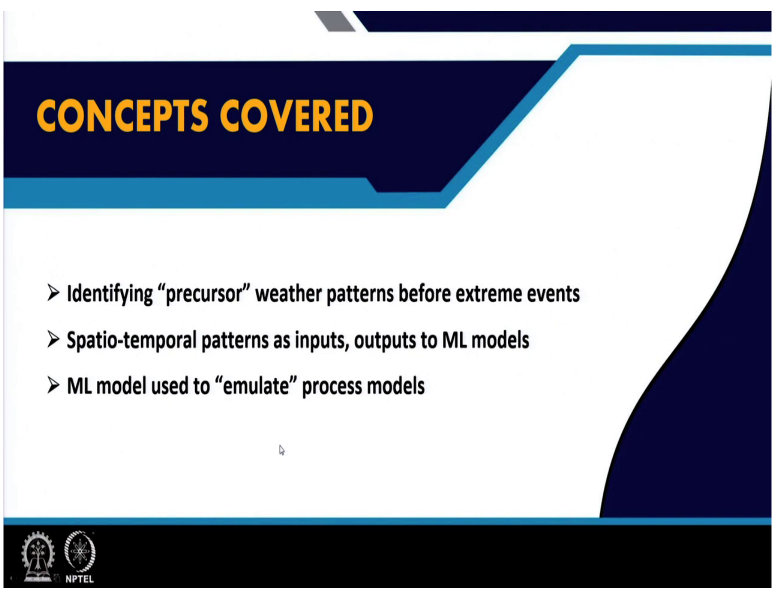
Machine Learning for Earth System Sciences
Prof. Adway Mitra
Department of Computer Science and Engineering
Centre of Excellence in Artificial Intelligence
Indian Institute of Technology, Kharagpur

Module - 03
Machine Learning for Discovering New Insights
Lecture - 24
Nowcasting of Extreme Weather Events

Hello everyone, welcome to lecture 24 of this course of Machine Learning for Earth System Science, we are still in module 3, where we are like exploring different applications of machine learning to answer various questions in and discover various insights in earth like various aspects of earth system sciences. So, today's lecture is going to be on Nowcasting of Extreme Weather Events.

So, like earlier, we had like discussed what are extreme events we had seen like how to discover such extreme events, like say tropical cyclones etcetera; where today's task is nowcasting of such events. Now, what is nowcasting? Nowcasting is the task of forecasting at short duration or like at short notice like of course, what is meant by not by short notice depends on the scale of the process.

(Refer Slide Time: 01:04)



So, like if the extreme event. So, in particular today we will consider two kinds of extreme events, one are heat waves or cold waves and the other is lightning.

Now, heat waves and cold waves these are basically things that last for several days. So, a nowcasting of them should be done a few days in advance. On the other hand lightning is a momentary phenomena and; so, forecast nowcasting of lightning should be something which is to be done like maybe a couple of hours ahead of time and so on.

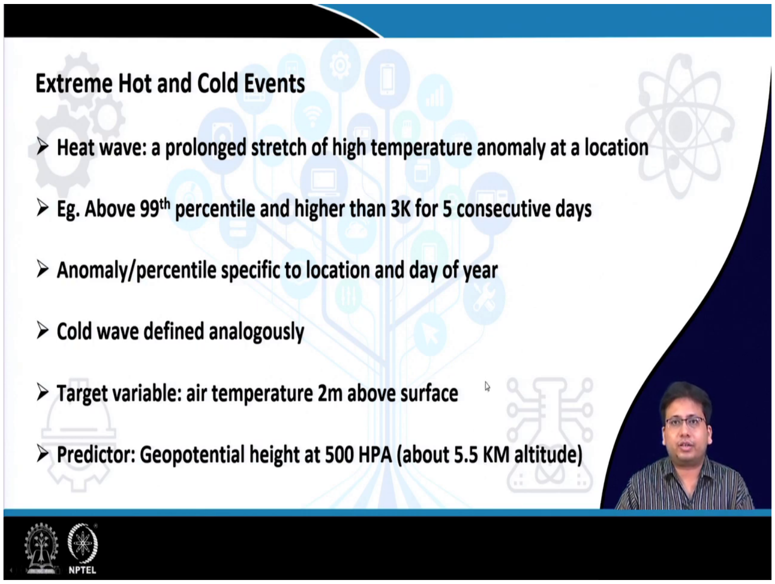
Now the concepts which we are going to cover today, are identifying precursor weather patterns before extreme events; that is, suppose an extreme event is going to happen, does the like are there some weather patterns which like which indicate that such a thing is going to happen, can we mine or discover such weather patterns.

Secondly the spatio-temporal patterns can be used as inputs and outputs to machine learning models. So, like the kind of predictors which we will be considering in this case, are mostly going to be a spatial maps, rather than observations at a particular location or at a particular time point. So, like we will need like most of the models which we our machine learning models that we are familiar with like they take like maybe a vector as input and so on, but in this case, the input is a map.

So, like in our when we are discussing various machine learning models, we had considered things like convolutional networks for spatial input and so on. So, somehow we have to utilize those properties. And thirdly, the third concept we will see is how machine learning models can be used to emulate various process models.

So, process models we have seen earlier also, we are like just in the previous lecture we were talking about the CMAQ model, earlier also we have talked about other process models. So, the like today we will briefly talk about, if machine learning can somehow be used as a surrogate for them; so, like of course, this concept will come back in greater details in the 5th module.

(Refer Slide Time: 03:35)



But now in today in this lecture, first of all let us start off by defining what are extreme hot and cold events; so, what is the heat wave? Heat wave we have all heard of heat waves, a heat wave is the prolonged stretch of high temporal and temperature anomaly at a location, that is, whatever is the expected temperature based on climatology at any given location, if the temperature remains significantly above that for a significantly long duration, then we say that is a heat wave.

So obviously, some concept of threshold is involved here; so, how long and how much above climate threshold etcetera. So, like for like was various standard definitions are used in different parts of the world, which may be like higher than 3 kelvin above the climatology or above the 99^{th} percentile for like again; for how many days this also can vary may be in some countries they may use 3 consecutive days some other countries they may go for 5 consecutive days and so on.

And whenever we are talking about these anomaly or the percentile and so on; I mean the amount of anomaly of this of 3 kelvin or the 99^{th} percentile, these are specific to location and the day of the year.

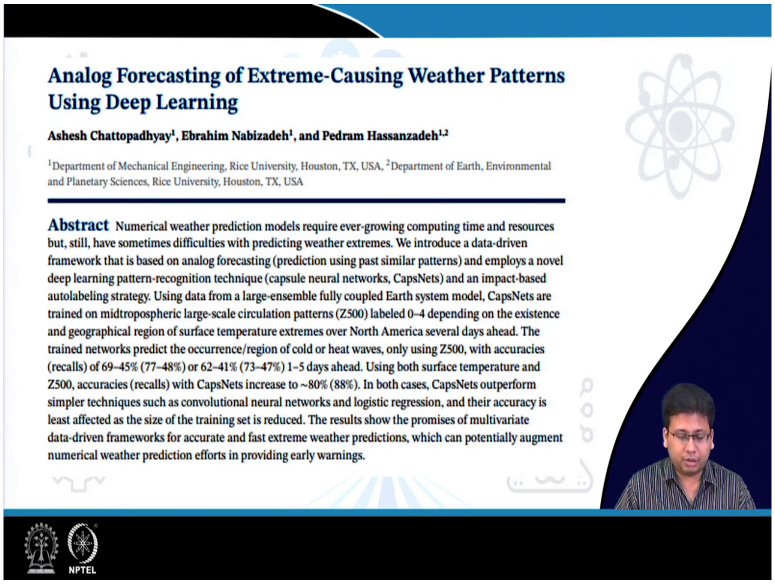
Say for example, like if the temperature is say 31 degree Celsius for 5 consecutive days in May that is probably not a heat wave; because it is actually expected that the temperature will be 31 or higher than that in that period, but if the temperature is 31 degree Celsius or more for 5 consecutive days in December, then that is a like a big thing, because in December, it the temperature is expected to be lower. So, these things are like these are usually not absolute, but they are specific to location and also the time of the year.

So, like what is a like what is the heat wave in Rajasthan may not be sorry what is the heat wave in Ladakh may not be a heat wave in Rajasthan. Similarly, we can define cold wave also analogously by like by consider instead of 99^{th} percentile if we consider the 1 percentile or instead of higher than 3 kelvin, if we consider lower than 3 kelvin and so on.

So, the target variable in this case, is the air temperature 2 meter above the surface. This is what the variable which we are focusing, which we will be focusing on to understand whether heat wave cold wave etcetera is happening or not. And the predictor is like the in the of course, there can be many variables which can be predictors, but in the paper that we are going to study, they have considered only one predictor, that is geopotential height at 500 HPA.

So, like that is basically it is like we know that, the pressure varies with altitude. So, a particular pressure say 500 hectopascal at what altitude will that be attained. This is the variable. So, it is like on an average this is 5.5 kilometers altitude, but it varies, from location to location and also from time to time. So, that is called the geopotential height. It can fall to maybe 5 kilometers or rise to 6 kilometers and so on.

(Refer Slide Time: 07:05)



So, the this is the paper where the analog forecasting of extreme causing weather patterns using deep learning.

So, note the choice of words in the heading, extreme causing weather patterns. So, extreme causing so, they are assuming that there are certain weather patterns, which cause extremes. So, and those weather patterns are mostly related to this geopotential height as they have considered in this paper. Numerical weather prediction models require ever growing computing times and resources, but still, sometimes have difficulties with predicting extreme weathers.

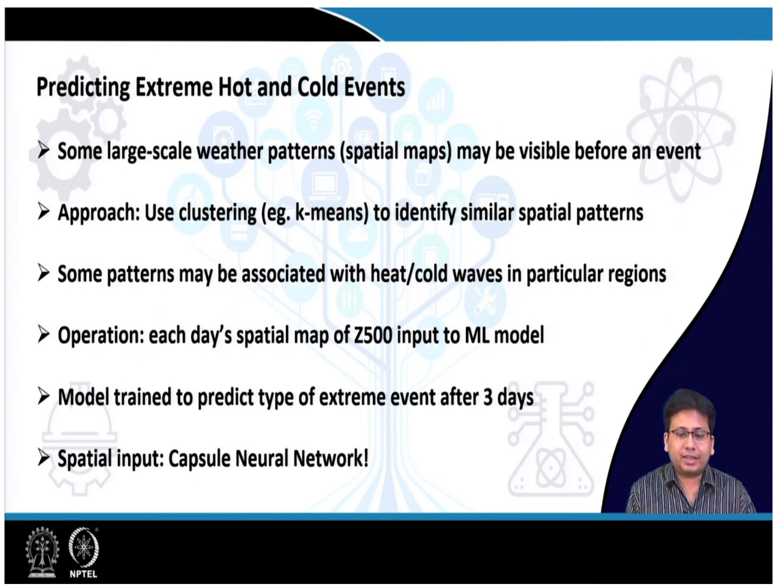
We introduce a data driven framework, that is based on analog forecasting or predicting using similar past similar patterns and employs a novel deep learning, pattern recognition technique, called capsule neural networks and an impact based auto labeling strategy.

Using data from a large ensemble, fully coupled earth system model, CapsNets are trained on mid tropospheric large scale circulation patterns of Z500, Z500 basically talks means the geopotential height of at hectopascals labeled 0 to 4, depending on the existence and geographical region of surfaced temperature extremes over North America, several days ahead like we will explain this in the coming slides.

The trained network, predict the occurrence or regions of cold or heat waves using only Z500 with accuracies of like these numbers, which are reasonably high 1 to 5 days ahead using both surface temperature and Z500 accuracy increased to 80 percent in both cases CapsNets, outperform simpler techniques like convolutional neural networks and logistic regression and their accuracy is least affected as the size of training set is reduced.

The results show the promises of multivariate data driven frameworks for accurate and fast extreme weather predictions, which can potentially augment numerical weather prediction efforts in providing early warnings.

(Refer Slide Time: 09:29)

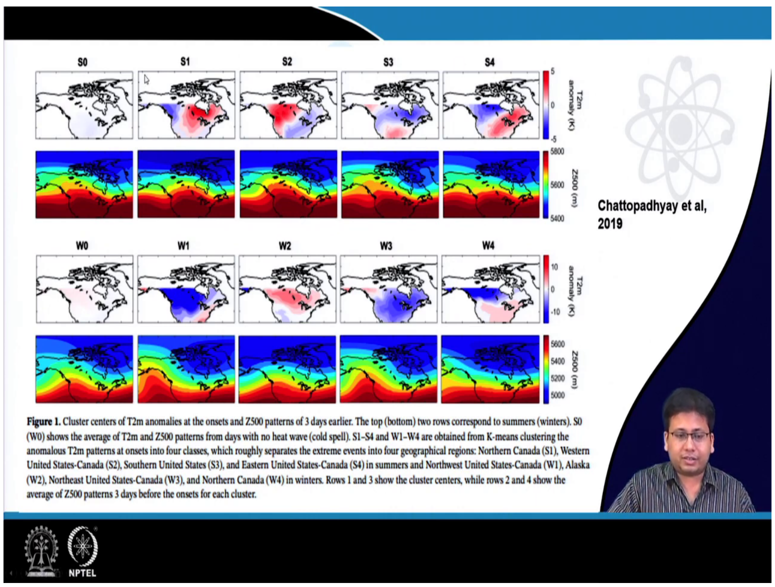


So, the basic approach of this paper is as follows there so, the idea is that there are some large scale weather patterns, that is the special maps of variables, which may be visible before an event. Before an event means like in this case, 3 to 5 days before the event.

The approach here first of all use clustering or k means, to identify similar spatial patterns. So, this is actually used this clustering is not part of their of the prediction algorithm, but this is used for labeling purposes. So, like there is a huge amount of data, which they have to label that is they have obtained the data set basically from some kind of model simulations.

So, that is a huge volume of data and if, but for training any prediction algorithm, you need labeled data. So, like since we cannot expect a human being to sit and label. So, many like so many pieces of data. So, what they have done is they have used auto labeling strategy, which means they first do the clustering of the data the clustering is so, I will show you like the clusters.

(Refer Slide Time: 10:45)



So, these are the examples of the clusters. So, you can see like it is a the clustering is done of both the air temperature as well as the geopotential heights. So, this is what the like if you do the clustering this is what; so, this is actually the temperature not the temperature itself, but the temperature anomaly. So, here you can see that each cluster is basically some kind of a spatial distribution of the temperature anomaly.

So, here you can see, like there is a positive strong positive temperature anomaly, in this region. In the like in the you can say the what is known as the Midwest region, while in the this regions there is a there seems to be a like a cold wave kind of thing. In this the this is the second cluster, where we see the strong positive anomaly in this region. In S3, there are we see the weak anomalies mostly, while in S4 we see a strong and positive anomaly along the eastern coast.

While in there is one more pattern, which does not show any significant anomaly in any place.

And similarly, if we do the clustering of the geopotential heights, there also we see like of course, these all these patterns are roughly the similar unlike in this case, but there are subtle differences also; that is as you can like if you look at the maps, you can understand the differences of these contours. And so, this kind of clustering is actually used; so, like this is in summer and this is in winter. So, these clusterings are used for prediction of heat waves and these ones for the prediction of cold waves.

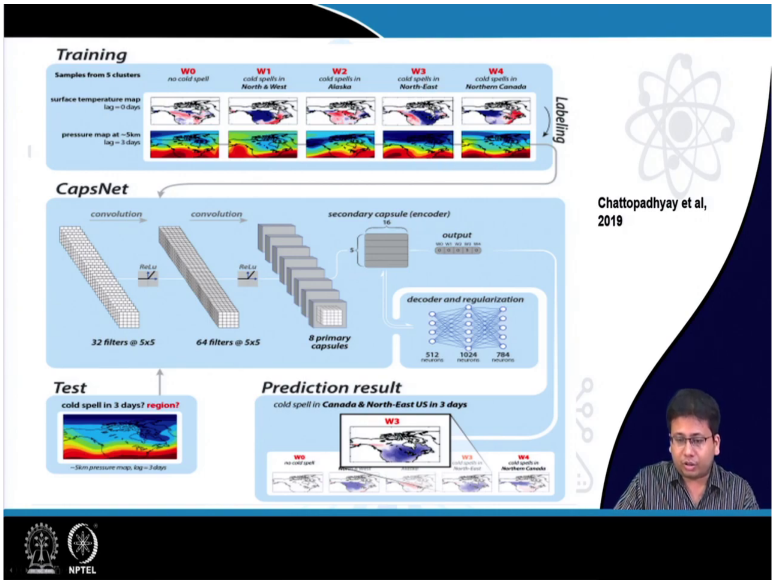
So, these clusters are created for the purpose of labeling that is a like every spatial map is to be labeled as like this is going to be like this is a like this is a heat wave in this part of the country this is a heat wave in that part of the country or this is not a heat wave at all and so on. So, one now some of the like some of these patterns are we which we obtain by this kind of claim is clustering, are associated with heat or cold waves in particular regions; as we just saw in the examples.

Now the operation of the prediction algorithm is as follows. So, every day's spatial map of Z500 is presented as an input to the machine learning model. In some cases along with Z500 the that days air temperature also may be presented as an input.

And the what the model is going to predict is, whether there will be an extreme event after 3 days and if so, in which region. I mean not point wise, I mean not at exactly in this city there is going to be a heat wave some not something like that, but in general, if is it the northeastern region or is it the south central region or which region the broad region, that they can predict.

And the spatial input is presented through the machine learning model. So, the machine learning model should be such a model, which is capable of taking in the spatial input. So, like we had discussed about convolutional neural networks for spatial models earlier. So, in this case they use a more sophisticated version of that the Capsule Neural Network.

(Refer Slide Time: 14:13)

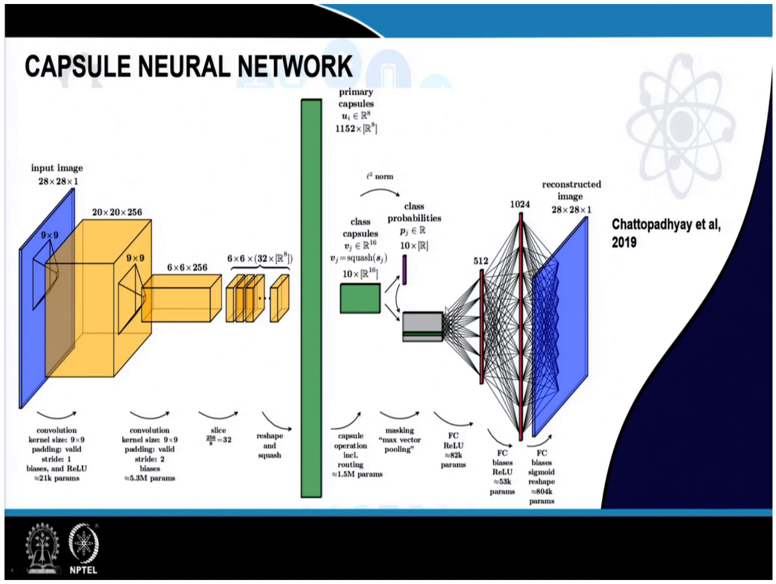


So, like just to illustrate this. So, this is the training purpose, here they have shown the labeling, where like basically they have the all the maps the other of the I mean the spatial maps of the different variables obtained from either the observations or simulations. So, for labeling purposes, they have utilized this clustering technique which we mentioned.

So, that way they get this kind of the label data, that is every map is accompanied by a label, which means like whether it is followed by a heat wave or not and if a heat wave then a heat wave in which region. Now, this kind of labeled data is presented as a CapsNet, which is used to present the CapsNet. So, the input is the spatial map of a like of the geo potential height and the prediction is going to be like any of these 5 classes right.

So, like this is the test input the geopotential height map and the what you like the question that is being asked is can will there be a hot spell or cold spell in 3 days. And if so, in which region and this is what the output is going to be like that the it seems that this pattern is a precursor to a W3 kind of event; W3 basically means a cold wave in the northeastern part of the country.

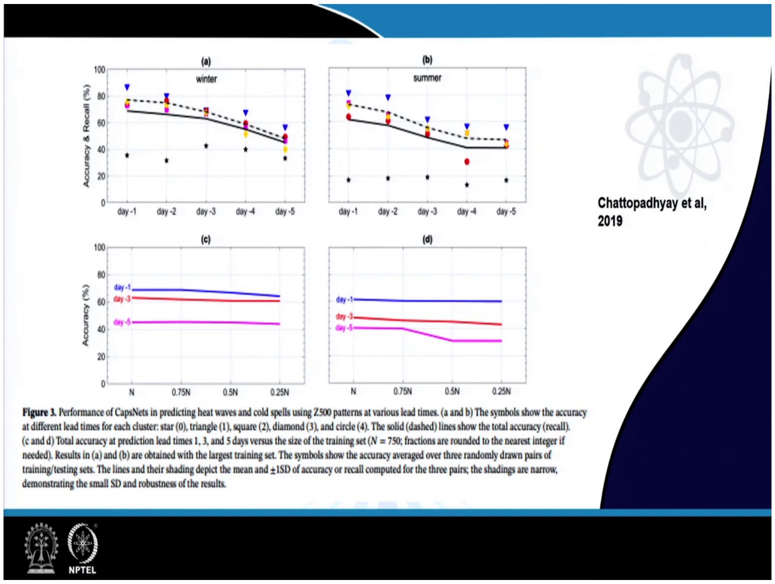
(Refer Slide Time: 15:39)



So, now what is this Capsule Neural Network? So, the capsule neural network is a like a special kind of convolutional neural network, which has all the components of the convolution such as the convolutional layer, the pooling layer etcetera; however, like all these convolutional input units, these are basically these are used to develop what are known as the capsules.

And these capsules are then like there is a hierarchy of capsules like that and all those capsules are arranged in a particular way and based on the arrangement of those capsules finally, the class is the class prediction is made.

(Refer Slide Time: 16:24)



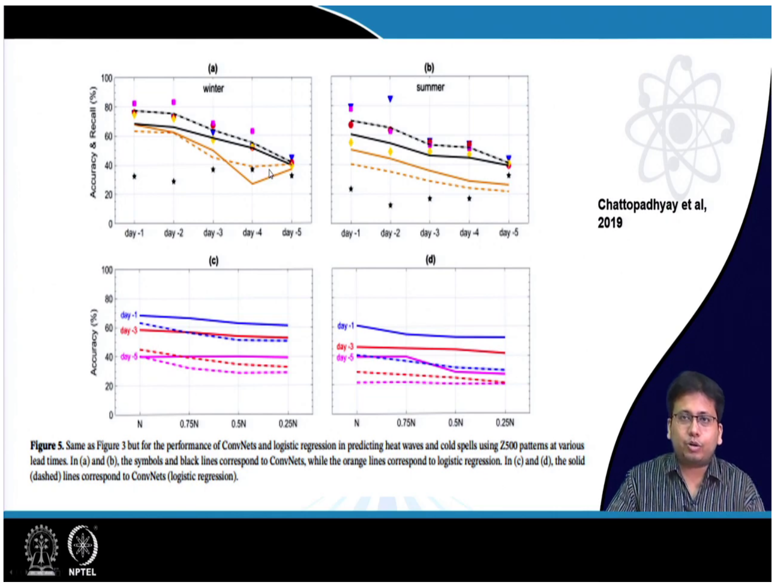
The reason why people use capsule neural networks is especially to preserve the spatial alignment, which unfortunately convolutional neural networks are not always used to like are not able to identify. Like in case, like suppose there is an input is presented in front of a convolutional neural network, which has all the features, but not in a specific spatial local layout. Say like for example, like when we consider a humans face, it is expected that the like the eyes will be on here then followed by the nose, followed by the mouth and so on.

But if there is a like if an image of a human face with the eye a one eye here, one eye here the nose on top etcetera I mean basically all the features are present all the, but they are misaligned the convolutional neural network will still think that it is a human face; which of course, it is not. So, this kind of a problem; however, can be can be avoided with a capsule neural network, which actually defines some kind of a spatial ordering of the different features that are being that are being represented by the different layers of the neural network.

So, like this is the these are the results which the prediction results which they get the accuracy as well as the recall and they show the accuracy at like different days ahead of time.

So, like; obviously, we can say that one day before the heat wave or cold wave, we can expect higher accuracy then after 2 days ahead we can expect slightly lower accuracy and so on. So, this is the that is the effect we see here, but we show that like in the case like when there is capsule neural networks are used, then the like in all days or at all lead values, we find that the accuracy is highest, compared to other approaches like say convolutional neural network or linear regression. Not sorry logistic regression.

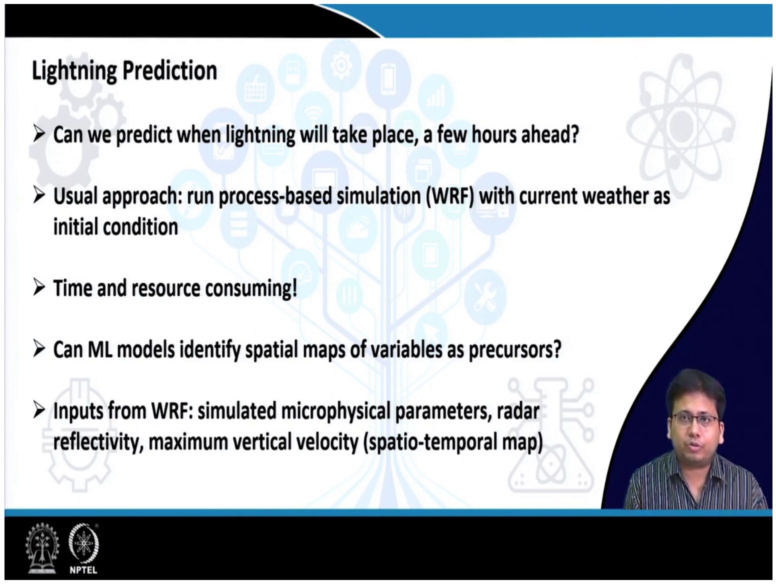
(Refer Slide Time: 18:29)



So, since this is a classification problem, we can also use non deep methods such as logistic regression, but we find that its prediction is not so great; and the reason for that; obviously, is that the this neural this convolutional or capsule neural network they are able to take the spatial map as input and it is actually able to represent the spatial relations, which something like logistic regression, which considers the whole thing as just as a long vector is unable to do that.

Not only that, it is found that the capsule neural network also does the best with least amount of training. So, like training examples; even if you use the only 25 percent of the training samples, then also its performance is better than the others.

(Refer Slide Time: 19:27)

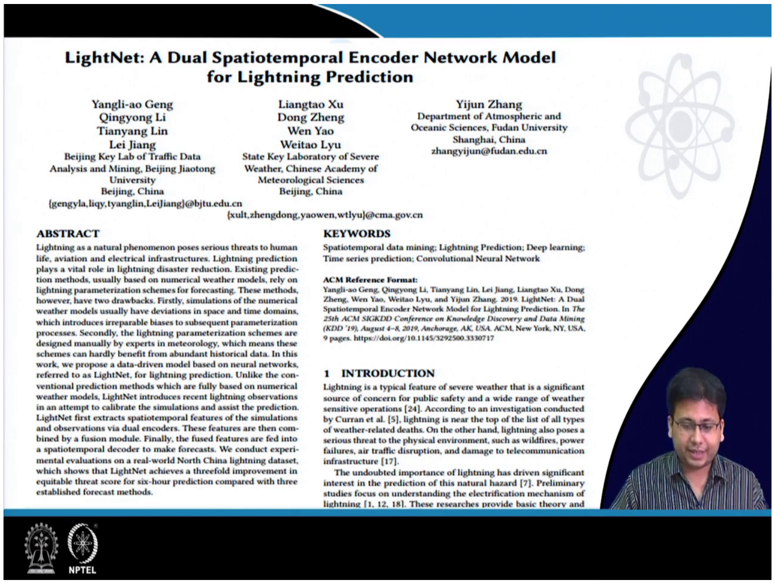


So, that is, so this is one of the papers we are discussing, the second one is related to as I said earlier, lightning prediction. So, the question here is can we predict when lightning is going to take place a few hours ahead of time. And the usual approach to do that the way meteorologists have been dealing is run process based simulation. So, there is a WRF model, it is a numerical weather prediction model, which is also known as NWT.

Like there is a set of neural like numerical equations, which it has to solve at various grid scales to do the simulation. So, it takes like at any given point, the current weather is pro including all the different variables is provided as the initial condition and it is allowed to run for a specific duration of time, and then it produces the like the outputs. Now the problem is that this is time and resource consuming, but can the machine learning models identify the spatial maps of variables as precursors. So, just like in the previous case.

So, like in this case, we will need inputs from the WRF itself, as well as will you will need some observational inputs. So, like and all these inputs are going to be in the form of spatio temporal map not just spatial maps, but also spatio temporal maps.

(Refer Slide Time: 20:45)



In the previous case, we have we just had one single map, but in this case we will need a series of or a sequence of such maps. So, lightning is a natural phenomena, poses serious threats to human life aviation and electrical infrastructure. Lightning prediction plays a vital role in lightning disaster reduction. Existing prediction methods usually based on numerical weather models rely on lightning parameterization schemes for forecasting.

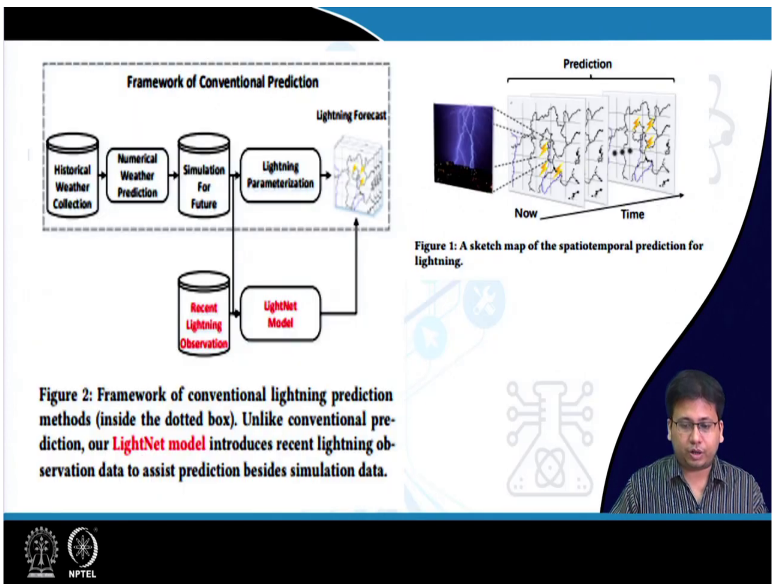
So, it is the lightning is not directly forecasted by the models like WRF, but certain enabling conditions are known due to like I mean some other variables are known, which are considered to be the immediate triggers to lightning. So, when we are running the NWP, we keep track of certain variables and whenever those variables like they reach some specific values, like for the and they are combined using some kind of a parameterized relation. Then we understand that a lightning is likely to be triggered.

So, these methods; however, have 2 drawbacks. Firstly, simulations of the numerical weather models usually have deviations in space and time domains which introduces irreparable biases to subsequent parameterization processes. Secondly, the lightning parameterization schemes are designed manually by experts, in meteorology, which means these schemes hardly benefit from abundant historical data. In this work, we propose a data driven model, based on neural networks referred to as LightNet for lightning prediction.

Unlike the conventional prediction methods, which are fully based on numerical weather model, LightNet introduces recent lightning observations, in an attempt to calibrate the simulations and assist the predictions. LightNet first extracts spatiotemporal features of the simulations and observations via dual encoders.

These features are then combined by a fusion module. Finally, the fused features are fed in by into a spatiotemporal decoder to make forecasts. We conduct experimental evaluations on a real world, North China lightning dataset, which shows that LightNet achieves a threefold improvement in equitable threat score for 6 hour prediction, compared to 3 established forecast methods.

(Refer Slide Time: 23:25)



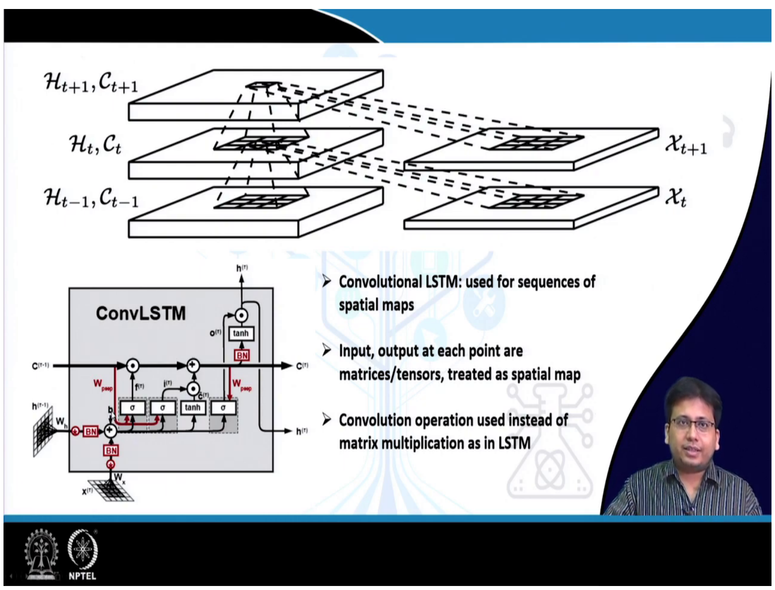
So, like when we are doing conventional lightning prediction, this is the way it happens. So, first of all like there is the database of historical weather collections, based on that numerical weather prediction is initialized and it like it is used to run the simulation for the future say in the next 6 hours or something like that.

Now, as I said, when that as the simulation runs, certain variables, which are known to be triggers of lightning and known I mean known to experts, those variables are tracked and like based on them there is some kind of lightning parameterization, which on the basis of those variables like or keeps calculating the probability of lightning at any given location and time.

Now, the way this paper, hopes to improve is by bringing in recent lightning observations along with the similar simulations of these various geophysical variables and developing this kind of a LightNet model, which like which in a sense is able to bypass the simulation I mean the simulation cannot be completely done away with, but like the this LightNet model, it will in a sense it will keep receiving the inputs from the simulations and as well as the past lightning data and it will be making the predictions.

So, the need of this parameterization is largely done away with. Now, now as I said earlier, in just like the previous paper, in this case also the inputs are going to be spatial maps of the different variables, which are obtained from the simulations as well as the historical distributions of the lightning.

(Refer Slide Time: 25:03)



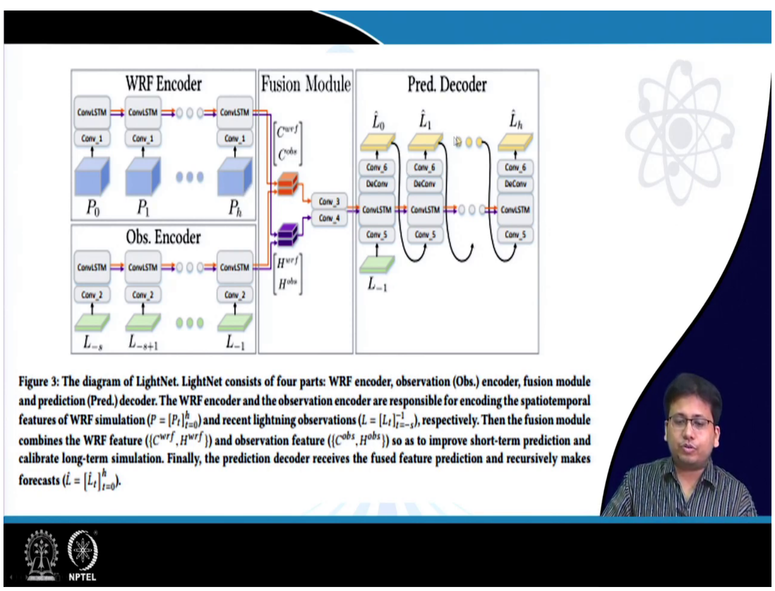
So, the these; so, once again we have the spatial maps as input. So, we will need some kind of convolution at the input. And the outputs are also going to be like this spatial maps, because unlike in the previous paper, where the predictions were like just a few discrete values like in fact, only 5 discrete values.

In this case, we actually the out want that output as a spatial map, where we want to provide at every location at every point, I want to give a predict probability of that a lightning may take place there ok. So, like both at both the input and the output, we need convolution and besides the at the input, we have a sequence of spatial maps coming in I mean as the simulation is running, in the WRF those spatial maps of the different variables those are also coming in at every point of time.

So, we have a sequence of spatial maps as input, unlike the previous case where we had only a single like spatial map as input we have a sequence of spatial maps as input. So, we need something like a ConvLSTM, LSTM for the temporal part and a convolution for the spatial part. So, ConvLSTM is like it is a method that was developed in 2015.

So, the idea is that it is such an LSTM where the inputs are not usual vectors as in a normal LSTM, but they are matrices or tensors and the like whenever in a normal LSTM, whenever there is the usual matrix or vector multiplication; that is, replaced by these convolution operation.

(Refer Slide Time: 27:06)

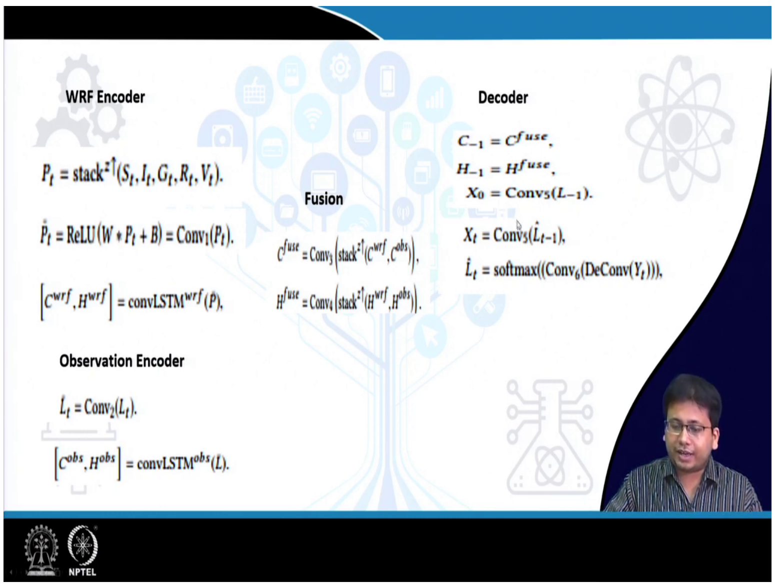


So, this is the scheme of the whole thing. So, there is a WRF encoder, which encodes the, which receives in the spatial maps from the WRF simulation and there is an observation encoder, which receives the like the lightning distribution as I mean the spatial map of lightning as the inputs.

So, the these two things are separately encoded by two separate networks, and then the encoded versions are somehow fused using a fusion module.

So, this is like basically some kind of a concatenation operation followed by convolutions. Now, so, finally, what we get at the is an encoded form of the past lightning as well as the simulations of the different weather variables; and these are then finally, used for the decoding purposes.

(Refer Slide Time: 28:01)



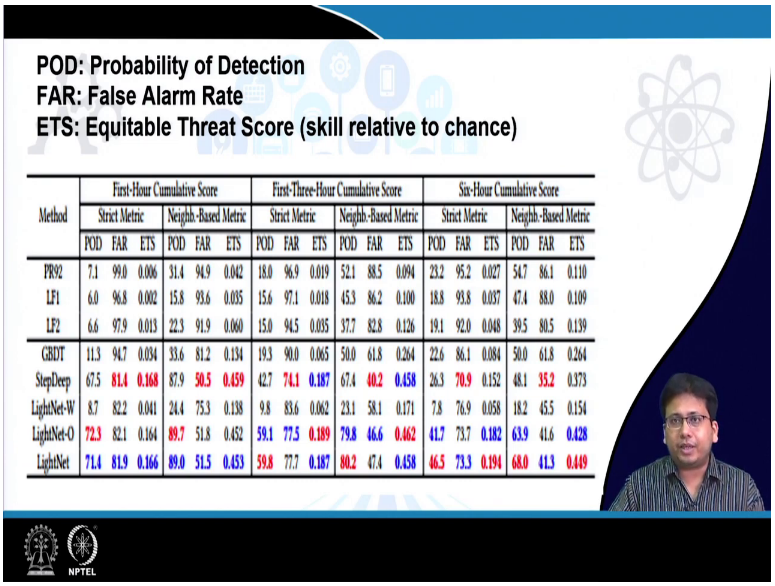
So, like at both the encoder at the fusion layer as well as the decoder, the convolution is used quite heavily as you can see. So, like, so, these S_t, I_t, G_t, R_t, V_t. So, these are basically the inputs which we get from the WRF at any given time. So, they like they are stacked together to form P_t like combined form and then this like they are passed through convolution of a convolution layer and then followed by a ConvLSTM.

So, we have a sequence of these things. So, they like they are present into the LSTM and; so the output of the LSTM is the C^{wrf} and H^{wrf}. So, it is like in a sense it is like the encoded hidden state of the sequence of maps that we receive from the WRF. Similar steps are done in the observation encoder also. So, from there also we get the these C and H maps. So, C and H for our

LSTM of course, refers to the cell state, I mean the long term and the short term memory as we have discussed in the lecture on LSTM or the sequential models.

And then, there is a fusion layer in which all these the these things are fused they are stacked together, concatenated together followed by convolution; and then the this thing is used as input to the decoder, which once again utilizes convolution to produce the output.

(Refer Slide Time: 29:32)

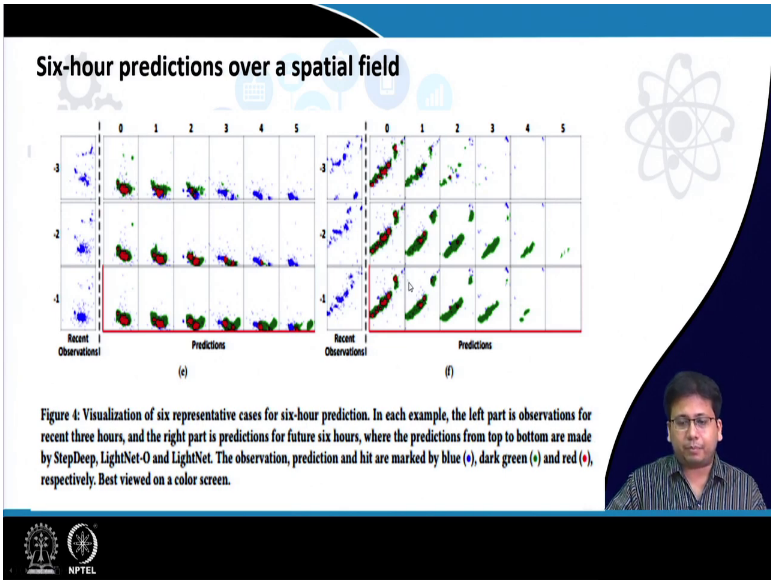


Method	First-Hour Cumulative Score						First-Three-Hour Cumulative Score						Six-Hour Cumulative Score					
	Strict Metric			Neighb.-Based Metric			Strict Metric			Neighb.-Based Metric			Strict Metric			Neighb.-Based Metric		
	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS
PR92	7.1	99.0	0.006	31.4	94.9	0.042	18.0	96.9	0.019	52.1	88.5	0.094	23.2	95.2	0.027	54.7	86.1	0.110
LF1	6.0	96.8	0.002	15.8	93.6	0.035	15.6	97.1	0.018	45.3	86.2	0.100	18.8	93.8	0.037	47.4	88.0	0.109
LF2	6.6	97.9	0.013	22.3	91.9	0.060	15.0	94.5	0.035	37.7	82.8	0.126	19.1	92.0	0.048	39.5	80.5	0.139
GBDT	11.3	94.7	0.034	33.6	81.2	0.134	19.3	90.0	0.065	50.0	61.8	0.264	22.6	86.1	0.084	50.0	61.8	0.264
StepDeep	67.5	81.4	0.168	87.9	50.5	0.459	42.7	74.1	0.187	67.4	40.2	0.458	26.3	70.9	0.152	48.1	35.2	0.373
LightNet-W	8.7	82.2	0.041	24.4	75.3	0.138	9.8	83.6	0.062	23.1	58.1	0.171	7.8	76.9	0.058	18.2	45.5	0.154
LightNet-O	72.3	82.1	0.164	89.7	51.8	0.452	59.1	77.5	0.189	79.8	46.6	0.462	41.7	73.7	0.182	63.9	41.6	0.428
LightNet	71.4	81.9	0.166	89.0	51.5	0.453	59.8	77.7	0.187	80.2	47.4	0.458	46.5	73.3	0.194	68.0	41.3	0.449

So, like these are the results which of this so, like when we are forecast making point wise forecast of something like lightning, it the fore or rather nowcast, it is usually evaluated using a certain metrics like POD the probability of detection that is suppose lightning actually happens.

What is the probability that I will be able to for nowcast it that is POD. Next is false alarm rate, which is of course, self explanatory; and there is then there is ETS or equitable threat score. So, like when I am successful in making a prediction it might have happened by chance also. So, can I somehow estimate how much better I am doing rather than just random predictions.

(Refer Slide Time: 30:18)



And so, like this is what the predictions may look like. So, like we can say that 3 hours ahead of time, these are some of the like the recent observations, of the lightning. So, these blue point these are like the distribution of lightning events at different locations and then based on that, in the coming hours, we make a map.

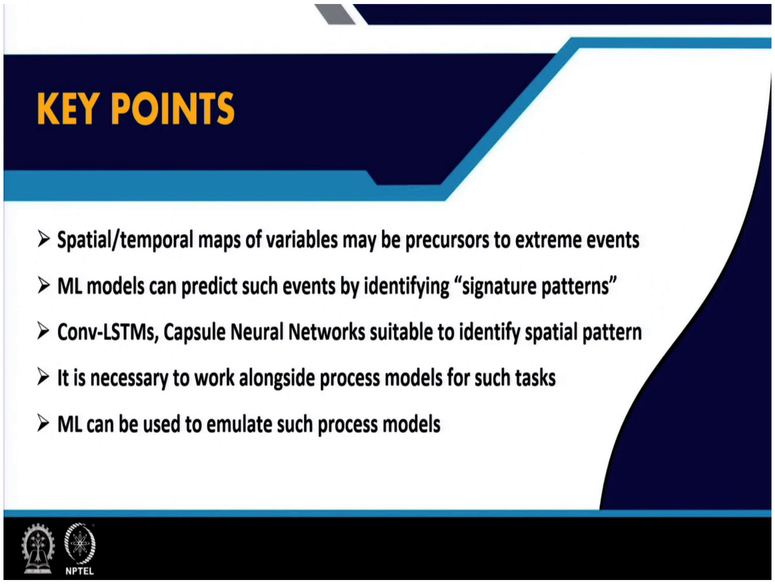
So, at every location, we actually plot the probabilities of the lightning happening or not and; so, like the what these different colors indicate is the observation prediction and the hits though the blue means like lightning is observed, green means it is predicted and red means hit means both observed as well as predicted.

So, as you can see that at least for the first 2-3 hours, there is a significant number of hits over the spatial field.

(Refer Slide Time: 31:17)

So, these are the different papers which we discussed today, this is the first paper and about the heat wave prediction and this is the second one about the LightNet and for ConvLSTM which this is here is the paper.

(Refer Slide Time: 31:34)



So, the key points from this lecture are first of all the spatial and temporal maps of variables, may be precursors to extreme events. Secondly, the machine learning models can predict such events by identifying certain signature patterns and Conv-LSTMs or Capsule Neural Networks, these are suitable for to identify like the spatial patterns so, of course, because the inputs are spatial maps.

So, we need convolution, but in both cases we are going beyond convolution, in one case we are going capsule neural network, which like consider the spatial ordering and then there is convolution LSTM where we have which are for receiving a sequence of spatial maps. And it is important to work alongside the process models such as the WRF in such cases and like the machine learning can also be used to emulate this kind of process models, this is again a point which we will see in greater detail, in the coming slides.

So, that brings us to the end of this lecture, we are nearing the end of the current module, which and in the coming lecture will be the last of this module 3. So, till then bye.