An Approach to Predict Glacial Lake Outburst Flood

In partial fulfillment of the requirements for the degree of Master of Technology in Computer Science

Submitted by
Partha Kayal (CS2037)
M.Tech.-CS

under the supervision of
Dr. Sarbani Palit

INDIAN STATISTICAL INSTITUTE
203, Barrackpore Trunk Rd
Kolkata-700108

Declaration

I hereby declare that the work presented in the project report entitled *An Approach to Predict Glacial Lake Outburst Flood* contains my work under the supervision of Dr. Sarbani Palit, ISI Kolkata. At places where ideas are borrowed from other sources, proper references, as applicable, have been cited. To the best of my knowledge, this work does not emanate from or resemble other work created by persons other than mentioned herein.

Partha Kayal
Roll:CS2037, M.Tech CS
ISI,Kolkata
Date: 05/07/2022

Indian Statistical Institute

Certificate of Approval

I hereby certify that the dissertation thesis entitled *An Approach to Predict Glacial Lake Outburst Flood* submitted by **Partha Kayal** in partial fulfillment for the award of the degree of Master of Technology in Computer Science is completed under my supervision.

Sarbani Palit

Dr. Sarbani Palit
Associate Professor, CVPR Unit
Indian Statistical Institute, Kolkata

ABSTRACT: Remote sensing data is a rich resource of information, as it provides a time-wise sequence of data, and therefore can be used for prediction purposes. In this paper, we addressed the challenge of using time series on satellite images to predict the Glacial Lake Outburst Flood(GLOF). In order to predict GLOF, we proposed two-step approach. In the first step, our aim is to extract the pixel-wise information about water, snow, and soil at different time stamps and prepare them for use in the training input. The second step we use is Long Short Term Memory (LSTM) network in order to learn temporal features and thus predict the future pixel value of water, snow, and soil.

Keywords: Glacial Lake Outburst Flood(GLOF),Normalized Difference Water Index(NDWI), Normalized Difference Snow Index(NDSI) , Normalized Difference Soil Index(NDSI), LSTM.

Contents

List of Tables

List of Figures

1 Introduction

World climate has been changing in the past hundred years but the last 5 decades have seen this change at an accelerated rate. For this reason, the temperature has been rising at a high rate. As a result, we see that glacial ices are melting very fast. A glacial is defined as a reserve of frozen water which is a valuable source of fresh water. Melting glaciers increase the water level and reduce the snow cover and soil in the upper level[6]. International Centre for Integrated Mountain Development (ICIMOD) and Asia-Pacific Network for Global Change Research (APN) jointly submitted a report[9] in 2005 and this reported more than 5218 glacial lakes in the Himalaya Mountains of India, Pakistan, and Tibet Autonomous Region. The lakes that formed from melting glacial ice are considered glacial lakes.

Expanding the areas of the lakes increases the outburst floods. A glacial lake outburst is a sudden increase in water flow in the downstream direction of a glacial lake, along with excessive sediment deposition. In the past few decades, we observed that the number of glacial lake outburst floods is increasing. As a result, there has been a loss of life and property due to these events. Current research's aim is to link climate change and glacial ice melting activity. In our work, we have attempted to predict glacial lake outburst flood based on satellite image data.

In the last few decades, we have seen that the use of time series forecasting has been increasing day by day. Remote sensing data is a rich resource of information, as it provides a timewise sequence of data, and therefore can be used for prediction purposes.

Figure 1: Chorabari Lake before disaster,Source:Google

Figure 2: Chorabari Lake after disaster,Source:Google

2 Literature Review

Remote sensing data have been widely used in order to predict the changes in land cover classification, urban expansion, and changes in forest structure across various geographical locations.

The literature review suggests various Supervised and unsupervised classification methods for lake identification. Huggel et al.[8] proposed a method that can detect glacier lakes from remote sensing data. In his work he suggested the Normalized Difference Water Index(NDWI), to automatically detect glacial lakes. Watanabe et al. (1994) evaluated the potential of GLOF in the Himalayan region based on changes in the water volume of the lake. Sonam Rinzin,Guoqing Zhang, and Sonam Wangchuk[12] reported that the glacial lakes area has expanded at a rate of 0.96km^2/year in Bhutan from 2016 to 2020.

Georg Veha et al(2018)[7] proposed a method that can detect GLOFs in the past. They used the "Change point Algorithm" to find the probability of changing pixel value from water to sediment. They took shrinking lakes and sediment tails downstream as key indicators for detecting GLOFs.

S. N. Remya et al.(2019)[13] proposed a method that can estimate the volume , area and depth of the glacial lake and also to predict further expansion of the glacial lakes.

Most of the work done in the past is for predicting glacier outburst in the past but there is negligible work done for predicting future glacier outburst which is the objective for our work.

By investigating the literature, we note that few research works have been conducted about GLOF. Since the LSTM network has the ability to remember long time past information, we can use it in order to predict such kinds of natural phenomena.

3 Preliminaries

3.1 Landsat data:

Landsat data is high-quality, multi-spectral imagery of the surface of the Earth, captured by the Landsat satellite. Landsat images have several bands based on their wavelength. A band represents a segment of the electromagnetic spectrum. Total number of bands in Landsat7 and Landsat8 images are 8 and 11 based on their wavelength. The details of Landsat7 and Landsat8 images are given in the below table.

S.N.	Landsat 7 Enhanced Thematic Mappers Plus (ETM +)			Band	Landsat 8 Operational Land Imagers (OLI) & Thermal Infrared Sensor (TIRS)		
	Resolution (meter)	**Wavelength (micrometer)**	**Band Name**		**Band Name**	**Wavelength (micrometers)**	**Resolution (meter)**
1	30	0.45-0.52	Blue	**Band 1**	Ultra-Blue	0.435-0.451	30
2	30	0.52-0.60	Green	**Band 2**	Blue	0.452-0.512	30
3	30	0.63-0.69	Red	**Band 3**	Green	0.533-0.590	30
4	30	0.77-0.90	NIR	**Band 4**	Red	0.636-0.673	30
5	30	1.55-1.75	SWIR 1	**Band 5**	NIR	0.851-0.879	30
6	60* (30)	10.40-12.50	Thermal	**Band 6**	SWIR 1	1.566-1.651	30
7	30	2.09-2.35	SWIR 2	**Band 7**	SWIR 2	2.107-2.294	30
8	15	0.52-0.90	Panchromatic	**Band 8**	Panchromatic	0.503-0.676	15
9				**Band 9**	Cirrus	1.363-1.384	30
10				**Band 10**	TIRS 1	10.60-11.19	100 * (30)
11				**Band 11**	TIRS 2	11.50-12.51	100 * (30)

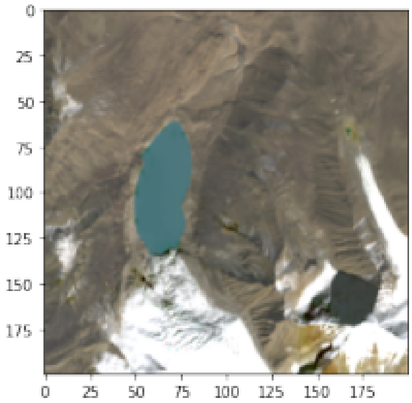
Table 1: Landsat 7 and Landsat 8 data's bands,resolution, wavelength [Source:Google]

3.2 Normalize Difference Index

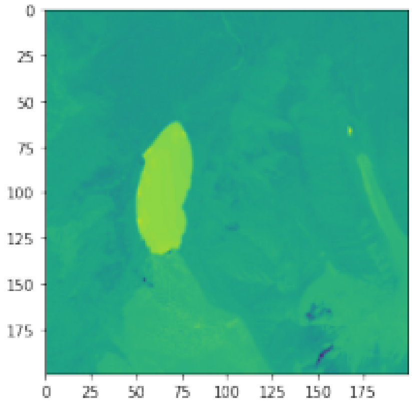
- **Normalized Difference Water Index(NDWI):** NDWI was proposed by Gao[3] in 1996 and is used for analysing the water bodies. It is defined by:

  NDWI=(NIR-SWIR)/(NIR+SWIR)

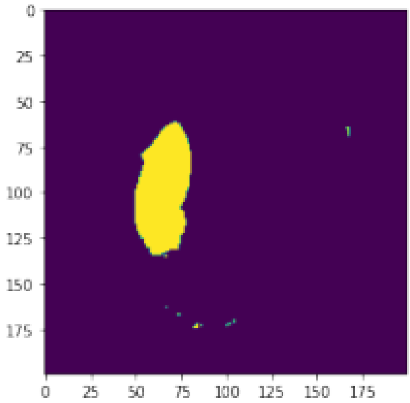
  This index is designed to maximize the water reflection by using the NIR spectrum and minimizing the low reflection of water by using the SWIR spectrum. To highlight the presence of water we used NDWI threshold greater than 0. If the NDWI value is greater than 0 then we get a Boolean value True for the water pixel and otherwise we get False for the water pixel.



(a)



(b)



(c)

Figure 3: Lemthang Tsho Lake, Bhutan(a) RGB image, (b) NDWI image, (c) binary image

- **Normalized Difference Snow Index(NDSI):**NDSI was proposed by Hall et al[4] in 2001 is used for mapping snow. It is defined by:

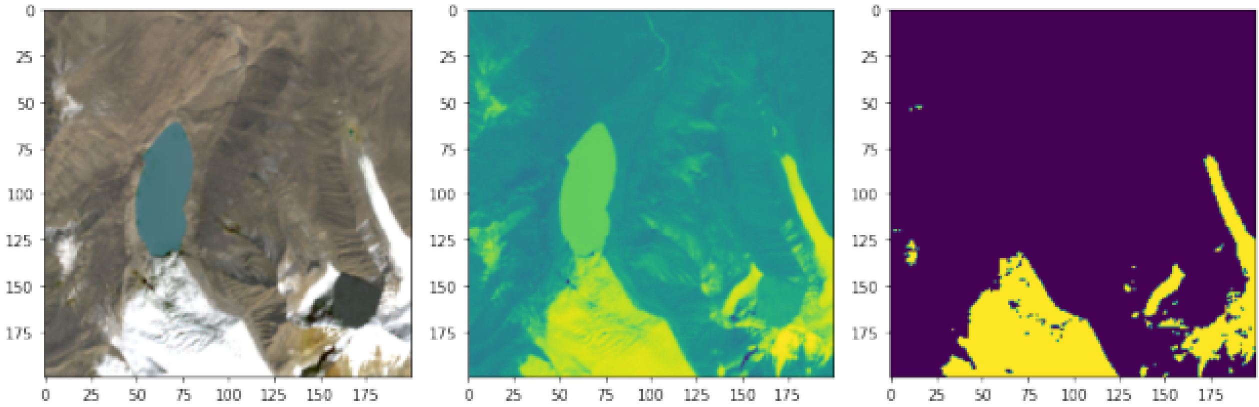
  NDSI=(GREEN-SWIR)/(GREEN+SWIR)

  Because snow is highly reflective in the GREEN spectrum and highly absorptive in SWIR spectrum. To highlight the presence of snow we used NDSI threshold value 0.4 . If the NDSI value is greater than the threshold then we get a Boolean value True for the snow pixel and otherwise we get False for the snow pixel.

- **Normalized Difference Soil Index(NDSI):**NDSI is used for mapping soil. It is defined by:

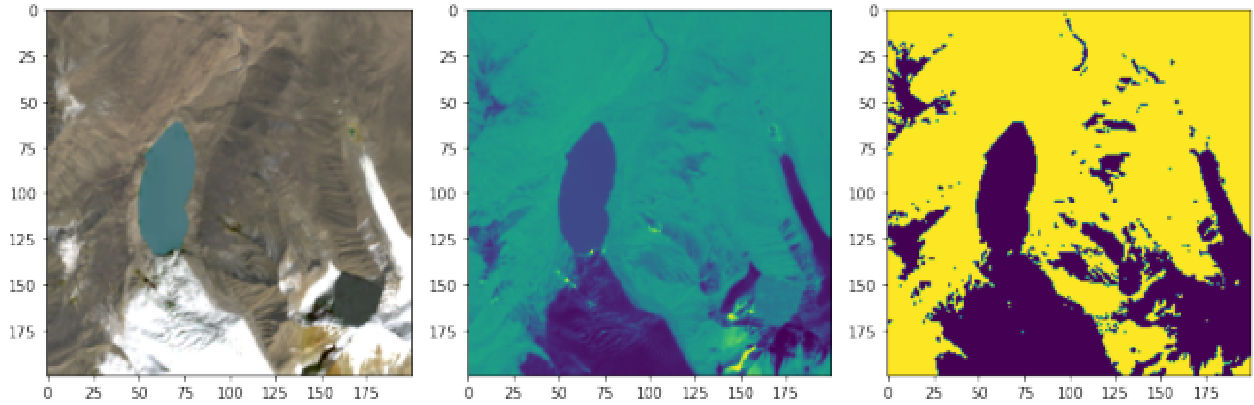
  NDSI=(SWIR-NIR)/(SWIR+NIR)

  Because soil is highly reflective in the SWIR spectrum and highly absorptive in the NIR spectrum.To highlight the presence of soil we use threshold value 0.1 . If the NDSI value is greater than 0 then we get a Boolean value True for soil pixel and otherwise we get False for soil pixel.



(a) (b) (c)

Figure 4: (a) RGB image, (b)Normalized Difference Snow Index image, (c) binary image for snow



(a) (b) (c)

Figure 5: (a) RGB image, (b) Normalized Difference Soil Index image, (c) binary image for soil

Index	Name	Landsat7	Landsat8
NDWI	Normalized Difference Water Index	(Band 4 − Band 5) / (Band 4 + Band 5)	(Band 5 − Band 6) / (Band 5 + Band 6)
NDSI	Normalized Difference Snow Index	(Band 2 − Band 5) / (Band 2 + Band 5)	(Band 3 − Band 6) / (Band 3 + Band 6)
NDSI	Normalized Difference Soil Index	(Band 5 − Band 4) / (Band 5 + Band 4)	(Band 6 − Band 5) / (Band 6 + Band 5)

Table 2: Formula for Normalized Difference Index

3.3 LSTM

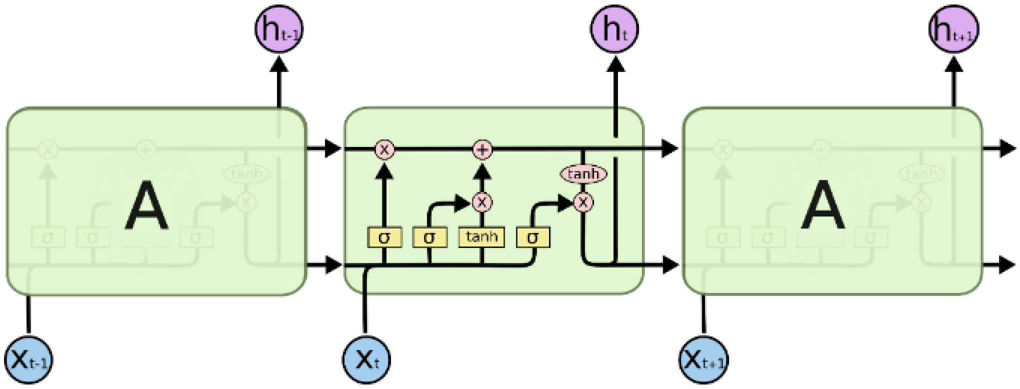
A recurrent Neural Network (RNN) is a type of neural network which contains memory and enables us to model sequential data. RNN suffers from short-term memory. If a sequence is too long then the network will not retain information from timesteps that are too old. RNN also suffers from vanishing gradient problems that can occur during the training. LSTM was created as the solution to short-term memory problems. They have internal mechanisms called gates that can regulate the flow of the information which enables the network to retain long-term memory.

LSTM was introduced by Sepp Hochreiter[5] in 1997. Every LSTM unit has three input-

- x_t−input at timestamp t
- c_{t-1}−cell state at timestamp $t-1$
- h_{t-1}−hidden state at timestamp $t-1$

and two output−

- c_t− current cell state
- h_t− updated hidden state



LSTM Architecture, Source:Colah Blog

Figure 6: LSTM has three gates Forget gate, Input gate, Output gate.

- First part in LSTM network is decide which information coming from the previous timestamp is going to throw away from the cell state . This decision is taken by a sigmoid function which is known as Forget gate.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{1}$$

  Therefore the value of f_t is a number between 1 and 0.

  If $f_t = 0$, then $c_{t-1} * f_t = 0$ completely forget the information from the previous timestamp.

  If $f_t = 1$, then $c_{t-1} * f_t = 1$ then keep whole information from the previous timestamp.

- Next part decided what information we store into the cell state c_t.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{c}_t = \tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{3}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{4}$$

  The equation(2) is known as Input gate .As it is use sigmoid layer so the value of the input gate i_t is a number between 0 and 1. 0 means the input value is not important and 1 means the input value is important

As tanh activation function squishes values to always be between -1 and 1 so the equation(3) squishes the input value between -1 and 1.

Now we go to the update the cell state.First we point wise multiplied the forget vector with cell state c_{t-1} with forget vector f_t.Then we do pointwise addition with output from the input gate.

- Finally we decide what information is going for output.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(c_t) \tag{6}$$

Equation(5) is known as output gate.Due to sigmoid function the value of o_t is a number between 0 and 1 and this gate decides what should be the next hidden state .

4 Dataset

To analyze the land cover classification remote sensing data is one of the most important resources and it is used on a global scale. Remote sensing data have been widely used in various fields like Glacier Lake Outburst Flood and Hazard Assessment, predicting the changes of Forest structure and vegetation, detecting urban growth etc at various geographical locations. In this work we use Landsat7 and Landsat8 data..

To verify the effectiveness of our proposed model we selected three glacier lakes - Chorabari Lake, South Sounak Lake and Lemthang Tsho Lake, Bhutan. We downloaded the images of these glacier lakes from the EarthExplorer web portal, and those with less than 25% cloud coverage. To avoid the snow and ice cover we take the timestamp between April and November.

The details of the dataset are given below:

Data sets					
Data set	latitude and longitude	Temporal coverage	Data Format	Resolution [m]	Source
Chorabari Lake	lat:$28°6'20''N$ lon:$089°53'57''E$	2012-2013	Landsat7	30	United States Geological Survey (USGS)
Lemthang Tsho Lake,Bhutan	lat:$30°44'06''N$ lon:$79°04'01''E$	2014-2015	Landsat8	30	United States Geological Survey (USGS)
South Lhonak lake	lat:$27°56^{'}44^{''}Nlon:$ $088°19^{'}56^{''}E$	2013-2014	Landsat7	30	United States Geological Survey (USGS)

Table 3: Dataset

5 Proposed Methods:

5.1 Processing:

First we select the lake area in EarthExplorer web portal and download the landsat images from this website with less than 25% cloud coverage. Then we crop the lake area from each image.

5.2 Method:

Then we find out the Normalized Difference Water Index (NDWI), Normalized Difference Snow Index(NDSI), and Normalized Difference Soil Index [3, 4]. Now we consider those water pixels from NDWI whose pixel values are greater than 0.0[3], snow pixels from NDSI whose pixel values are greater than 0.4[4] and soil pixels whose pixel values are greater than 0. Now we count all the water pixels, snow pixels, and soil pixels and normalize them. Then we prepare these pixel values as an input to the LSTM network.

The input x which is a 3rd order tensor containing a series of a vector containing proportions of water, snow, and soil is fed to layer 1 LSTM from where it is further fed to layer 2, layer 3, and layer 4 LSTMs each containing 7 units. Now, this output is further fed to a series of fully connected layer 5 containing 7 neurons, layer 6 containing 7 neurons, and finally to layer 7 containing 3 neurons. Thus the final output y is a vector of dimension 3 with individual dimensions representing the proportion of pixels having water, snow, and soil respectively.

Our predicted information is added to our data frame. Then we use threshold 0.0625[7] to observe how the soil pixel changes between predicted information and previous information. If the change is greater than the threshold value then the GLOF can happen otherwise it cannot happen.

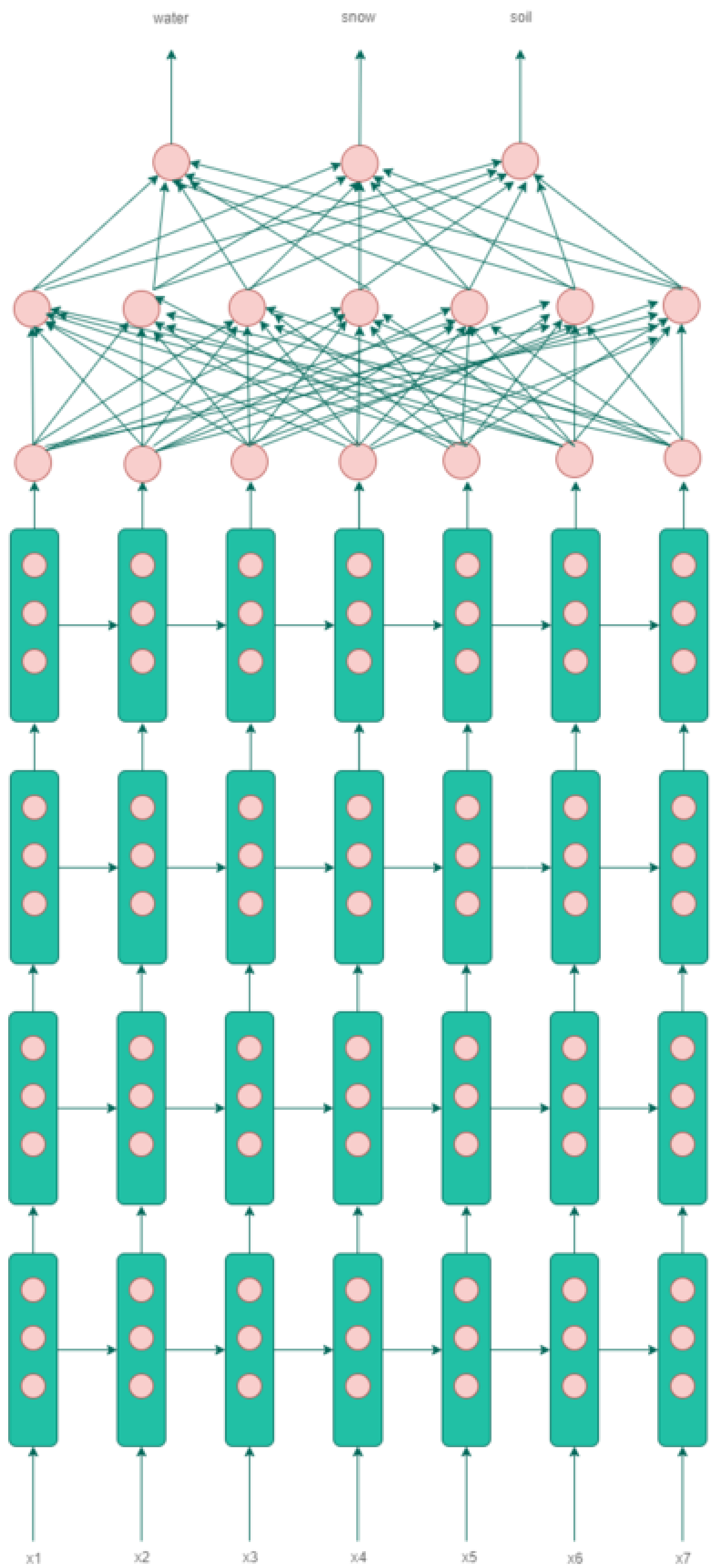


Figure 7: LSTM model architecture

Figure 8: Model Diagram

6 Results

We use the threshold value of 0.0625[7]to observe the changes in soil pixel if the changes are greater than the threshold value then GLOF can happen otherwise not. Based on this threshold we got two correct results and two false results. The details of the precision, recall, accuracy, and F1 score based on these threshold values of each Lake are given below-

Result					Evaluation metrics			
Data set	Number of Images	Temporal Coverage	Actual Result	Predicted Result	Precision	Recall	Accuracy	F1 score
Chorabari Lake	31	2012-2013	1	1	0.55	.84	0.69	.66
Chorabari Lake	25	2010-2011	0	1	.5	.32	.65	0.39
Lemthang Tsho Lake,Bhutan	23	2014-2015	1	0	0.47	0.40	0.78	0.43
South Lhonak lake	17	2012	0	0	.5	.41	.82	0.45

Table 4: Results

7 Conclusion and future works

In this work, we faced certain problems in predicting the occurrence of the glacier outburst. The first one is the lack of uniform data because of the presence of frozen snow in the winter. So, we were unable to use the winter season data. The second one is the presence of clouds in the atmosphere during the rainy season, so when we take the cloud coverage between 0 to 25% then there was a gap of data. The third one is that in Landsat7 images there were many parallel lines which were created as noise in the data . Interestingly our proposed method required lots of sequential data but the increasing the number of satellite data makes it very difficult to handle the images in our computer.

By using deep learning in contrast to the previous works which predicted in the past we are able to forecast future occurrences of glacier outburst with fair predictions as indicated by evaluation metrics in table 4. With sufficient amount of data we could even get better predictions. Future work may include use of convolutional LSTMs for prediction. Since we can predict future proportions of land area we can have a time series of prediction of land proportions in the future. Now these new predictions can be used to train another neural network which would predict the probability of having a glacier outburst for the given period. By sliding the window of inputs we would get different probability of glacier outburst. Now, the year of glacier outburst would correspond to the maximum output probability by the given model.

References

[1] AMIT ANAND ,"GLACIAL LAKE OUTBURST FLOOD HAZARD ASSESSMENT IN A PART OF UTTARAKHAND, INDIA", 2014.

[2] Bo-cai Gao,"NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space",1996

[3] Phila Sibandze , Paidamwoyo Mhangara1, John Odindi, Mahlatse Kganyago1, "A comparison of Normalised Difference Snow Index (NDSI) and Normalised Difference Principal Component Snow Index (NDPCSI) techniques in distinguishing snow from related land cover types ", 2014.

[4] Sepp Hochreiter,Jurgen Schmidhuber, "LONG SHORT-TERM MEMORY",1997.

[5] Ian JoughinBenjamin, E. Smithand Brooke Medley,"Marine Ice Sheet Collapse Potentially Under Way for the Thwaites Glacier Basin, West Antarctica",2014

[6] J. Gabriel Campbell, "Inventory of Glaciers and Glacial Lakes and the Identification of Potential Glacial Lake Outburst Floods (GLOFs) Affected by Global Warming in the Mountains of India, Pakistan and China/Tibet Autonomous Region",2005

[7] Georg Veh,Oliver Korup ,Sigrid Roessner Ariane Walz,"Detecting Himalayan glacial lake outburst floods from Landsat time series",2018.

[8] Huggel et al. (2002),"Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps,2002.

[9] Yingbin Deng, Changshan Wu,, Miao Li c, Renrong Chen,"International Journal of Applied Earth Observation and Geoinformation",2015

[10] C. ScottWatson et al., " Optimising NDWI supraglacial pond classification on Himalayan debris-covered glaciers",2018.

[11] Dennis C. Duro et al,"A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery",2011.

[12] Sonam Rinzin et al.,"Glacial Lake Area Change and Potential Outburst Flood Hazard Assessment in the Bhutan Himalaya",2021.

[13] S. N. Remya et al," Volume estimation of existing and potential glacier lakes: A case study in Sikkim Himalaya"2019.