

LANDSLIDE PREDICTION USING RAINFALL ANALYSIS

Submitted in partial fulfillment of the requirements for the award of

Bachelor Degree in Computer Science

by

Karthika A (Reg.No:38290034)

Mahalakshmi D (Reg.No:38290044)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU.**

APRIL 2021



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Karthika A (Reg.No:38290034) and Mahalakshmi D (Reg.No:38290044) who carried out the project entitled "**LANDSLIDE PREDICTION USING RAINFALL ANALYSIS**" under my supervision from October 2020 to April 2021

Internal Guide

Dr.K. Mohana Prasad, M.E., Ph.D

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I, Karthika A (Reg.No:38290034) and Mahalakshmi D (Reg.No:38290044) hereby declare that the Project Report entitled “**LANDSLIDE PREDICTION USING RAINFALL ANALYSIS** ” done by me under the guidance of Dr. K. Mohana Prasad at Sathyabama Institute Of Science And Technology, Chennai is submitted in partial fulfillment of the requirements for the award of Bachelor degree in Computer Science .

DATE:
PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M. E., Ph.D., Dean**, School of Computing, **Dr. S. Vigneshwari M. E., Ph. D., and Dr. L. Lakshmanan, M. E., Ph. D., Heads of the Department** of Computer Science for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my honor and gratefulness to my Project Coordinator **Dr. K. Mohana Prasad, M.E., Ph.D**, Department of Computer Science and Engineering, who was beside me throughout the course of the project providing me all the support with keen interest.

I also thank my friends and my family members who had been providing me with the moral support. Last, but not the least, I thank God Almighty for giving me the strength and intellect to finish this project successfully.

ABSTRACT

Regional precipitation-induced landslide is one of the major geologic hazards in our country and it is mainly affects in large areas, sudden breakout and serious consequences. Our aim is to develop landslide early warning systems for India, the several researchers developed different models to assess landslide occurrence risk based on daily rainfall forecasts and rainfall thresholds. The existing system employs Artificial Neural Networks (ANN) to predict rainfall intensity and then assesses the risk of landslide occurrence by comparing it with rainfall thresholds. But, here the prediction accuracy is low and computational time is high.

To overcome the drawbacks in landslide prediction, classification mechanism called Support Vector Machine (SVM) has been employed. The dataset has been taken for the Cherapunjee region which receives the highest intensity of rainfall in India. The aim is to predict the landslide occurrence and classify the risk level associated with the landslide. To improve the reliability in landslide prediction, the proposed model uses pre-processing for removing null values in the dataset. After getting the pre-processed dataset, it will apply normalization, then SVM training and finally the Testing process. Support Vector Machines can exhibit a large degree of flexibility in handling tasks of varied complexities because of the non-linear boundary functions.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	i
	LIST OF FIGURES	iv
	LIST OF ABBREVIATIONS	v
1	INTRODUCTION	1
	1.1 OVERVIEW	2
	1.2 ORGANISATION OF THESIS	3
2	LITERATURE REVIEW	4
3	AIM & SCOPE OF THE PROJECT	8
	3.1 SCOPE OF THE PROJECT	8
	3.2 OBJECTIVE	8
	3.3 PROBLEM DEFINITION	8
	3.4 EXISTING SYSTEM	8
	3.4.1 Issues in the existing system	9
	3.5 PROPOSED SYSTEM	9
	3.5.1 Advantages	10
4	METHODS AND ALGORITHMS USED	11
	4.1 HARDWARE REQUIREMENTS	11
	4.2 SOFTWARE REQUIREMENTS	11
	4.2.1 Overview of Java	11
	4.2.2 Hadoop	12
	4.2.3 Design and Development	13
	4.3 SYSTEM DESIGN	13
	4.4 SYSTEM ARCHITECTURE	15
	4.5 FLOW OF OVERALL PROJECT	15
	4.6 MODULE DESIGN	16
	4.6.1 Landslide dataset preparation and preprocessing	16
	4.6.2 Normalization and SVM training	17
	4.6.3 Landslide prediction using SVM	18
	4.6.4 Landslide Stage Identification	18
	4.7 PROPOSED ALGORITHMS	19

4.7.1	Landslide Prediction Algorithm	19
4.7.2	Advantage	20
5	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS	21
5.1	RESULTS AND DISCUSSION	21
5.1.1	Landslide dataset preparation and preprocessing	21
5.1.2	Normalization and SVM training	22
5.1.3	Landslide prediction using SVM	23
5.1.4	Landslide Stage Identification	24
5.2	PERFORMANCE ANALYSIS	25
6	SUMMARY AND CONCLUSION	27
6.1	SUMMARY	27
6.2	FUTURE ENHANCEMENT	27
	REFERENCES	28
	APPENDIX	29
a.	SAMPLE CODE	29
b.	PUBLICATIONS	30

LIST OF FIGURES

FIGURE NO.	NAME OF THE FIGURE	PAGE NO.
4.1	System Architecture	15
4.2	Flow of overall project	15
4.3	Landslide dataset preparation and preprocessing	16
4.4	Normalization and SVM training	17
4.5	Landslide prediction using SVM	18
4.6	Landslide Stage Identification	18
5.1	Home page	21
5.2	Uploading of dataset	22
5.3	Selected dataset	22
5.4	SVM Training	23
5.5	Landslide prediction	23
5.6	Landslide Result	24
5.7	Landslide Stage Identification	24
5.8	Comparison with Real, ANN, SVM	25
5.9	Comparison with Real, ANN, SVM	25

LIST OF ABBREVIATIONS

SVM	- Support Vector Machine
ANN	- Artificial Neural Network
RBF	-Radial Basis Function
LIBSVM	-Library Support Vector Machine
HDFS	-Hadoop Distributed File System
FFT	-Fast Fourier Transform
FFNN	-Feed Forward Neural Network

CHAPTER 1

INTRODUCTION

The occurrence of extreme rainfall in a short time causes serious damage to economy and sometimes even loss of lives due to flood. Insufficient rainfall for long period causes drought. This can effect to economic growth of developing countries. Thus, rainfall estimation is very important because of its effects on human life, water resources and water usage. However, rainfall affected by the geographical and regional variations and features is very difficult to estimate.

According to statistics, natural hazards are believed to account for up to 4 percent of the total annual deaths worldwide, besides causing enormous economic losses and uprooting habitation, of which landslides form a significant component. The term landslide in the strict sense may be defined as a process of involving the downward and outward movement of a part of the slope forming materials, due to the action of gravity, other forms of mass movements like falls, flows, topples and creeps are generally included in landslides. The occurrence of landslides is a function of spatial variability of site conditions, local geology and rainfall and hence is highly uncertain. Among the various factors triggering the occurrence of landslides, rainfall holds unprecedented significance. India is among the top ten countries with highest percentage of landslide fatalities for past few years.

Cherrapunjee, located in the sub divisional town in the East Khasi Hills district in the Indian state of Meghalaya which receives the highest rainfall in India is chosen to be the study area for this piece of work. It is studied to receive a very high intensity rainfall which approximately crosses 11,754.5 mm per year during the monsoon.

Cherrapunjee receives the rains from Bay of Bengal during the monsoon. The monsoon clouds hit the Khasi Hills before which they are found to be unhindered over the plains of Bangladesh for about 400 km. The deep valleys channels along with the geography of the hills with the low-flying moisture-laden clouds from a wide area to converge over Cherrapunjee. Most of Cherrapunjee's rain is the result of air being lifted as a large body of water vapour. The extreme amount of rainfall at Cherrapunjee is perhaps the best-known feature of rain in northeast India. Occasionally, cloudbursts can occur in one part of Cherrapunjee while other areas may be totally or relatively dry, reflecting the high spatial variability of the rainfall. Atmospheric humidity is extremely high during the peak monsoon period.

The researchers have employed various models which include both empirical as well as mathematical concepts such as ANN to predict rainfall intensity and assess the risk of landslide occurrence in comparison with the rainfall threshold. But here the prediction accuracy is observed to be low and computational time is high.

The very motive is to develop a landslide prediction model and generating the risk level associated with the landslide. This can increase the preparedness level to reduce the severity of the disaster thereby having a control on the human loss and also the infrastructural damages. As far as this area of study is concerned, there is no landslide prediction system in this region. The work employs the Support Vector Machine Concept to develop an early warning approach to the landslide. The risk level (high, medium, low) is also evaluated.

1.1 OVERVIEW

The major problem in the existing system is associated with the use of artificial neural networks to predict rainfall intensity, one day in advance. The predicted rainfall intensity is then compared to the daily rainfall thresholds developed for the study area and the risk of landslide occurrence is evaluated. Predicting rainfall intensity with antecedent rainfall data alone, without the use of other meteorological parameters, also adds to the novelty of the approach.

The suitability of ANNs for daily rainfall forecasting, as all the back propagation neural network models gave satisfactory performance in terms of landslide risk assessment based on the daily forecasted rainfall. The study is done with an aim to enlighten the respective authorities about a feasible way of developing a landslide early warning systems for landslide prone area.

Because of the nature of the feature space in which these boundaries are found, Support Vector Machines can exhibit a large degree of flexibility in handling classification and regression tasks of varied complexities.

1.2 ORGANISATION OF THESIS

Chapter 1 deals with an introduction to the project where the existing system is been discussed. It also gives an overview of how ANN has been incorporated for the landslide prediction.

In Chapter 2, a detailed description of the literature survey of the papers which are referred during the course of the project was summarized.

Chapter 3 gives a brief explanation on the aim and scope of the project. Here proposed system has been compared with the existing system. The issues in the existing system and the advantages of the proposed system are also discussed in this chapter

Chapter 4 deals with the methods and algorithms used. The hardware and software requirements are provided along with the system design, architecture and flow of overall project. The module design and the landslide prediction algorithm is also mentioned here.

In chapter 5, the Results and Discussion along with the screen layouts of each module has been depicted. It also includes the performance analysis of SVM and ANN.

Chapter 6 deals with the summary and conclusion of the project. It also includes the future scope of the project

CHAPTER 2

LITERATURE REVIEW

The literature review includes the papers which we have referred for the implementation of the project. It gives a brief idea on the issues in the existing techniques and their implementation.

According to S. Renuga Devi et.al (2014), different techniques have been applied for the task of rainfall prediction worldwide. But the prediction accuracy obtained still lies below the satisfaction level due to large uncertainties involved in the climatic processes. With an aim to develop landslide early warning systems for Nilgiris, they developed different models to assess landslide occurrence risk, based on daily rainfall forecasts and rainfall thresholds. The paper employs Artificial Neural Networks to predict one day in advance rainfall intensity and then assesses the risk of landslide occurrence by comparing it with rainfall thresholds. The data set comprises of daily recorded rainfall intensities at 14 rain gauge stations located in and around Coonoor. The results obtained and sensitivity analysis performed establishes the efficiency and adequacy of rainfall data as a supplement

to different meteorological parameters and suitability of artificial neural networks in forecasting rainfall and hence evaluating the risk of landslide occurrence.

In his paper, author K. Anbarasu et.al (2010) has given an implication of the proposed model such that slide movement at LantaKhola is dictated by the collapse of the contact zone and that material from this region is extruded downslope from the point of intersection with the surface. As a consequence, material is consistently removed from within the weak contact zone and transported downslope as slide debris. Mass depletion leads to volume reduction within the zone, causing subsidence under the weight of the overlying rock overburden. This causes instability in the overlying gneiss unit and blocks of the gneiss collapse onto and are ultimately incorporated within the debris of the slide. A series of mitigatory measures including the construction of gabion walls two rows of pipe piles, and an under drainage system behind the wall have been envisioned to facilitate safe drainage of surface and ground water and prevention of migration of fines at three different elevations along the LantaKhola slope between the contact zone and the road level.

According to author A. Sengupta et.al (2010), Himalayan regions are characterized by extremely steep and rugged slopes where landslides are a major hazard. Very few site-specific studies have been conducted on landslides in these regions because of the very difficult terrain, though a number of landslide susceptibility zonation studies have been carried out to identify the vulnerable areas. These studies mostly include construction and superimposition of thematic maps that supposedly correspond to contributing causative factors. In view of this record, a comprehensive study was recently undertaken to find out the cause for the activation of this landslide and to monitor the landslide

In his paper, author S. S. Chandrasekaran et.al (2011), have investigated three case histories from Nilgiris district: the slope failure of a railway track at Aravankadu, failure of retaining walls supporting buildings at Coonoor, failure of the slope and retaining wall along national highway (NH67) at Chinnabikatty. Laboratory investigations are carried out on soil samples collected at these sites. Soils at all the three locations have high fine content and low values of coefficient of permeability. Finite element analysis of all the three case histories was carried out using software in order to understand the failure mechanism and contributing factors. Slope stability analysis using strength reduction technique is carried out for the slope at Aravankadu to determine the critical slip surface and factor of safety.

According to author Dimitri A. Lisin et.al (2013), slope stability analysis using strength reduction technique is carried out for the slope at Aravankadu to determine the critical slip surface and factor of safety. Results reveal that the increase in pore pressures led to a reduction in shear strength of the soil and consequently resulted in progressive failure of

slope at Aravankadu site. Displacement analysis is carried out for Coonoor and Chinnabikatty sites. The results show that combined effect of surcharge load of building and high pore pressure led to intense shearing behind the retaining walls at Coonoor site. Results indicate occurrence of large displacements along the face

In his paper, author S. S. Chandrasekaran et.al (2010) suggests the increasing demand for coal of the industrial society has led to more and more serious coal mining. The Mining-induced landslide hazard has seriously influenced the sustainable development of mining areas. So it is important for us to select a suitable method to predict mine slope stability. However, there are many influencing factors for landslide and the effects of the same factors in various areas are different. The mathematical relationship between the factors which impact landslide and the landslide stability prediction is hard to obtain. Therefore, it is a comparatively accurate method to get a statistical analysis model with the historical data. SVM can get solution by solving a convex quadratic programming question. The solution is global optimal solution, and its ratio is high. The Prediction with SVM will use structural risk minimization principle instead of the empirical risk minimization principle, maximize the generalization ability of learning machine, make sure that the independent test set which was gotten from a limited sample of training set remains a small error, and get a non-linear mathematical relationship with a higher dimension at the same time.

In his paper, author Biswajeet Pradhan et.al (2009), the combination of Fast Fourier Transform (FFT) and Feed Forward Neural Network (FFNN) is applied for next one year rainfall prediction. Fast Fourier transform with filtering is performed on interpolated rainfall data to separate periodic components. These periodic components and delayed periodic components are given as input and desired output respectively to an FFNN for training. While testing the output of FFNN, inverse FFT gives the predicted rainfall value by amount of training input-output delay.

According to the work of author Seema Mahajan et.al (2013), there is regional variation in rainfall distribution of India and separate rainfall prediction model for each subdivision is desired. Global climatic factors like, distance from sea, latitude the Himalayan Mountain, distribution of land and water, surface pressure and wind, upper air circulation and western cyclones are affecting Indian sub-divisional rainfall. Rainfall is complex functions of these parameters. Researchers are correlating these independent parameters to predict rainfall. These parameters are embedded in rainfall time series.

Rainfall prediction in the proposed model uses effect of these parameters as part of rainfall time series.

In his paper, author X. Yao et.al (2006), proved the need to assess landslide hazard for the purpose of land-use planning and hazard mitigation has led to the development of many statistical and process-based models, with increasing emphasis on the use of GIS. Statistical modelling approaches have generally taken the form of multivariate statistical analyses of physical characteristics associated with past or weighted hazard ratings based on environmental attributes related to whereas the process based models have generally used infinite slope stability theory coupled with geomorphological, hydrological, geological and vegetation data to estimate slope, strength parameters and pore pressure distribution and then to model the factor of safety of the slopes. In general, GIS-based statistical models have been able to indicate general regions of landslide susceptibility, whereas GIS-based process models specifically include the physical processes involved in land sliding and therefore can often better describe the physical processes of mass movement. Theoretically, process models commonly use site-specific data and should produce a more realistic result with slope instability modelling than most statistical approaches.

Thus from the literature survey, the issues in the existing system has been studied which helped in improving the proposed work. The aim and scope of the project is been discussed in the next chapter

CHAPTER 3

AIM & SCOPE OF THE PROJECT

3.1 SCOPE OF THE PROJECT

The Support Vector Machine concept has been used in accordance with the constructive machine learning procedure based on the statistical learning theory. It provides non-linear solutions to regression as well as classification problems by transforming the inputs which

occupy a large dimension space. In this project, SVM model has been used for both training and testing process

The forecasting of landslide has been dealt using factors which affects the landslide occurrence. We have used the training of data samples to determine several classes for prediction and to remove the null values. Also, the testing process has been done to improve efficiency and accuracy.

3.2 OBJECTIVE

The objective of the project is to develop a prediction model by using Support Vector Machine (SVM) technique which is meant to possess a strong capability to predict landslides by forecasting rainfall dataset using BigData concept and also to predict the landslide occurrence and classify the risk level associated with the landslide

3.3 PROBLEM DEFINITION

The existing system used Artificial Neural Networks which has low accuracy and high computational time. To overcome these drawbacks, Support Vector Machine model concept has been approached.

3.4 EXISTING SYSTEM

The existing system employs artificial neural networks to predict rainfall intensity, one day in advance. The predicted rainfall intensity is then compared to the daily rainfall thresholds developed for the study area and the risk of landslide occurrence is evaluated. Predicting rainfall intensity with antecedent rainfall data alone, without the use of other meteorological parameters, also adds to the novelty of the approach.

The suitability of ANNs for daily rainfall forecasting, as all the back propagation neural network models gave satisfactory performance in terms of landslide risk assessment based on the daily forecasted rainfall. The existing systems owe to high cost in case of usage of sensor networks

3.4.1 Issues In Existing System

- ❖ There is no risk level prediction for landslide
- ❖ Landslide prediction accuracy is low.
- ❖ Classification process takes more time.
- ❖ Complexity is high

3.5 PROPOSED SYSTEM

To improve the reliability in landslide prediction, our proposed system use preprocessing for removing null values in the dataset. After getting the preprocessed dataset, we apply normalization, SVM training and testing process. This method performs regression and classification tasks by constructing nonlinear decision boundaries. Because of the nature of the feature space in which these boundaries are found, Support Vector Machines can exhibit a large degree of flexibility in handling classification and regression tasks of varied complexities.

By using SVM classification, we predict the landslide level for the given dataset. In output, there are two possibilities

- ❖ No landslide
- ❖ If there is landslide occur then it will predict the risk level of landslide such low, medium or high.

For this, we utilize climatic threshold method and the attributes.

3.5.1 Advantages

- ❖ Landslide Prediction accuracy is high.
- ❖ Risk level of landslide also predicted.
- ❖ Reduced complexity
- ❖ Performance is very high.

Thus the aim and scope of the project has been discussed and the issues with the existing system are also studied. The advantages in the proposed work are also mentioned. The methods and the algorithms implemented in the project are discussed in the next chapter.

CHAPTER 4

METHODS AND ALGORITHMS USED

4.1 HARDWARE REQUIREMENTS

Processor	: i5
Clock speed	: 550MHz
Hard Disk	: 500 GB
RAM	: 4 GB
Cache Memory	: 512KB
Operating System	: Windows8

4.2 SOFTWARE REQUIREMENTS

FRONT END	: JAVA
BACK END	: HADOOP

4.2.1 Overview of Java

One of the most compelling reasons to move to java is its platform independence. Java runs on most major hardware and software platforms, including windows 98, 95, and NT Macintosh and several varieties of UNIX. Originally java called oak by James Gosling, one of the inventor of the language. The java development team which included Patrick Naughton discovered that the existing language like C and C++ had limitations in terms of both reliability and portability. However, they modeled their new language java on C and C++ but removed a number of features of C and C++ that were considered as sources of problems and thus made java a really simple, reliable, portable and powerful language.

Specifically, this overview will include a bit of the history of java platform, touch of the java programming language, and the ways in which people are using java applications and applets, now and in the likely future. After going a while down the path of consumer – electronics devices, they realized that they had something particularly cool in the java language and focused on it as a language for network computing. Sun formed the java soft group which in a little over three years has grown to over six hundred people working on java related technologies.

Java quickly became a hot buzzword of the computing industry. People wanted to know Java - it was said to be great for creating dynamic interactive content for webpages. Yet the true power of Java lies not in applets, but in its many other uses. Java is used for developing standalone applications, and for server-side programming. The face of Java has changed, but the core language remains the same. In this tutorial series, I'll teach you the basics of Java programming. You'll still need a good book as a companion to this tutorial series, but for those who are dabbling in Java, this should be enough to get your feet wet.

Java is an object-orientated language, and may not be suitable for first time programmers. Learning a new language takes some time, but learning your first object-orientated language can be exceedingly difficult. Nonetheless, if you've done some C programming before, the shift into Java shouldn't be unreachable, providing you obtain a good reference book. There are also those that believe programmers should start with an object-orientated language first, and many universities have adopted this practice. Java is two things: a programming language and a platform.

4.2.2 Hadoop

Hadoop is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common and should be automatically handled by the framework.

The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS), and a processing part called MapReduce. Hadoop splits files into large blocks and distributes them across nodes in a cluster. To process data, Hadoop transfers packaged code for nodes to process in parallel based on the data that needs to be processed. This approach takes advantage of data locality nodes manipulating the data they have access, to allow the dataset to be processed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system where computation and data are distributed via high-speed networking

4.2.3 Design & Development

System design is a process by which the software requirements are translated into a representation of software components, interfaces and data necessary for the implementation phase. Implementation is the final phase of the project which reveals us the real outcome of the previous steps. Here the software based tasks directly dealing with the projects target is carried out. The implementation explains about the way the project work. This includes the software support, modules and their implementation.

4.3 SYSTEM DESIGN

Design is a meaningful engineering of something that is to be built. Software Design sites at the technical kernel of software engineering. Software design is a process through which the requirements are translated in to a representation of the software i.e. the blue print for constructing software. Design provides us with representation of software that can be assesses for quality. Design is the only way that we can accurately translating a customer's requirements in to a finished software product.

The design concept of the landslide prediction model has been explained in accordance with the algorithms mentioned above. Initially Collect Landslide Datasets from the universal database and store the datasets into database. Then apply pre-processing to replace all the null values (only applicable for numerical dataset). Pre-processing reduce the time complexity. Then pre-processed data were taken as an input for training process.

Apply normalization process for normalizing the pre-processed dataset. Radial Basis Function is one of the major kernel functions is utilized for SVM training process.

In the testing phase, Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training. Then proceed by taking the query dataset for testing phase. Apply pre-processing for input query dataset. Support Vector Machine compares the query dataset with trained dataset. It displays the predicted status (landslide or not). Apply climatic threshold based prediction mechanism if the classification result is relevant to landslide. Finally display the risk level of landslide such as (low, medium and high)

INPUT

In this proposed system, the input of this project is landslide prediction dataset

BEHAVIOUR

The proposed system analyzes the given input after SVM training and testing phases.

OUTPUT

The output of this project is to display the predicted risk level of landslide.

PERFORMANCE

The performance of the proposed system will be high. This is due to better accurate results when compared to ANN

USABILITY

It efficiently provides valuable new information for landslide assessment and hazard mitigation planning.

4.4 SYSTEM ARCHITECTURE

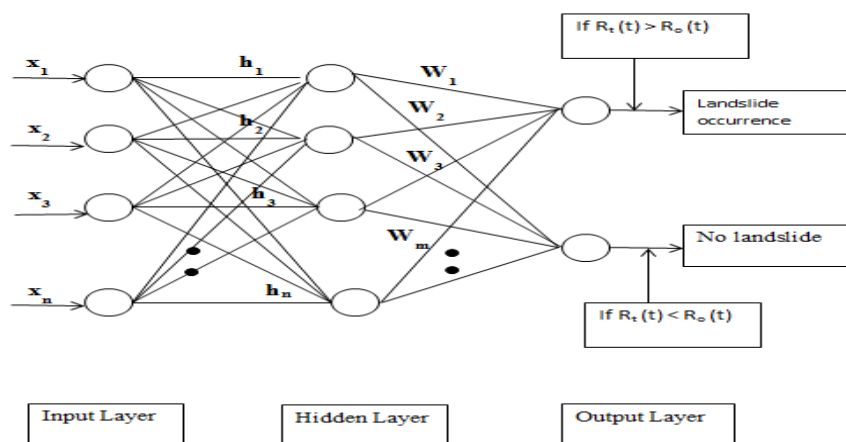


Fig. 4.1: System Architecture

The parameters taken as input includes $X_1, X_2, \dots, X_n$ where n represents the number of input nodes. The Rainfall Threshold for the day is represented as $R_t(t)$ and the Rainfall Predicted for the day as $R_o(t)$. The system architecture is specified in fig 4.1.

4.5 FLOW OF OVERALL PROJECT

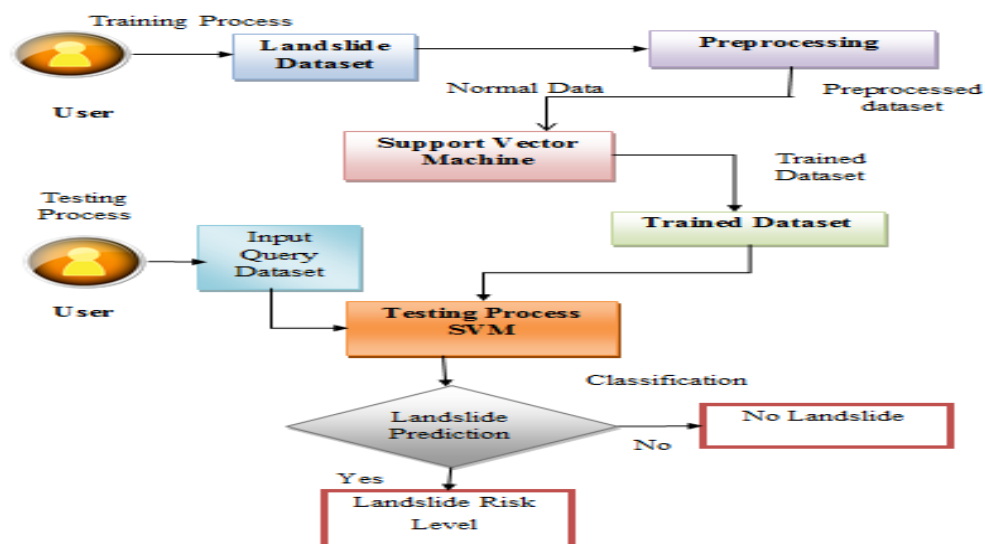


Fig. 4.2: Flow of Overall Project

The flow of overall project is depicted in Fig 4.2 which provides how the entire project is incorporated using SVM.

4.6 MODULE DESIGN

The proposed work has been implemented as in 4 modules. They are described as follows:

- Landslide Dataset Preparation and Pre-processing
- Normalization and SVM Training
- Landslide Prediction using SVM
- Landslide Stage Identification

4.6.1 Landslide Dataset Preparation and Pre-processing

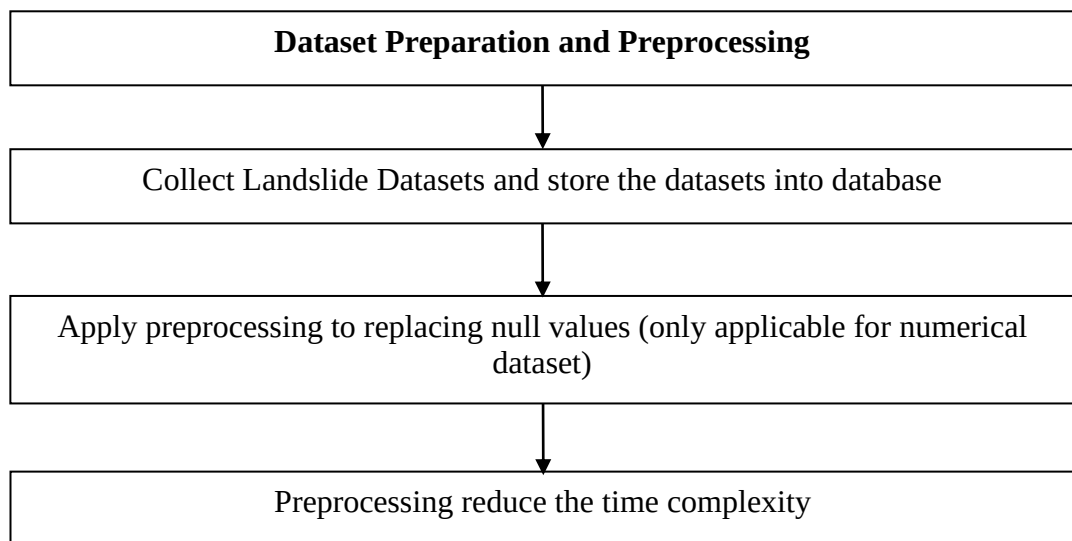


Fig. 4.3: Landslide Dataset Preparation and Pre-processing

The Prediction of landslide based on the previous year's data and records is shown in fig 4.3. In the first step, for predicting the landslide, need to collect all the previous record for preprocessing function, which makes a specific result for the future Prediction. The preprocessing method removes the null value.

4.6.2: Normalization and SVM Training

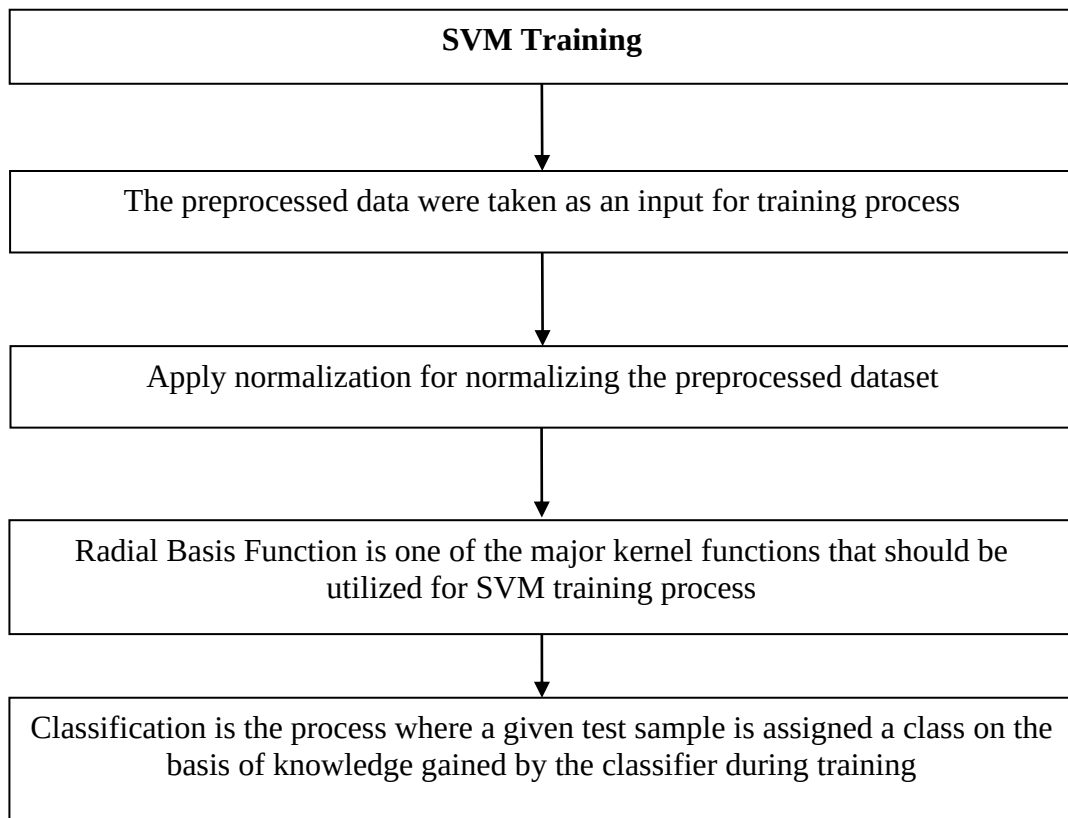


Fig. 4.4: Normalization and SVM Training

The support vector machine (SVM) is training the dataset by using Radial Basis Function in fig 4.4. Radial Basis Function is one of the major kernel functions that should be utilized for SVM training process. The trained data is need to classify and testing by SVM. Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training.

4.6.3: Landslide Prediction using SVM

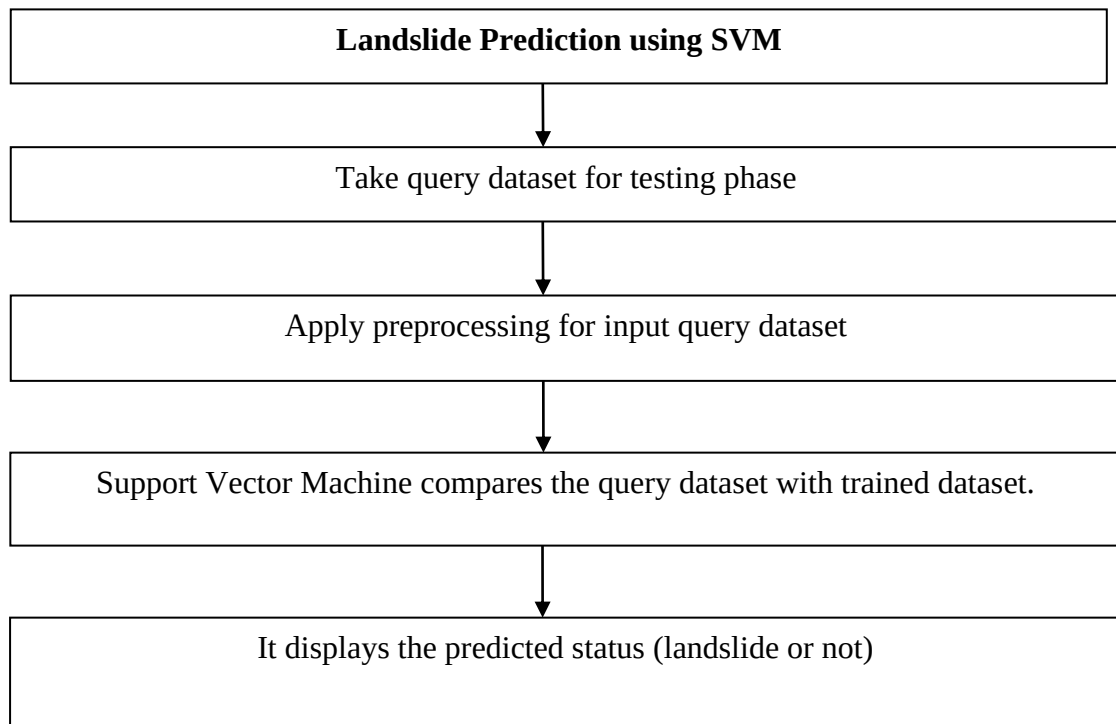


Fig. 4.5: Landslide Prediction using SVM

The users are making a request for the testing process by Support Vector Machine technique. The testing is essential process for predicting the landslide based on the classification. To make the classification results comparable and for exhaustive data analysis, we have used leave one out classification method for the SVM classifier. In classification, SVM efficiently classifies the landslide dataset as shown in fig 4.5.

4.6.4: Landslide Stage Identification

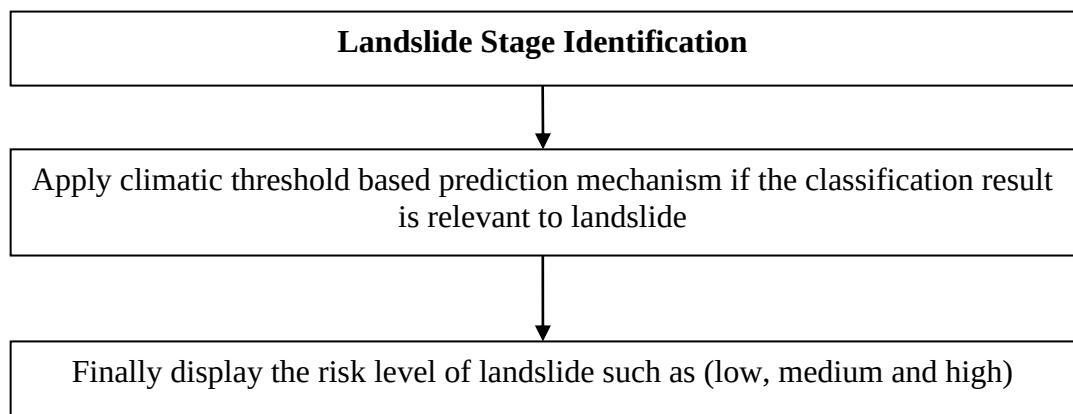


Fig. 4.6: Landslide Stage Identification

Predicted landslide image is taken as an input for this module in fig 4.6. In this module, the risk level of future prediction landslide dataset is analyzed. Consider three types of risk level. The climatic threshold based mechanism attempts to classify the dataset into 3 stages.

- ❖ Low
- ❖ Medium

❖ High

Finally display the risk level of landslide such as (low, medium and high).

4.7 PROPOSED ALGORITHMS

4.7.1: Landslide prediction algorithm

The landslide prediction algorithm has three phases, the algorithm for predicting landslide is briefly explained below

Algorithm 1: Landslide_prediction_Algorithm

Input: Climatic data from the Indian Meteorological Observatory

Output: Display the landslide occurrence dates and classify the risk level of it.

The main Detection procedure for landslide prediction algorithm is given below:

Steps

1. Start
2. {
3. Rainfall dataset as an input
4. Initial input size
5. Set threshold value of Rainfall
6. initialize threshold to zero
7. {
8. Initialize the input format
9. Training of the rainfall dataset with SVM
10. Calculate the error rate of trained dataset
11. Testing of the dataset with SVM
12. Normalization of Trained dataset
13. Set maximum and minimum fraction
14. Start the Yearly evaluation of the landslide
15. End the yearly evaluation of the landslide
16. {
17. Compute Threshold
18. Predict the risk level of land slide
19. }
20. }
21. }
22. End

4.7.2 Advantage

- ❖ *Better efficiency*
- ❖ *Low computational time*
- ❖ *Better accuracy*

The design for each module has been described and the algorithm for the project work has been mentioned. The result and the performance analysis are being discussed in the next chapter

CHAPTER 5

RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS

5.1 RESULTS AND DISCUSSION

In this section, result of the proposed methodology is discussed. The proposed technique has proved to improve the reliability in landslide prediction where it uses pre-processing for removing null values in the dataset. After getting the pre-processed dataset, it will apply normalization, then SVM training and finally the Testing process.

5.1.1 Landslide Dataset Preparation and Pre-processing

SCREENSHOTS

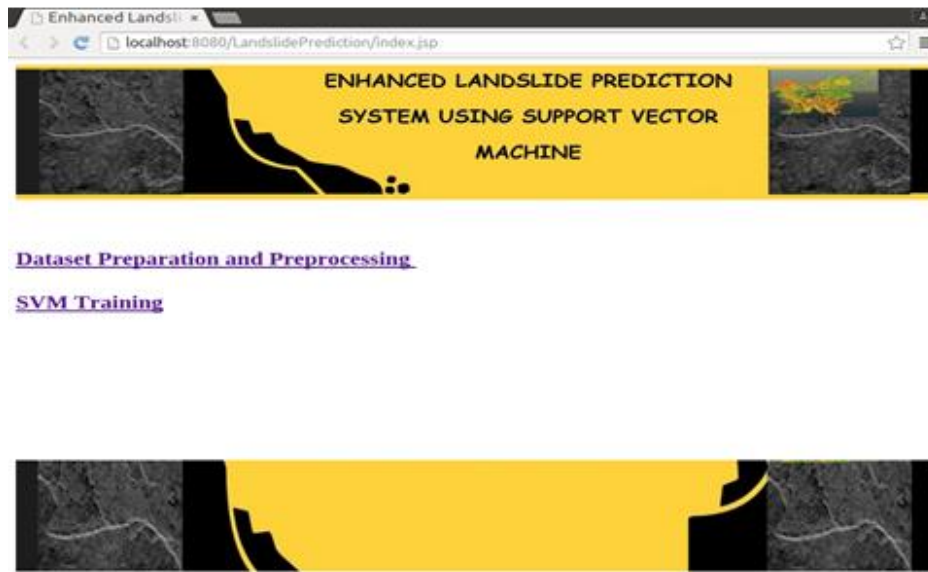


Fig: 5.1 Home page



Fig: 5.2 Uploading of dataset

The Prediction of landslide based on the previous year's data and records. In first step, for predicting the landslide, the dataset is collected for pre-processing function, which makes a specific result for the future Prediction. In the home page, when clicked on the dataset preparation and preprocessing, it will direct to the dataset selection page. The cheeradataset.csv file is then chosen and uploaded into the hive server. The pre-processing method removes the null value. This is depicted in fig 5.1 and fig 5.2

5.1.2 Normalization and SVM Training

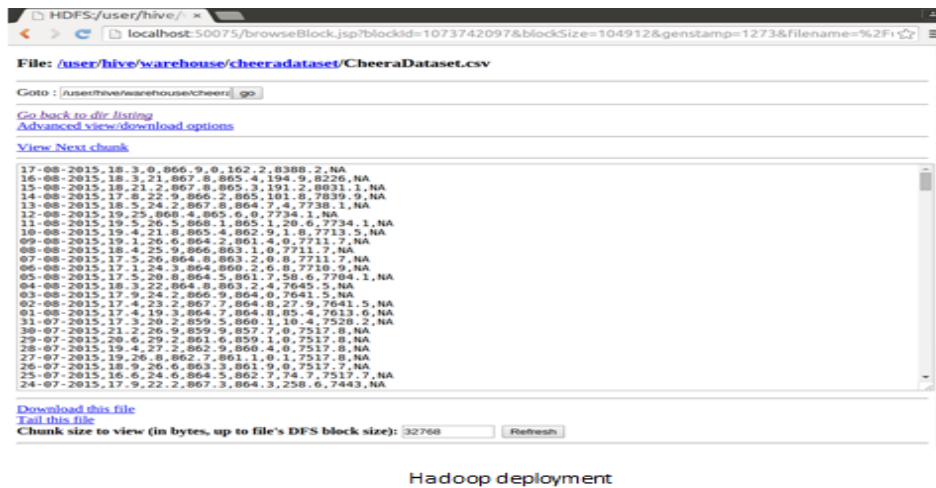


Fig 5.3: Selected Dataset

The support vector machine (SVM) is training the dataset by using Radial Basis Function (RBF). RBF is one of the major kernel functions that should be utilized for SVM training process. The uploaded dataset from hive server is displayed in fig 5.3

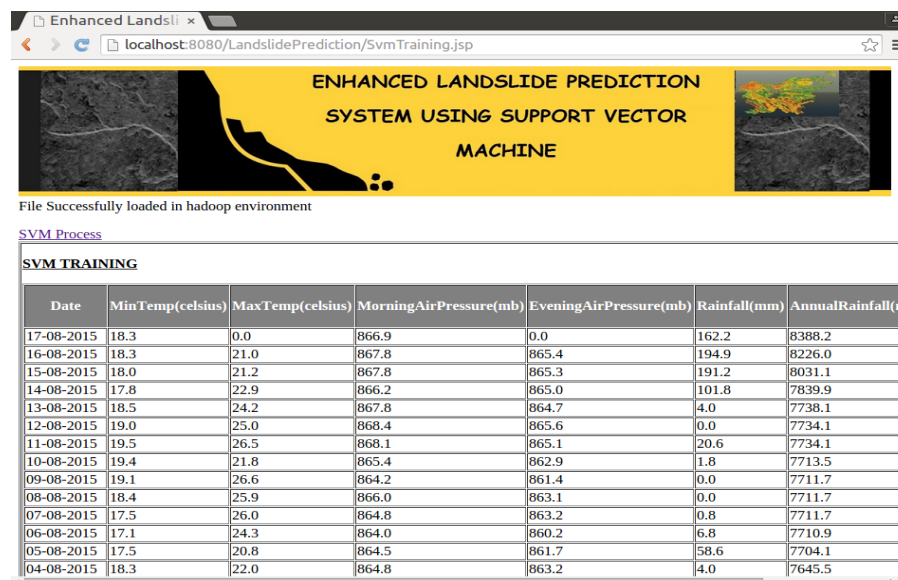


Fig: 5.4 SVM Training

The trained data is classified and tested by SVM. This is shown in fig 5.4. After the file is been retrieved from the hadoop environment, training of the selected dataset is carried out. Several parameters such as minimum temperature, maximum temperature, rainfall etc are trained here. Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training.

5.1.3 Landslide Prediction using SVM

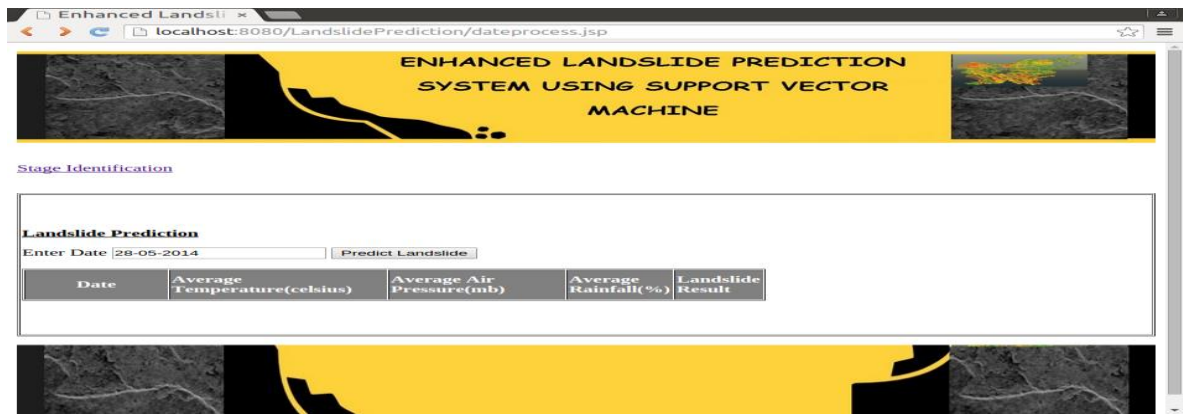


Fig: 5.5 Landslide prediction

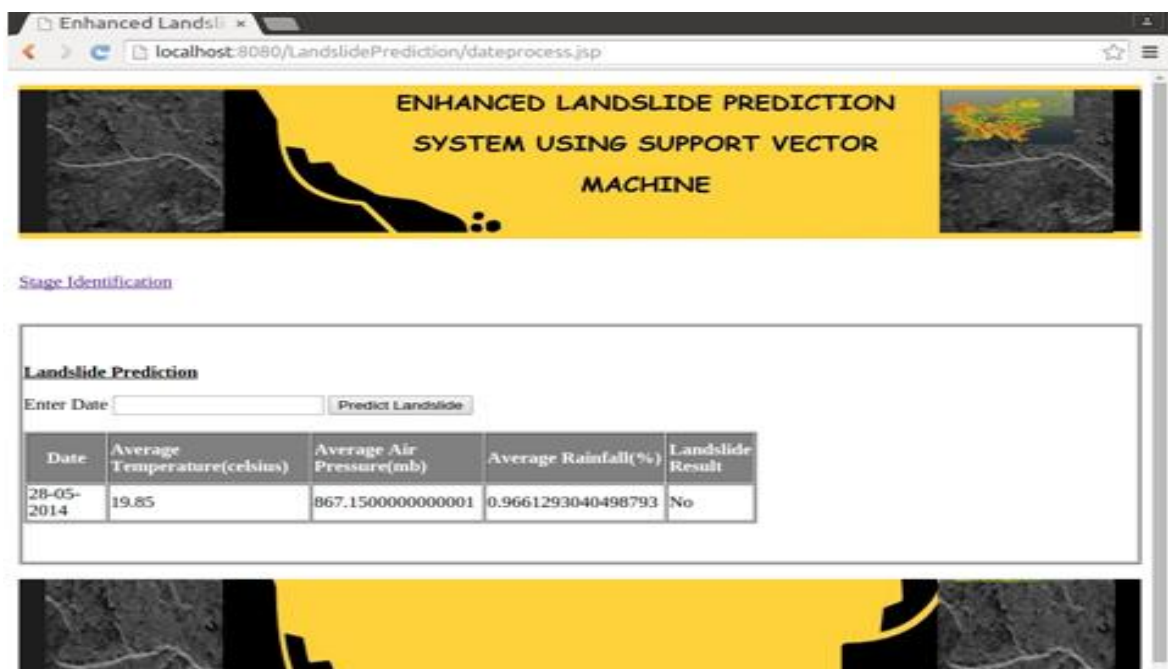


Fig: 5.6 Landslide result

By clicking on SVM process, it directs to a page where the user needs to enter the date on which the landslide occurrence has to be predicted. This is shown in fig 5.5. The users make a request for the testing process by Support Vector Machine technique. To make the classification results comparable and for exhaustive data analysis, we have used classification method for the SVM classifier. In classification, SVM efficiently classifies the landslide dataset and the result that there is no occurrence of landslide for the given date is finally shown in fig 5.6.

5.1.4 Landslide Stage Identification

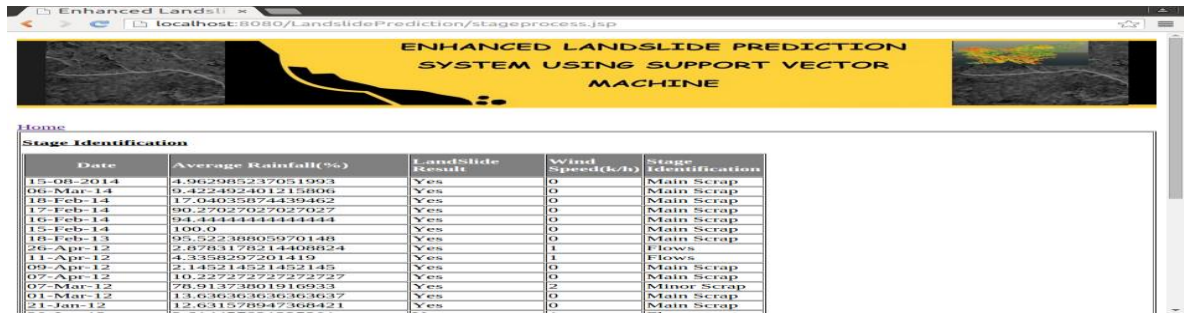


Fig: 5.7 Landslide stage identification

Predicted landslide image taken as an input for this module. In this module, the risk level of future prediction landslide dataset is analyzed as in fig 5.7. Consider three types of risk level. If the result is “yes” then the link to stage identification is clicked where it depicts the risk level associated with the landslide. Finally display the risk level of landslide such as (low, medium and high).

5.2 PERFORMANCE ANALYSIS

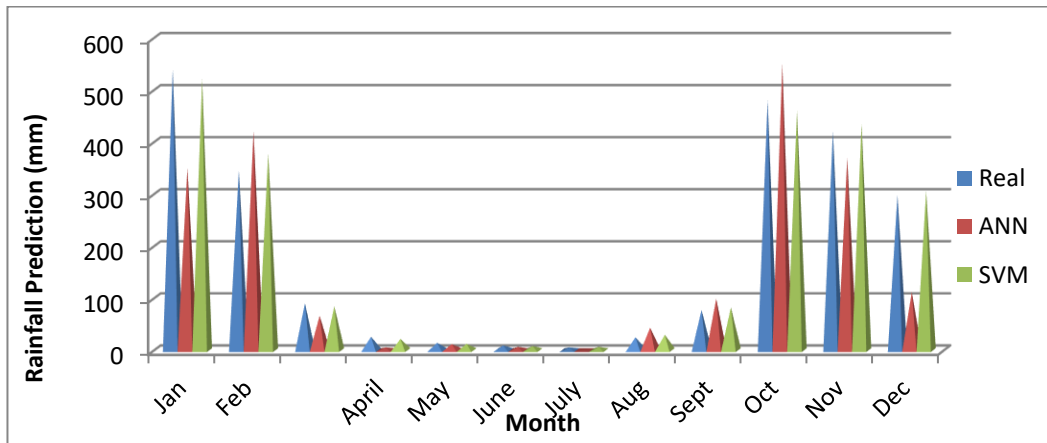


Fig: 5.8 Comparison with Real, ANN, SVM

The above graph fig 5.8 is showing the prediction of rainfall comparison with real, ANN and proposed technique SVM. The result is showing that the proposed technique SVM is better than ANN and it produces more relevant and approximate result.

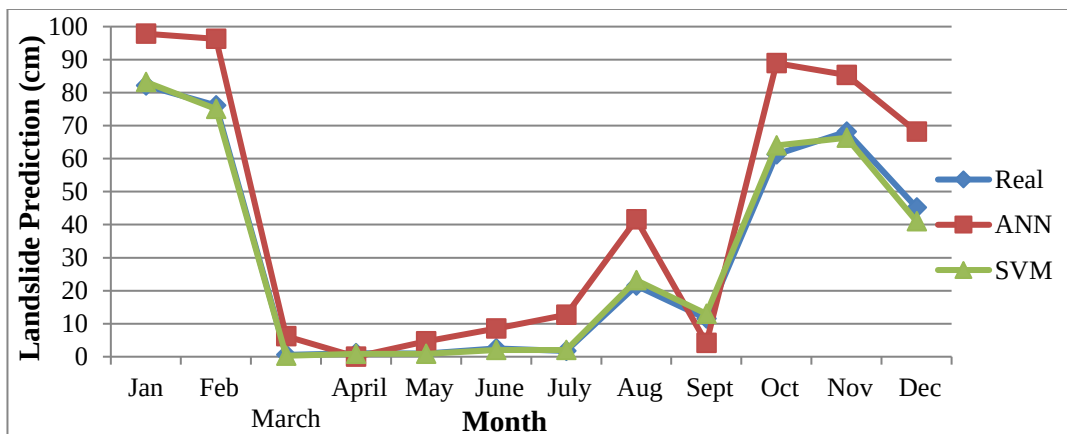


Fig: 5.9 Comparison with Real, ANN, SVM

The above graph fig 5.9 is showing the landslide prediction accuracy within the rainfall condition, the proposed technique SVM is producing more relevant and close to exact or real prediction of the landslide. From subsequent above tasks, the training and the testing samples for the study area was accomplished. Different parameters in training process were inputted. Construction of the prediction model was done with the best parameters and thus the model was used to pick up the support vectors in the samples. We verified from various literature surveys that using Library Support Vector Machine (LIBSVM) for training the dataset sample would provide better results.

The screen layouts of the output and the graphs showing the comparison of ANN with SVM are also discussed in this chapter. The summary and conclusion of the project with the future enhancement is discussed in the next chapter.

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 SUMMARY

Natural hazards like landslides are beyond the control of humans but the destruction and loss of resources can be minimized if their occurrence is known before time. Innumerable researchers worldwide are striving hard to develop early indicators for such hazards. This project proposed an approach to predict the risk level of landslide. In this project, we have used Support Vector Machine (SVM) for training the dataset and for testing process, where the radial basis function is a major functionality that must be used in the SVM to implement the training process and testing for the landslide dataset. The Support vector Machine

classifier automatically classifies the landslide. SVM classifier accurately classifies the user input and predicts the landslide. Our climatic threshold method predicts the risk level. Thereby concludes that SVM proved to be an efficient technique to forecast the landslides by predicting the rainfall in advance. The study has been done specifically for Cherrapunjee region and can be implemented for any landslide prone area.

6.2 FUTURE ENHANCEMENT

The application of the landslide prediction model is an exploration always and it includes many things to be researched further. The future scope of the work is in providing an initial selection of the kernel parameter in order to save the time on finding the best parameters.

REFERENCES

- 1.a) Anbarasu K, Sengupta A, Gupta S, and Sharma SP(2010), “Mechanism of of the LantaKhola landslide in Sikkim Himalayas”, *Landslides*, vol. no. 7, pp. 135-147.
- 1.b) Anbarasu K , Sengupta A, Gupta S, (2010), “Rainfall thresholds for the initiation in north Sikkim, India”, *National Hazards*, vol. no. 52, pp. 31-42
2. Biswajeet Pradhan and Saro Lee(2009), “Landslide risk analysis using artificial neural network model focusing on different training sites”, 2009, *International Journal of Physical Sciences*, January, Vol. 4 (1), pp. 001-015
3. Chandrasekaran S.S, Elayaraja S, and Renugadevi S(2011), “Damages to transport facilities by rainfall induced landslides during November 2009 in Nilgiris, India,” *Proceedings of the Second World Landslide Forum, Rome*.
4. National disaster management guidelines – “Management of landslide and snow avalanches”, 2009. A publication of the national disaster management authority, Government of India, June 2009, New Delhi.
5. Renuga Devi S, Pranay Agarwal, Venkatesh C, Arulmozhivarman P(2014), “Daily Rainfall Forecasting using Artificial Neural Networks for Early Warning of Landslides”,*IEEE,International Conference on Advances in Computing, Communications and Informatics*.
6. Seema Mahajan, Himanshu Mazumdar(2013), “Rainfall Prediction using Neural Net based Frequency Analysis Approach”, *International Journal of Computer Applications* (0975 – 8887) Volume 84 – No 9, December

7. Xiaowen Zhao a, Min Ji a, Xianguo Cui a(2010), "Research On Landslide Prediction Model Based On Support Vector Model", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part II
8. Yao X, Dai F.C(2006), "Support vector machine modelling of landslide susceptibility using a GIS: A case study", The Geological Society of London 2006, IAEG Paper number 793
9. <http://www.cherrapunjee.com/weather-info/daily-weather-data-2009-2013/>
10. <http://www.cherrapunjee.com/weather-info/daily-weather-data-2014/->

APPENDIX

a. SAMPLE CODE

```

<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1, minimal-ui">
<title>Enhanced Landslide Prediction System using Support Vector Machine</title>
<%
String msg=(String)request.getSession().getAttribute("msg");
if(msg==null){
    msg="";
}
%>
</head>
<body>
<div class="row"></div>
<div id="menu-container">
<div id="menu-1" class="homepage home-section container">
<div class="home-projects">
<div class="row">
<div class="project-title animated fadeInUp">
<br/>
<br/>
<h2><span class="post-1"><a href="fileUpload.jsp" class="btn">

```

Dataset Preparation and Preprocessing

 SVM Training

</h2></div> </div>

</div>

</body>

</html>

a. PUBLICATIONS

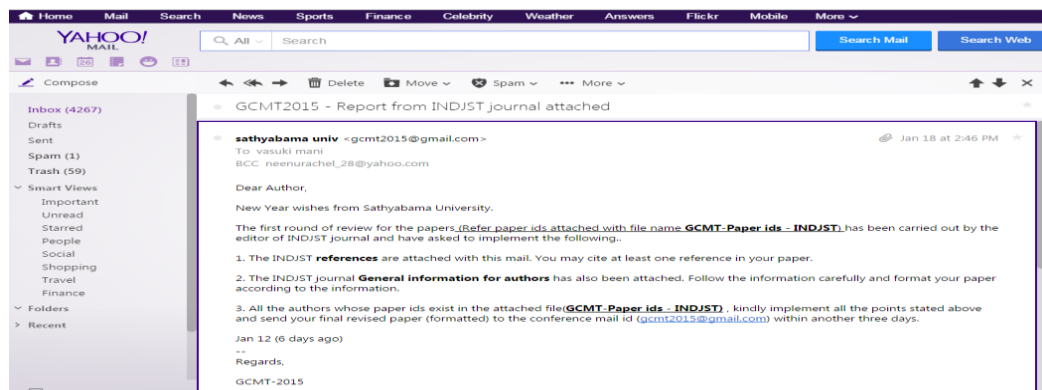
1. Acceptance Letter



2. Certificate of Presentation



3. Publication details



Accepted by the Indian Journal Of Science and Technology for publishing

4. Publication paper

Global Congress on Computing and Media Technologies (GCMT-2015)

Landslide Prediction with Rainfall Analysis using Support Vector Machine

Neenu Rachel⁽¹⁾ and Dr.M.Lakshmi⁽²⁾

neenurachel_28@yahoo.com, laks@icadsindia.com

⁽¹⁾PGStudent, Department of Computer Science and Engineering, Sathyabama University, Chennai

⁽²⁾Professor and Head, Faculty of Computing, Sathyabama University, Chennai

Abstract

Objective:

The paper aims in presenting a prediction model by using Support Vector Machine (SVM) technique which is meant to possess a strong capability to predict landslides by forecasting rainfall dataset using BigData concept.

Methods

The dataset has been taken for the Cherapunjee region which receives the highest intensity of rainfall in India. The aim is to predict the landslide occurrence and classify the risk level associated with the landslide. To improve the reliability in landslide prediction, the proposed model uses pre-processing for removing null values in the dataset. After getting the pre-processed dataset, it will apply normalization, then SVM training and finally the Testing process. Thus the Support Vector Machine concept proved to exhibit a large degree of flexibility in handling tasks of varied complexities because of the non-linear boundary functions.

Findings

The study concludes that SVM proved to be an efficient technique to forecast the landslides by predicting the rainfall in advance. The comparative results of SVM in regard with Artificial Neural Networks were proven. The study has been done specifically for Cherrapunjee region and can be implemented for any landslide prone area

Novelty/Improvement

Researchers worldwide are having a great pace to develop early prediction mechanisms for natural hazards. The study uses Radial Basis Function as an initial parameter for predicting the risk level classification of landslide. The novelty is in providing an initial selection of the kernel parameter in order to save the time on finding the best parameters.

Keywords: SVM, BigData, Rainfall Data, Hadoop

-

1.Introduction

The rainfall induced hazard has seriously influenced the sustainable development in many parts of the country. The effect of damages and losses mainly incurred regularly as there has been very less considerations of the potential problems in the planning of land usage and slope stability management. The landslide term can be describes as the movement of the slope downwards along with debris due to a sudden outbreak in the rainfall leading to mass flow of land and other substances. Landslide occurrence can be defined as a function of both temporal and spatial variability of the site scenarios ¹.

The regional precipitation induced landslides in Uttharakhand, Meghalaya, Sikkim and the Western Ghats ^{2, 3, 4}. A statistical approach to slope stability analysis with the strength reduction methodology is carried for the Nilgiri region⁴. The authors have proposed for a

modular ANN mechanism which is coupled with the data-pre-processing strategies. It aims in improving the accuracy of rainfall forecasting. MANN is comprised of 3 local models which are part of the 3 subsets being clustered by the Fuzzy C-means (FCM) clustering technique.

The recent occurrence with the highest devastation which occurred in Utharakhand in 2013 caused by heavy rainfall took away the lives of more than thousand lives. Landslide phenomenon presents a significant pronouncement during the monsoon period ⁵.

The researchers have employed various models which include both empirical as well as mathematical concepts such as ANN to predict rainfall intensity and access the risk of landslide occurrence in comparison with the rainfall threshold. But here the prediction accuracy is observed to be low and computational time is high.

With the advancement of computational efficiency which when combined with several sophisticated statistical approaches⁶, various techniques have been used for the prediction of rainfall induced landslides. The approach of a linear system is quite complex when climatic phenomenon are considered. The ANN which was invented in 1964 was made use in landslide prediction using the daily rainfall analysis. The back propagation algorithm was made used for the mapping of the input to the output with the appropriate back propagation of errors .

Several risk analysis are performed using the remote sensing data. The authors have used the Digital Elevation Model technique ⁷ to explore the most hazardous zones of the landslides area. The DEM has been used to calculate the various factors such as slope aspects, slope angle and the slope curvature.

An extensive study on the geological, geotechnical and the geophysical investigations ⁸ were carried out to study the major cause of the landslide activation in the Lanta Khola region of the Sikkim Himalayas. The authors have strengthened their model by proving that the sliding and the debris generation are initiated in the zone only after the heavy rainfall ⁹.

An Artificial Neural Network model which is categorized on frequency analysis using the Fast Fourier Transform (FFT) technique¹⁰ has been used. The study also has examined the roles of Feed Forward Neural Network (FFNN) for the rainfall prediction. A detailed accuracy classification of the prediction of landslide occurrence has been depicted using SVM and ANN technique

Thus several mathematical and empirical models have been extracted to achieve different precision in several regions across the world. Many studies are also concentrated on both the temporal as well as spatial parameters. The paper exhorts a cluster validity index for

rough fuzzy c-means¹¹ clustering algorithm called rough fuzzy Bayesian like validation method which roots on probabilistic metric.

The very motive is to develop a landslide prediction model and generating the risk level associated with the landslide. This can increase the preparedness level to reduce the severity of the disaster thereby having a control on the human loss and also the infrastructural damages. As far as this area of study is concerned, there is no landslide prediction system in this region. This paper employs the Support Vector Machine Concept to develop an early warning approach to the landslide. The risk level (high, medium, low) is also evaluated in this paper.

2. Methodology

2.1. Research motivation

The existing system uses artificial neural network to predict the rainfall intensity, one day in advance. The daily threshold values are used for comparison with the rainfall intensity to forecast the occurrence of landslide. ANN removes the complex parameters in the physical process and thus considered to be black-box. The enhancement of performance and volume management are the 2 major goals of research works in the domain of Bigdata.

2.2. Support Vector Machine

Support Vector Machine (SVM) originally developed by Vapnik, in 2005 is considered to be a new generation of learning algorithms. The Support Vector Machine concept has been used in accordance with the constructive machine learning procedure based on the statistical learning theory. It provides non-linear solutions to regression as well as classification problems by transforming the inputs which occupy a large dimension space. In this paper, SVM model has been used for both training and testing process.

2.3. Application of SVM in the study area

The forecasting of landslide has been dealt using factors which affects the landslide occurrence. We have used the training of data samples to determine several classes for prediction and to remove the null values. Also, the testing process has been done to improve efficiency and accuracy.

2.3.1 Dataset

The Cherrapunjee region of Meghalaya has been selected for the study which has the maximum intensity of rainfall in the country. The dataset comprises of 7 years daily recorded data for rainfall from 2009 to 2015^{12, 13}

The sample dataset is given below:

	A	B	C	D	E	F	G	H
1								
2	Date	Min Temp (Celsius)	Max Temp (Celsius)	Morn. Air Pressure (mb)	Even. Air Pressure (mb)	Rainfall (mm)	Annual Rainfall (mm)	Wind Speed (k/h)
3	17/08/2015	18.3	NA	866.9	NA	162.2	8388.2	NA
4	16/08/2015	18.3	21	867.8	865.4	194.9	8226	NA
5	15/08/2015	18	21.2	867.8	865.3	191.2	8031.1	NA
6	14/08/2015	17.8	22.9	866.2	865	101.8	7839.9	NA
7	13/08/2015	18.5	24.2	867.8	864.7	4	7738.1	NA
8	08-12-2015	19	25	868.4	865.6	0	7734.1	NA
9	08-11-2015	19.5	26.5	868.1	865.1	20.6	7734.1	NA
10	08-10-2015	19.4	21.8	865.4	862.9	1.8	7713.5	NA
11	08-09-2015	19.1	26.6	864.2	861.4	0	7711.7	NA
12	08-08-2015	18.4	25.9	866	863.1	0	7711.7	NA
13	08-07-2015	17.5	26	864.8	863.2	0.8	7711.7	NA
14	08-06-2015	17.1	24.3	864	860.2	6.8	7710.9	NA
15	08-05-2015	17.5	20.8	864.5	861.7	58.6	7704.1	NA
16	08-04-2015	18.3	22	864.8	863.2	4	7645.5	NA
17	08-03-2015	17.9	24.2	866.9	864	0	7641.5	NA
18	08-02-2015	17.4	23.2	867.7	864.8	27.9	7641.5	NA
19	08-01-2015	17.4	19.3	864.7	864.8	85.4	7613.6	NA
20	31/07/2015	17.3	20.2	859.5	860.1	10.4	7528.2	NA
21	30/07/2015	21.2	26.9	859.9	857.7	0	7517.8	NA
22	29/07/2015	20.6	29.2	861.6	859.1	0	7517.8	NA
23	28/07/2015	19.4	27.2	862.9	860.4	0	7517.8	NA
24	27/07/2015	19	26.8	862.7	861.1	0.1	7517.8	NA
25	26/07/2015	18.9	26.6	863.3	861.9	0	7517.7	NA
26	25/07/2015	16.6	24.6	864.5	862.7	74.7	7517.7	NA
27	24/07/2015	17.9	22.2	867.3	864.3	258.6	7443	NA
28	23/07/2015	17.9	21.8	867.7	865.5	51.2	7184.4	NA

Fig 1: Sample dataset

2.3.2 Selection of parameters

The several other parameters in concern with the rainfall data included the maximum and minimum temperature, air pressures during morning and evening and wind speed. A proper class based approach has been taken to distinguish the several parameters.

To build a SVM model with much efficiency and accuracy, the parameters (C and γ) of the prediction model need to be chosen in advance. The parameter C is used to determine the trade-off cost in between for minimizing the error in training and also the complexity of the SVM model. The predictive accuracy of the training sample will be higher if the value of C increases. This however might cause an over-training issue. The parameter used in the Radial Basis Function determines a non-linear mapping from the input space to the high-dimensional feature space. Shape of RBF function is determined by the value of γ . Therefore, the parameters (C and γ) have a strong influence on the generalization performance and efficiency of SVM model.

Computing procedure

The procedure of computing the landslide prediction model is accomplished using the following steps:

- Transform data to remove null values
- Perform scaling on the transformed data
- Consider the RBF kernel $K(X_i, X_j) = \exp(-\gamma \|x_i - x_j\|^2)$
- Find the best parameter C and γ using cross-validation
- Train the whole training set by using the best parameter C and γ
- Test the data

3. Algorithm

3.1 Landslide prediction algorithm

The landslide prediction algorithm has three phases, the algorithm for predicting landslide is briefly explained below

Algorithm 1: Landslide_prediction_Algorithm

Input: Climatic data from the Indian Meteorological Observatory

Output: Display the landslide occurrence dates and classify the risk level of it.

The main Detection procedure for landslide prediction algorithm is given as:

```
23. Start
24. {
25. Rainfall dataset as an input
26. Initial input size
27. Set threshold value of Rainfall
28. initialize threshold to zero
29. {
30. Initialize the input format
31. Training of the rainfall dataset with SVM
32. Calculate the error rate of trained dataset
33. Testing of the dataset with SVM
34. Normalization of Trained dataset
35. Set maximum and minimum fraction
36. Start the Yearly evaluation of the landslide
37. End the yearly evaluation of the landslide
38. {
39. Compute Threshold
40. Predict the risk level of land slide
41. }
42. }
43. }
44. End
```

- Landslide Stage Identification

1) Landslide Dataset Preparation and Pre-processing

The Prediction of landslide based on the previous year's data and records. In first step, for predicting the landslide, need to collect all the previous record for pre-processing function, which makes a specific result for the future Prediction. The pre-processing method removes the null value.

The proposed work has been implemented as in 4 modules. They are described as follows:

- Landslide Dataset Preparation and Pre-processing
- Normalization and SVM Training
- Landslide Prediction using SVM

2) Normalization and SVM Training

The support vector machine (SVM) is training the dataset by using Radial Basis Function. Radial Basis Function is one of the major kernel functions that should be utilized for SVM training process. The trained data is classified and tested by SVM. Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training.

3) Landslide Prediction using SVM

The users make a request for the testing process by Support Vector Machine technique. The testing is essential process for predicting the landslide based on the classification. To make the classification results comparable and for exhaustive data analysis, we have used leave one out classification method for the SVM classifier. In classification, SVM efficiently classifies the landslide dataset.

4) Landslide Stage Identification

Predicted landslide image taken as an input for this module. In this module, analyse the risk level of future prediction landslide dataset. Consider three types of risk level. The climatic threshold based mechanism attempts to classify the dataset into 3 stages: low, medium, high. Finally display the risk level of landslide such as (low, medium and high).

4. Architecture

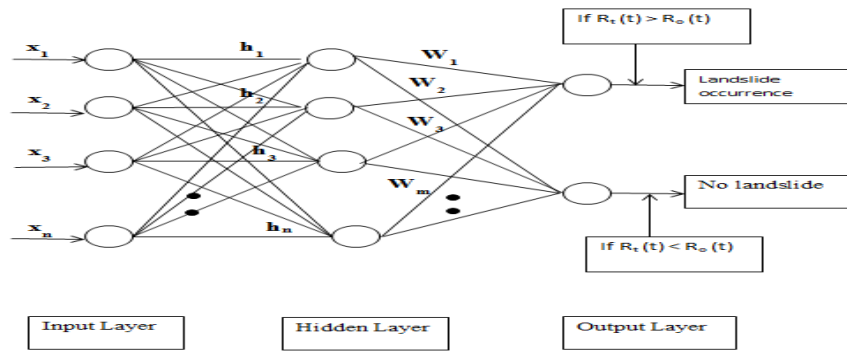


Fig 2: SVM Architecture
The parameters taken as input includes $X_1, X_2, \dots, X_n$ where n represents the number of input nodes. The Rainfall Threshold for the day is represented as $R_t(t)$ and the Rainfall Predicted for the day as $R_o(t)$.

5. Design and implementation:

5.1 Design

The design concept of the landslide prediction model has been explained in accordance with the algorithms mentioned above. Initially Collect Landslide Datasets from the user and store the datasets into database. Then apply pre-processing to replace all the null values (only applicable for numerical dataset). Pre-processing reduce the time complexity. Then pre-processed data were taken as an input for training process.

Apply normalization process for normalizing the pre-processed dataset. Radial Basis Function is one of the major kernel functions is utilized for SVM training process.

In the testing phase, Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training. Then proceed by taking the query dataset for testing phase. Apply pre-processing for input query dataset.

Support Vector Machine compares the query dataset with trained dataset. It displays the predicted status (landslide or not). Apply climatic threshold based prediction mechanism if the classification result is relevant to landslide. Finally display the risk level of landslide such as (low, medium and high)

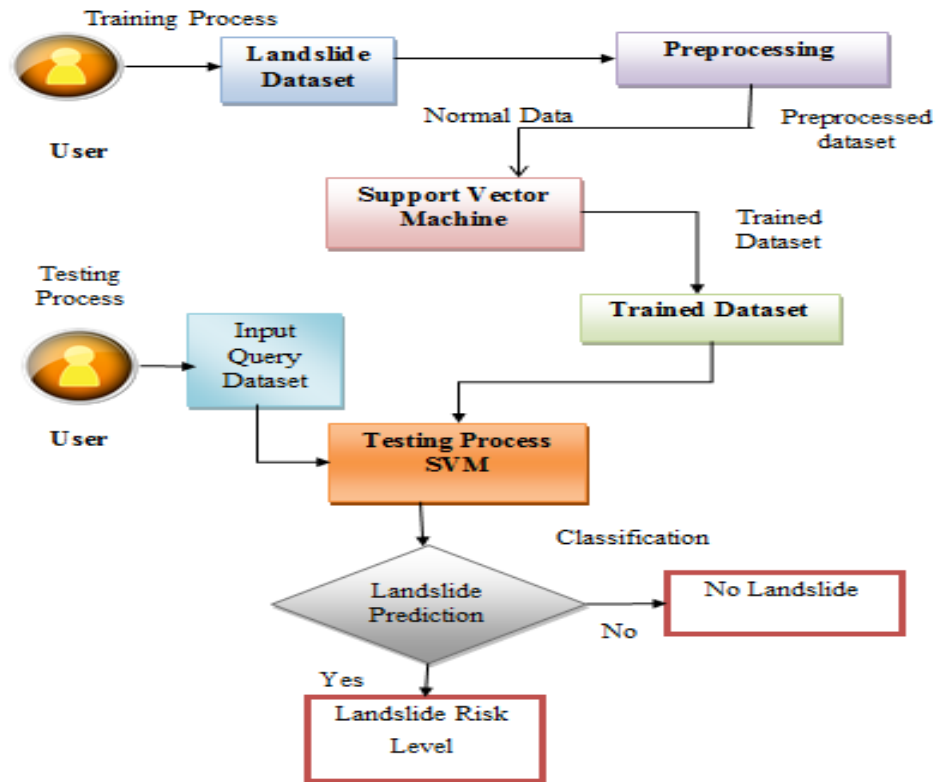


Fig 3: Flow diagram of Landslide prediction

5.2 Implementation

The proposed algorithm is implemented using hadoop with java language. The algorithm is implemented considering the nature of the spatial database. The details of the nodes are stored in the hive database. Hadoop is an open-source software framework that supports data-intensive distributed applications to facilitate the storage and processing of large and rapidly growing data sets. Since it is highly scalable, they are used with increased no of attributes.

8. Results

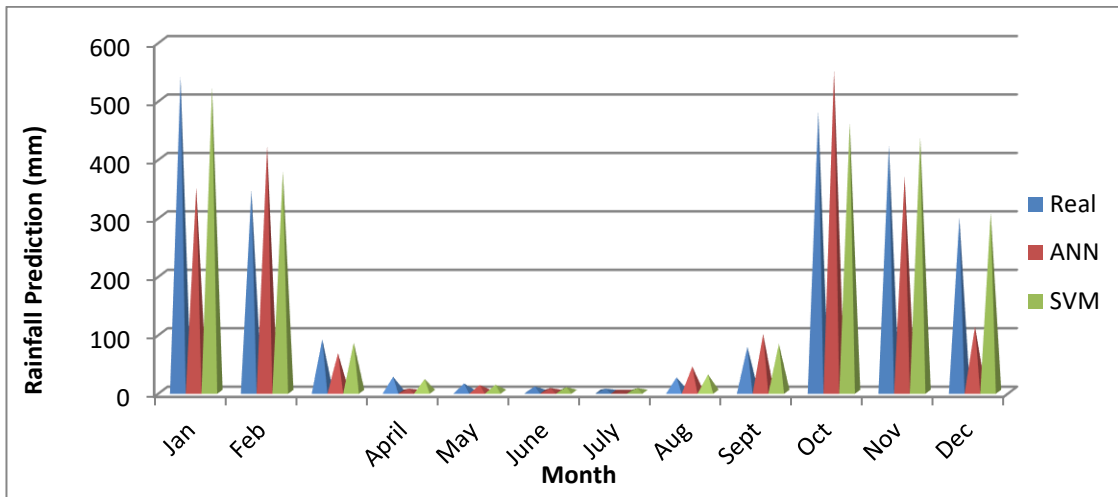


Fig 4: Comparison with Real, ANN, SVM

The above graph is showing the prediction of rainfall comparison with real, ANN and proposed technique SVM. The result is showing that the proposed technique SVM is better than ANN and it produces more relevant and approximate result.

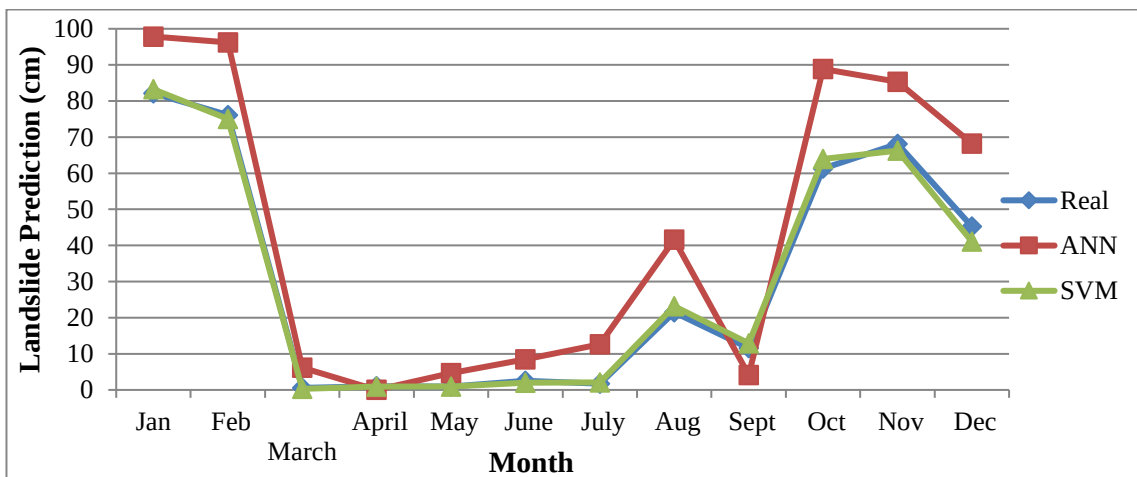


Fig 5: Comparison with Real, ANN, SVM

The above graph is showing the landslide prediction accuracy within the rainfall condition, the proposed technique SVM is producing more relevant and close to exact or real prediction of the landslide. From subsequent above tasks, the training and the testing samples for the study area was accomplished. Different parameters in training process were inputted. Construction of the prediction model was done with the best parameters and thus the model was used to pick up the support vectors in the samples. We verified from various literature surveys that using Library Support Vector Machine (LIBSVM) for training the dataset sample would provide better results.

6. Conclusion and future work

The natural hazards are beyond human control but their destruction can be reduced if prediction mechanisms are carried out in advance. Researchers worldwide are having a

great pace to develop early prediction mechanisms for such natural hazards. It was hard to use traditional mathematical methods for analyzing. Thus SVM helped in getting non-linear relationships using the historical data.

Thereby, the study concludes that SVM proved to be an efficient technique to forecast the landslides by predicting the rainfall in advance. The study has been done specifically for Cherrapunjee region and can be implemented for any landslide prone area

The application of the landslide prediction model is an exploration always and it includes many things to be researched further. The future scope of the work is in providing an initial selection of the kernel parameter in order to save the time on finding the best parameters.

Acknowledgement

The access to rainfall dataset of Cherrapunjee datasets from the Indian Meteorological Observatory has been acknowledged. We are also thankful to the Sathyabama University for encouragement.

References

1. S. Renuga Devi, Pranay Agarwal, C. Venkatesh, P. Arulmozhivarman, "Daily Rainfall Forecasting using Artificial Neural Networks for Early Warning of Landslides", 2014 IEEE, International Conference on Advances in Computing, Communications and Informatics.
2. K. Anbarasu, A. Sengupta, S. Gupta, and SP Sharma, "Mechanism of activation of the Lanta Khola landslide in Sikkim Himalayas", Landslides, vol. no. 7, pp. 135-147, 2010.
3. A. Sengupta, S. Gupta, and K. Anbarasu, "Rainfall thresholds for the initiation of landslide at Lanta Khola in north Sikkim, India," National Hazards, vol. no. 52, pp. 31-42, 2010.
4. S. S. Chandrasekaran, S. Elayaraja, and S. Renugadevi, "Damages to transport facilities by rainfall induced landslides during November 2009 in Nilgiris, India," Proceedings of the Second World Landslide Forum, Rome, 3-7 October 2011.
5. National disaster management guidelines – Management of landslide and snow avalanches, 2009. A publication of the national disaster management authority, Government of India, June 2009, New Delhi.
6. Xiaowen Zhao a, Min Ji a, Xianguo Cui a, " Research On Landslide Prediction Model Based On Support Vector Model", 2010, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part II

7. Biswajeet Pradhan and Saro Lee, “Landslide risk analysis using artificial neural network model focusing on different training sites”, 2009, International Journal of Physical Sciences, January, 2009, Vol. 4 (1), pp. 001-015
8. K. Anbarasu, A. Sengupta, S. Gupta. S. P. Sharma, “Mechanism of activation of the Lanta Khola landslide in Sikkim Himalayas”, 27 January 2010, Springer-Verlag 2010, Landslides (2010) 7:135–147
9. Seema Mahajan, Himanshu Mazumdar, “Rainfall Prediction using Neural Net based Frequency Analysis Approach”, International Journal of Computer Applications (0975 – 8887) Volume 84 – No 9, December 2013
10. X. Yao, F.C. Dai, “Support vector machine modelling of landslide susceptibility using a GIS: A case study”, The Geological Society of London 2006, IAEG2006 Paper number 793
11. S. Revathy, B. Parvaathavarthini, S. Rajathi, “Futuristic Validation Method for Rough Fuzzy Clustering”, Indian Journal of Science and Technology, 2015 Jan, 8(2), Doi no:10.17485/ijst/2015/v8i2/58943
12. Cherrapunjee dataset: <http://www.cherrapunjee.com/weather-info/daily-weather-data-2009-2013/> Date accessed: 11/09/2015
13. Cherrapunjee dataset: <http://www.cherrapunjee.com/weather-info/daily-weather-data-2014-/> Date accessed: 11/09/2015