VIVEKANAND EDUCATION SOCIETY'S

INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

Catchment Control and Water Supply Management

In partial fulfillment of the Fourth Year (Semester–VIII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2018-2019

Project Mentor

Richard Joseph, Assistant Professor

Submitted by
(D17C)
Samujjwaal Dey - 15
Ashish Joshi- 23
Gurpreet Singh - 39
(2018-19)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that ***Samujjwaal Dey, Ashish Joshi, Gurpreet Singh*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "***Catchment Control and Water Supply Management***" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Prof.Richard Joseph*** in the year 2018-2019 .

This thesis/dissertation/project report entitled ***Catchment Control and Water Supply Management*** is approved for the degree of B.E. in Computer Engineering.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:
Project Guide:

Project Report Approval

For

B. E (Computer Engineering)

This project report entitled **Catchment Control and Water Supply Management** by **Samujjwaal Dey, Ashish Joshi, Gurpreet Singh Nagpal** is approved for the degree of **B.E, Computer Engineering**.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---------- ----------

(Signature) (Signature)

---------- ----------

(Samujjwaal Dey (D17 - C , Roll no - 15)) (Ashish Joshi (D17 - C , Roll no - 23))

(Signature)

(Gurpreet Singh Nagpal (D17 - C , Roll no - 39))

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mr. Richard Joseph** (Project Guide) for his kind help and valuable advice during the development of project synopsis and for his guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO 2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

PROJECT ABSTRACT

There has been a significant amount of change in the state of water bodies in the past decade. The shrinkage of the water bodies due to urbanization and changes in the climatic conditions has led to the scarcity of water. Hence there is a need of other alternatives to meet the need of water. One of the ways is to increase the water catchment area in a particular region. Our system is based on this approach, it will use Machine Learning to predict areas which could be used for the catchment purpose and will lead to optimum distribution of water.

This project focuses on the issues related to water scarcity in the rural parts of Maharashtra , where there is limited water resources like lakes and rivers. The ground water level of these areas have also reduced drastically on account of various environmental issues like global warming, etc. The project aims at increasing the catchment areas of Maharashtra that are hit by drought like conditions and thereby increasing the ground water level, which will eventually help in the long run. We aim at providing solution to make a region more sustainable.

INDEX

1. Introduction	**12**
1.1. Introduction to the project	**13**
1.2. Motivation for the project	**13**
1.3. Drawback of the existing system	**13**
1.4. Problem Definition	**14**
1.5. Relevance of the Project	**14**
1.6. Methodology	**14**
2. Literature Survey	**16**
2.1. Papers Reviewed	**17**
2.2. Articles Referred	**21**
2.3. Patent Search	**22**
3. Requirement Of Proposed System	**23**
3.1 Functional Requirements	**24**
3.2. Non-Functional Requirements	**24**
3.3. Constraints	**25**
3.4. Hardware & Software Requirements	**25**
3.5. Techniques, Tools and Algorithms used till date	**26**
3.6. Project Proposal	**26**
4. Proposed Design	**27**
4.1 Block Diagram	**28**
4.2 Modular Design	**29**
4.3 Detailed Design :	**30**
a. Data Flow Diagram (Level 0,1)	**30**
b. Flowchart for the proposed system	**32**
4.4. Project Scheduling & Tracking using Gantt Chart	**33**
5. Implementation Details	**34**
5.1. Algorithms considered for study	**35**

5.2. Comparative Analysis with the existing algorithms	**41**
6. Testing	**42**
6.1 . Definition of testing	**43**
6.2. Types of tests	**43**
6.3. Type of Testing considered with justification	**46**
6.4 Various test case scenarios considered	**53**
6.5. Inference drawn from the test	**53**
7. Result and Analysis	**54**
7.1. Module(s) under consideration	**55**
7.2. Screenshots of User Interface (UI) for the respective module	**56**
7.3. Evaluation of the developed system	**60**
7.4. Graphical outputs of the various scenarios considered	**61**
8. Conclusion	**62**
8.1 Limitations	**63**
8.2 Conclusion	**63**
8.3 Future Scope	**63**
9. References	**64**
9.1. Newspaper articles referred	**65**
9.2. Research Papers Referred	**65**
9.3. Patents referred	**66**
10. Appendix	**67**
10.1. Paper I	**68**
a. Paper published	**68**
b. Plagiarism report and Certificate	**73**
c. Project review sheet	**75**

LIST OF FIGURES

Figure Number	Heading	Page no.
Fig 1	Block Diagram	28
Fig 2	Modular Design	29
Fig 3	Level 0 DFD	30
Fig 4	Level 1 DFD	31
Fig 5	Flowchart	32
Fig 6	Gantt Chart	33
Fig 7	K Means implementation	35
Fig 8	Satellite image before highlighting water body	36
Fig 9	Satellite image after highlighting water body	36
Fig 10	Satellite image before highlighting vegetation	36
Fig 11	Satellite image after highlighting vegetation	36
Fig 12	Gathering elevation for given place	36
Fig 13	Gathering weather-data for given place	37
Fig 14	Initial accuracy	47
Fig 15	Accuracy after removing “Latitude” and “Longitude”	48
Fig 16	Accuracy after removing “Elevation”	49
Fig 17	Accuracy after removing “Soil Score”	50
Fig 18	Accuracy after removing “Groundwater”	51
Fig 19	Accuracy after removing “Rainfall”	52
Fig 20	Single Input	56

Fig 21	Single input result	56
Fig 22	Single input result on map	57
Fig 23	Batch input	57
Fig 24	Batch input result	58
Fig 25	Batch input result on map	59
Fig 26	Accuracy of trained model	60
Fig 27	Dataset	61
Fig 28	Rainfall vs Monsoon Groundwater K-means clustering	61
Fig 29	Rainfall vs Pre-Monsoon Groundwater K-means clustering	61
Fig 30	Plagiarism Report	73
Fig 31.	Certificate of Paper Publication	74

LIST OF TABLES

Table Number	Table name	Page no.
Table 1	Comparison of different algorithms considered for study	41
Table 2	Impact analysis of attributes and comparison with weights assigned previously	52
Table 3	Comparison of training and testing accuracy of algorithms considered for study	53

CHAPTER 1
INTRODUCTION

1.1. INTRODUCTION

There has been a significant amount of change in the state of water bodies in the past decade. The shrinkage of the water bodies due to urbanization and changes in the climatic conditions has led to the scarcity of water. Hence there is a need of other alternatives to meet the need of water .One of the ways is to increase the water catchment area in a particular region. Our system is based on this approach,it will use AI to predict areas which could be used for the catchment purpose and will lead to optimum distribution of water.This project is proposed with the idea to analyse the difference between the past and present condition of the water bodies and provide a solution to the water supply problem by detecting areas suitable for catchment

1.2. MOTIVATION

The Marathwada region of Maharashtra, a perennially drought-plagued area, is once again staring at a severe water scarcity. Over 41 percent of the villages in this region of central Maharashtra have reported 'average yield' below 50 paise. This will force the government to mull augmenting foodgrain and water supply in the region. A data compiled by the revenue department revealed that more than 3,500 villages in Marathwada have reported average yield less than 50 paise.This means the scarcity of basic resources such as foodgrain production and water is less than 50 percent of its actual capacity..Erratic showers are generally the main cause of less average yield, a senior revenue official said.Of the 153 days of the monsoon in Marathwada last year, 94 days were dry. This clearly means there were more dry days in the monsoon season than wet. There are many such reasons in India that are agricultural and need a good supply of water but due to scarcity, these areas are becoming barren and the conditions are not worth living for any living creature. The biggest challenge in such areas is to harvest the water and to increase the number of catchment areas that can lead to more storage of water.

1.3 DRAWBACKS OF THE EXISTING SYSTEM

The existing system requires extensive manual work to perform land surveys of regions to identify drought prone regions and diminishing water bodies. There have been significant changes in the water bodies which have not been frequently surveyed in regular intervals. The unchecked expansion of land usage for various purpose has increased which has led to water body depletion. The existing system is not very accountable as it is not transparent and can be a victim of corruption. No solution has been provided for this problem and as a result drought conditions are fairly common in interior parts of Maharashtra.

1.4 PROBLEM DEFINITION

The magnitude of climate change in a particular region impacts the local soil properties like water retention capacity, rate of evaporation water, etc. These along with other factors like altitude and average humidity directly impact the amount of water present in aquifers and surface water catchment of the region. Thus we aim to build a system to successfully identify actual and potential drought prone areas of Maharashtra from remote sensed satellite images. Taking Machine Learning as base we propose a system that would predict that if a particular region has or will have drought in due course of time. The system will also identify a suitable area for water catchment control around such regions.

1.5 RELEVANCE

This project focuses on the issues related to water scarcity in the rural parts of Maharashtra, where there are limited water resources like lakes and rivers. The ground water level of these areas are prone to reducing drastically on account of various environmental issues like global warming, etc. The project aims at increasing the catchment areas in places which have actual or potential drought like conditions and thereby increasing the ground water level, which will eventually help in the long run. We aim at providing solution to make such regions more sustainable. Such areas not only lack rainfall but also the way the water is stored and its supply managed.
These results would not only benefit our work, but also the work of thousands of other practitioners and researchers,who find it difficult to analyse the changes in the environment. The application is intended to be open-source.

1.6 METHODOLOGY

The proposed system works in the following phases:

A. Data Gathering and Mapping

This step is about gathering the data related to environmental factors from different sources. This data is collected w.r.t. Latitude and Longitude, to understand the environment of a particular location. Here we gathered data related to following parameters: Rainfall, Temperature, Soil Type, groundwater level, Humidity and Altitude. The data for parameters temperature and altitude was collected with the help of weather API's. Whereas the data for parameters humidity, soil type, rainfall and groundwater level has been collected from Government weather portals and Indian Meteorological Department datasheets.

B. Data Pre - Processing and Feature Extraction

Data extracted from API's won't need much pre - processing and can be directly used as features. It will just need to be transformed into a suitable format for further analysis. The

rainfall data obtained was monthly data from 1996 to 2017 with a number of missing values. Firstly missing values were handled by least square regression. Further for clustering we used mean value of rainfall over the years for a district. For classification we required yearly total from 1996 to 2017 for each district of Maharashtra. Similar set of operations were conducted for groundwater data

C. Identification of Drought Prone Areas

Feature extracted will now be used to train the neural network algorithm to identify the drought prone areas. In this section, we present how the Two Class Support Vector Machine will train the model to identify the drought prone regions w.r.t. Latitudes and Longitudes.

We use neural networks as its goal approaches to combinatorial optimization to formulate the desired objective function being optimized, such that it can be viewed as a "natural" energy minimization problem. We have many parameters that are very closely related in this scenario for example the evaporation rate is directly proportional to temperature, Humidity in soil is related to soil type, etc. All of these parameters with combinations will give different results and we need an optimized result that would help us detect the drought prone area.

D. Identification of areas suitable for catchment

As we identify the areas that are hit by drought, our research to find a source of water starts exactly from the point where we found the hit for drought. We can help such areas in two ways. We can either help such areas to increase the goud water level or we can build some kind of reservoir for the area. All of these depends on many factors that will eventually help to harvest and conserve water for long duration of time. There are several factors that leads to drought like conditions. Few are natural and few are man made. With the help of unsupervised learning and by using competitive learning rule, we can come out with combinations of parameters that might give us the best result for the region as all the combinations that are considered as output compete with each other and finally the winner takes all concept is implemented.

CHAPTER 2
LITERATURE SURVEY

2.1 PAPERS REVIEWED

1) Application of multispectral remote sensing technology in surface water body extraction

Authors: Zhang Ding , Ni Qi , Fang Dong , Li Jinhui , Yao Wei , Yuan Shenggui

Abstract: This paper addresses an improved water-body extraction method by the multispectral remote sensing image of four-bands. Firstly, traditional multispectral water body extraction technologies are reviewed, and their features are compared. It shows that finding the most difference of the reflection ratio of the water body and the background is the core to improve the identification rate. Furthermore, combining data's feature of GF-01 and GF-02 satellites, an improved normalization index method by introducing the weighted average of the blue and green light bands was proposed. Finally, the comparative results illustrating the improvement of water body extraction are provided and analyzed.

Inference: On the basis of reviewing multispectral water body extraction technology, physical mechanism, performance feature, and existing defect of single band, multiband, normalization index method, and other typical extraction methods are analyzed. Then an improved normalization index method was proposed. Compared with the original methods, the method can more accurately and effectively make use of water body strong reflectivity band information. Furthermore, it inherits and adopts weak reflectivity band and normalization calculation form to improve identification accuracy and to guarantee threshold stability. Finally, a comparative test was performed in Wuxi and Beijing with the high resolution multispectral images acquired by domestically made high resolution satellite, while advantages of the new algorithm were preliminarily demonstrated.

2) Decision Jungles : Compact and Rich Models for Classification

Authors: Jamie Shotton, Toby Sharp, Pushmeet Kohli, Sebastian Nowozin, John Winn, Antonio Criminisi

Abstract: Randomized decision trees and forests have a rich history in machine learning and have seen considerable success in application, perhaps particularly so for computer vision. However, they face a fundamental limitation: given enough data, the number of nodes in decision trees will grow exponentially with depth. For certain applications, for example on mobile or embedded processors, memory is a limited resource, and so the exponential growth of trees limits their depth, and thus their potential accuracy. This paper proposes decision jungles, revisiting the idea of ensembles of rooted decision directed acyclic graphs (DAGs), and shows these to be compact and powerful discriminative models for classification. Unlike conventional decision trees that only

allow one path to every node, a DAG in a decision jungle allows multiple paths from the root to each leaf. We present and compare two new node merging algorithms that jointly optimize both the features and the structure of the DAGs efficiently. During training, node splitting and node merging are driven by the minimization of exactly the same objective function, here the weighted sum of entropies at the leaves. Results on varied datasets show that, compared to decision forests and several other baselines, decision jungles require dramatically less memory while considerably improving generalization.

Inference: This paper has presented decision jungles as ensembles of rooted decision DAGs. These DAGs are trained, level-by-level, by jointly optimizing an objective function over both the choice of split function and the structure of the DAG. Two local optimization strategies were evaluated, with an efficient move-making algorithm producing the best results. Our evaluation on a number of diverse and challenging classification tasks has shown jungles to improve both memory efficiency and generalization for several tasks compared to conventional decision forests and their variants. Decision jungles can be extended to regression tasks.

3) A Generalized Flow for Multi-class and Binary Classification Tasks: An Azure ML Approach

Authors: Matthew Bihis, Sohini Roychowdhury

Abstract: The constant growth in the present day real-world databases pose computational challenges for a single computer. Cloud-based platforms, on the other hand, are capable of handling large volumes of information manipulation tasks, thereby necessitating their use for large real-world data set computations. This work focuses on creating a novel Generalized Flow within the cloud-based computing platform: Microsoft Azure Machine Learning Studio (MAMLS) that accepts multi-class and binary classification data sets alike and processes them to maximize the overall classification accuracy.

Inference: The MAMLS platform provides classification modules that can be fine-tuned and combined with dimensionality reduction and decision-making modules for classification tasks on scalable “Big Data” sets. In this work, we demonstrate and benchmark the performance of a novel Generalized Flow in the MAMLS platform that maximizes overall classification accuracies independent of computation device limitations.

4) Decision Forests, Convolutional Networks and the Models in-Between

Authors: Yani Ioannou, Duncan Robertson, Darko Zikic, Peter Kontschieder, Jamie Shotton, Matthew Brown, Antonio Criminisi

Abstract: This paper investigates the connections between two state of the art classifiers: decision forests (DFs, including decision jungles) and convolutional neural networks (CNNs). Decision forests are computationally efficient thanks to their conditional computation property (computation is confined to only a small region of the tree, the nodes along a single branch). CNNs achieve state of the art accuracy, thanks to their representation learning capabilities. We present a systematic analysis of how to fuse conditional computation with representation learning and achieve a continuum of hybrid models with different ratios of accuracy vs. efficiency.

Inference: This paper has investigated similarities and differences between decision trees/forests and convolutional networks. This has led us to introduce a hybrid model (namely conditional network) which can be thought both as: i) trees which have been augmented with representation learning capabilities, and ii) CNNs which have been augmented with explicit data routers and a rich, branched architecture.

5) Determination of The Water Catchment Area in Semarang City Using a Combination of Object Based Image Analysis (OBIA) Classification, InSAR and Geographic Information System (GIS) Methods Based On a High-Resolution SPOT 6 Image and Radar Imagery

Authors: Yudo Prasetyo, Setyo Ardi Gunawan, Zia Ul Maksum

Abstract: Semarang is the biggest city in central Java-Indonesia which has a rapid and massive infrastructure development nowadays. In order to control water resources and flood, the local government has been built east and west flood canal in Kaligarang and West Semarang River. One of main problem in Semarang city is the lack of fresh water in dry season because groundwater is not rechargeable well. Rechargeable groundwater ability depends on underground water recharge rate and catchment area condition. The objective of the study is to determine condition and classification of water catchment area in Semarang city. The catchment area conditions will be determine by five parameters as follows soil type, land use, slope, ground water potential and rainfall intensity. In this study, we use three methods approach to solve the problem which is segmentation classification to acquire land use classification from high resolution imagery using nearest neighborhood algorithm, Interferometric Synthetic Aperture Radar (SAR) to derive DTM from SAR Imagery and multi criteria weighting and spatial analysis using GIS method. There are three types optical image (ALOS PRISM, SPOT-6 and ALOS

PALSAR) to calculate water catchment area condition in Semarang city. For final result, this research will divide the water catchment into six criteria as follows good, naturally normal, early critical, a little bit critical, critical and very critical condition.

Inference: The result shows that water catchment area condition is in an early critical condition around 2607,523 Ha (33,17 %), naturally normal condition around 1507,674 Ha (19,18 %), a little bit critical condition around 1452,931 Ha (18,48 %), good with 1157,04 Ha (14,72 %), critical with 1058,639 Ha (13,47 %) and very critical with 75,0387 Ha (0,95 %). The distribution of water catchment area conditions in West and East Flood Canal have an irregular pattern. In northern area of watershed consists of begin to critical, naturally normal and good condition. Meanwhile in southern area of watershed consists of a little bit critical, critical and very critical condition.

6) Water bodies identification from multispectral images using Gabor filter, FCM and canny edge detection methods

Authors: T. Vignesh , K. K. Thyagharajan

Abstract: Water bodies identification using multispectral images is a very useful application of image processing. This paper proposed a novel method for water bodies identification from multispectral images using Gabor filter, Fuzzy c-means and canny edge detection algorithm. Gabor filter is a combination of low-pass filter and bandpass filter. This two filters extracting the importance features from satellite images. From the extracted features fuzzy c-means algorithm clustered the various land use and land cover classes. Finally water bodies are identified from land use and land cover classes with the use of canny edge detection methods. The proposed approach was experimented with the use of Landsat-7, Landsat-8 satellite images. Our experimental results proved that proposed methods provides better result for water identification with high efficiency.

Inference: In this paper efficient method for water bodies identification from land use and land cover classification has been proposed. The proposed method successfully extracted the water bodies with very low error rate and high accuracy specifically 1.94 % error rate for Landsat-8 data set and 2.01 % error rate for Landsat-7 dataset and it is proved that proposed method having high accuracy rate than other methods.

2.2 ARTICLES REFERRED

1) 12 districts in Maharashtra stare at drought this year

Twelve districts of Maharashtra — mainly in Marathwada and north Maharashtra — have received deficient rainfall this season and may once again face drought. The region was ravaged by drought in 2016 too.

The 12 districts are Aurangabad, Beed, Hingoli, Jalna, Nanded, Latur, Osmanabad, Parbhani, Dhule, Nandurbar, Ahmednagar and Jalgaon. Officials said some of these areas may start facing a shortage of drinking water within days. The state's relief and rehabilitation department has begun the process to declare the areas as drought-hit. The issue was discussed in the state cabinet meeting on Wednesday, when state revenue minister Chandrakant Patil said drought will need to be declared in the 12 districts. "We have already started the process of analysing the seriousness of the situation by conducting panchnamas. Drought will be declared by October 15," Patil told HT. He added that a detailed report will be sent to the Centre to get financial assistance. Between June and the end of September, Maharashtra got about 77% of its average rainfall, which means it is facing a 23% deficit.

However, there are 26 tehsils that received less than 50% of the average rainfall. Of them, seven are from Marathwada , and four tehsils are from north Maharashtra. For now, the state has deployed 329 tankers to supply drinking water to districts across Maharashtra. Of them, 182 tankers have been deployed for the Marathwada region alone; 118 tankers are supplying water to north Maharashtra. The maximum number of tankers 158 have been deployed to Aurangabad, according to statistics from the state water supply and sanitation department.

In terms of water stock, Marathwada has only 28.44% water in its dams. The Jayakwadi dam, the lifeline of Marathwada region, situated at Paithan, has just 42.51% water stock. Last year this time, it had more than 100% water stock.

2) Monitoring land use changes associated with urbanization: An object based image analysis approach

Land use/land cover (LULC) change occurs due to natural and anthropogenic causes. In developing countries, rapid industrialization and urbanization imposes a major threat to natural environment. Present study was carried out to monitor the LULC changes due to urbanization in a rapidly changing river basin, India. The purpose of choosing the river basin was to analyze past changes and predict possible consequences within a defined natural boundary. Multi-temporal images acquired from Landsat and Indian Remote Sensing (IRS) satellites between 1992-2009 as well as a digital elevation model was used to generate historical and current LULC pattern in the basin. An object based image

analysis technique was employed for precise classification of multi-temporal images followed by GIS-based change detection studies. The study reveals that the built up area has increased significantly and added 288 km2 between 1992 - 2009. Increase in built up area is attributed to decrease in wastelands and agricultural land. The expansion of built up area along major transportation networks, specifically after the year 2000 shows the rapid rate of urbanization in the basin.

2.3 PATENT SEARCH

1) Real-time detection method for drought status in detected area based on miniature unmanned plane

CN20151185314 20150418

Inventor(s): ZHENG ENHUI, FENG YIHUA, FU YAQIONG, CHEN LE

The invention discloses a real-time detection method for drought status in a detected area based on a miniature unmanned plane. By employing the miniature unmanned plane and a matched remote controller and a host computer, the host computer and the miniature unmanned plane are connected through a wireless transmission module, a cloud deck mechanism with a camera is loaded on the miniature unmanned plane; sub areas are divided, image acquisition points, coordinate position, height and aerial photography angles can be set, and are sent to the miniature unmanned plan according to a sampling sequence, the coordinate position information received by the miniature unmanned plane is used, automatic cruise acquisition detection is carried out, the plane images in the detected area can be obtained by splicing images, and the drought status can be determined through a meteorology drought grade standard. The drought status detection is carried out by employing the camera of the unmanned plane, the camera can substitute a satellite for real-time monitoring, acquisition of high-resolution image is realized, disadvantage that image can not be obtained due to shielding of a cloud layer by satellite remote sensing can be made up, and problems of long revisiting period and untimely emergency of traditional satellite remote sensing can be solved.

CHAPTER 3
REQUIREMENTS

3.1 FUNCTIONAL REQUIREMENTS

1. **Identification of water deficient districts :**

   This module focuses on classifying the districts of Maharashtra as water sufficient or deficient based on rainwater and ground water level.

2. **Indexing the data :**

   The data includes various parameters like rainfall , ground water , temperature, humidity which need to be indexed according to their importance.

3. **Training the model:**

   Using various classification algorithms models need to be trained to identify which would best classify a location to be suitable for catchment or not.

4. **Comparative study of various algorithms:**

   Evaluating the efficiency of various algorithms to determine which gives the best result

5. **Giving a distribution plan :**

   After the trained machine learning model predicts whether a given location is suitable for catchment or not it looks for areas which are already self sufficient for water or need water and are in 35 km radius of the location to receive or supply water according to the prediction.

3.2 NON-FUNCTIONAL REQUIREMENTS

1. Generation of interactive report and visualization for the performed analysis.
2. Customizable color code system which can be set by the end user.
3. Easy retrieval of previously generated reports for satellites images of other timelines.

4. Highly optimized algorithm with minimal time and space complexity.
5. Water supply management for urban areas.
6. Identifying urban regions susceptible to water shortage

3.3 CONSTRAINTS

1. Ready-made datasets of satellite images are not easily available.
2. High speed of computations is essential to process large amount of data.
3. The resolution of the obtained satellite images should be high.
4. Accuracy of datasets used.
5. Accurate geographical and climatic conditions of the region.
6. Unavailability of urbanization data.

3.4 HARDWARE AND SOFTWARE REQUIREMENTS

HARDWARE:

- Virtual Machine (GPU)
- Virtual Machine (CPU)
- Operating System (Windows 10)
- RAM: 8GB or more.
- Hard Drive Space Required: at least 3 TB
- Processor: Intel Core i5 or higher

SOFTWARE:

Technologies to be used:

- Python
- Flask Web Framework
- Azure Machine Learning Studio
- OpenCV
- Scikit-Learn

Tools to be used:

- Pentaho
- Google Earth Pro
- UiPath Studio
- Azure SQL Database
- Azure Web Service
- Power BI Embedded

3.5 TECHNIQUES, TOOLS AND ALGORITHMS USED TILL DATE

- The OpenCV library can take advantage of multi-core processing. Enabled with OpenCL, it can take advantage of the hardware acceleration of the underlying heterogeneous compute platform.
- Azure ML studio to train model
- TensorFlow is used to create large-scale neural networks with many layers. Classification, Perception, Understanding, Discovering, Prediction and Creation and similar machine learning problems majorly makes use of TensorFlow
- Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries.
- pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.
- Pandas is a NumFOCUS sponsored project. This will help ensure the success of development of pandas as a world-class open-source project, and makes it possible to donate to the project.

3.6 PROJECT PROPOSAL

Various weather parameters will be collected for the state of Maharashtra. These include rainfall, temperature, humidity, groundwater, etc. Further the elevation of a place is identified using APIs. The soil type of a place is identified and scored according to water soaking capacity. All of the collected data is cleaned and missing values are handled using least square regression. Further the parameters are evaluated and a score is generated for a region these scores are used to label the data . Finally labelled data is trained and model tested for accuracy for various algorithms on Azure ML studios.The model with best accuracy is selected and REST API is generated for the same. The new points are tested and results are reported.

CHAPTER 4
PROPOSED DESIGN

4.1 Block Diagram

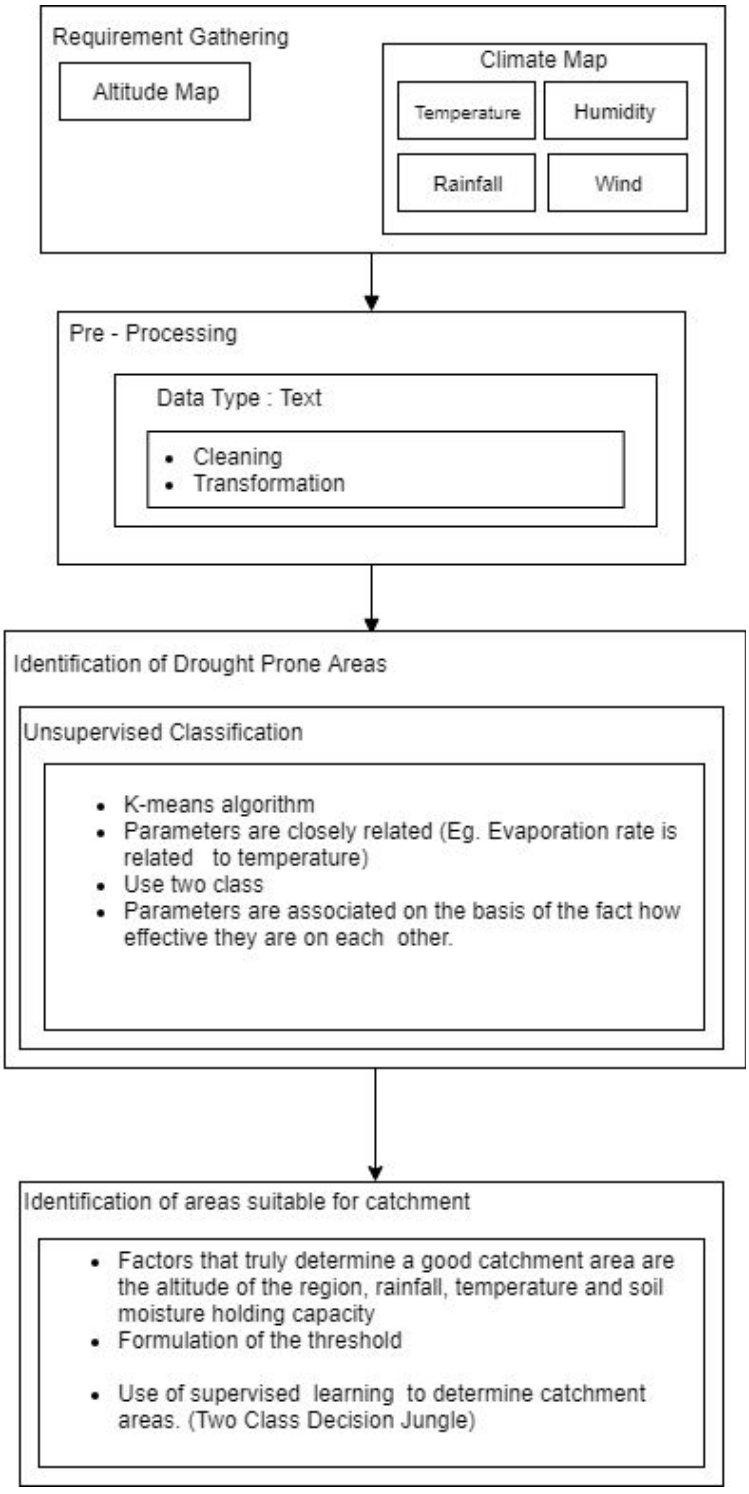


Figure 1 : Block Diagram

4.2 Modular Design

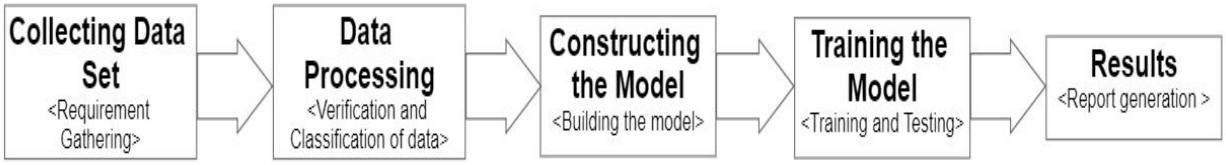


Figure 2: Modular Design

The above diagram is the modular diagram of our system. It is a five layered architecture. Through this modular design we are trying to show the entire flow of process and interaction between the different layers of our system, mainly dividing into five main components as mentioned above.

The main task in any machine learning project is collection of dataset, so the very first step in the process is collection of the dataset. It is possible that the collected dataset may contain unnecessary data, vague data, data that is not in the correct required format. The next step of data processing converts the collected data into the required format.

After collecting and processing the data the next thing that is to be done is selecting the algorithms and constructing the model.

Once the model is ready, it needs to be trained and tested using the processed dataset created in the 2nd step and provide the desired results.

4.3 DETAILED DESIGN

4.3.a DFD

1) Level 0 –

The DFD Level 0 briefly shows the overall functionality of the system. The single process represents the entire system. Climatic information and groundwater level are given as input to the system. The system consisting of various other components has the ability to generate a report with causes of drought and an optimal water catchment suggestion.

Figure 3 : Level 0 DFD

2) Level 1 –

Initially dataset has to be collected. Only the required features are to be extracted from datasets and rest to be discarded. A suitable algorithm is selected and the model is constructed. Training of the model takes place on the training dataset to predict the drought affected and prone regions. Finally a report is generated with causes of drought and an optimal water catchment suggestion.

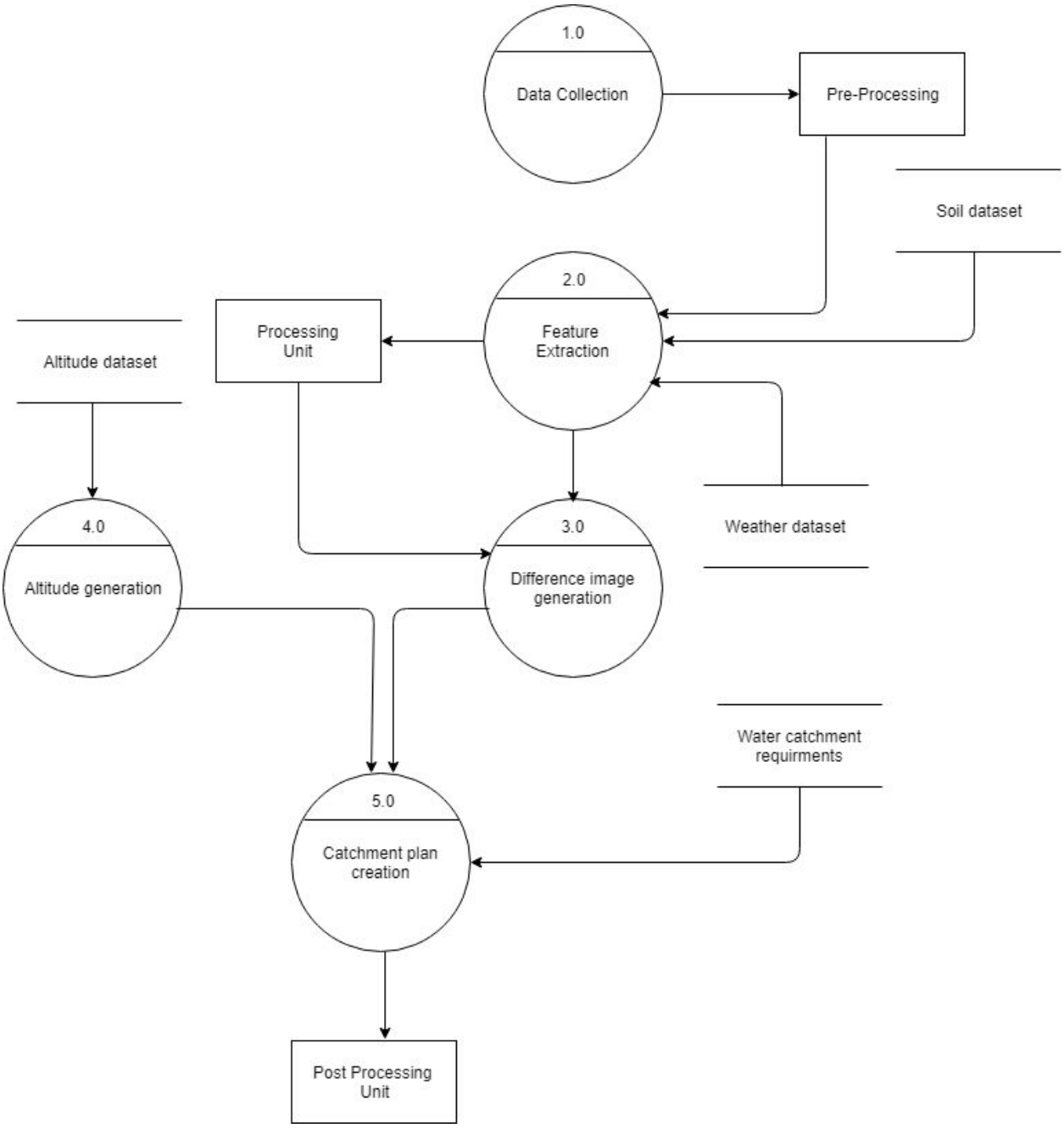


Figure 4 : Level 1 DFD

4.3.b Flowchart for the proposed system

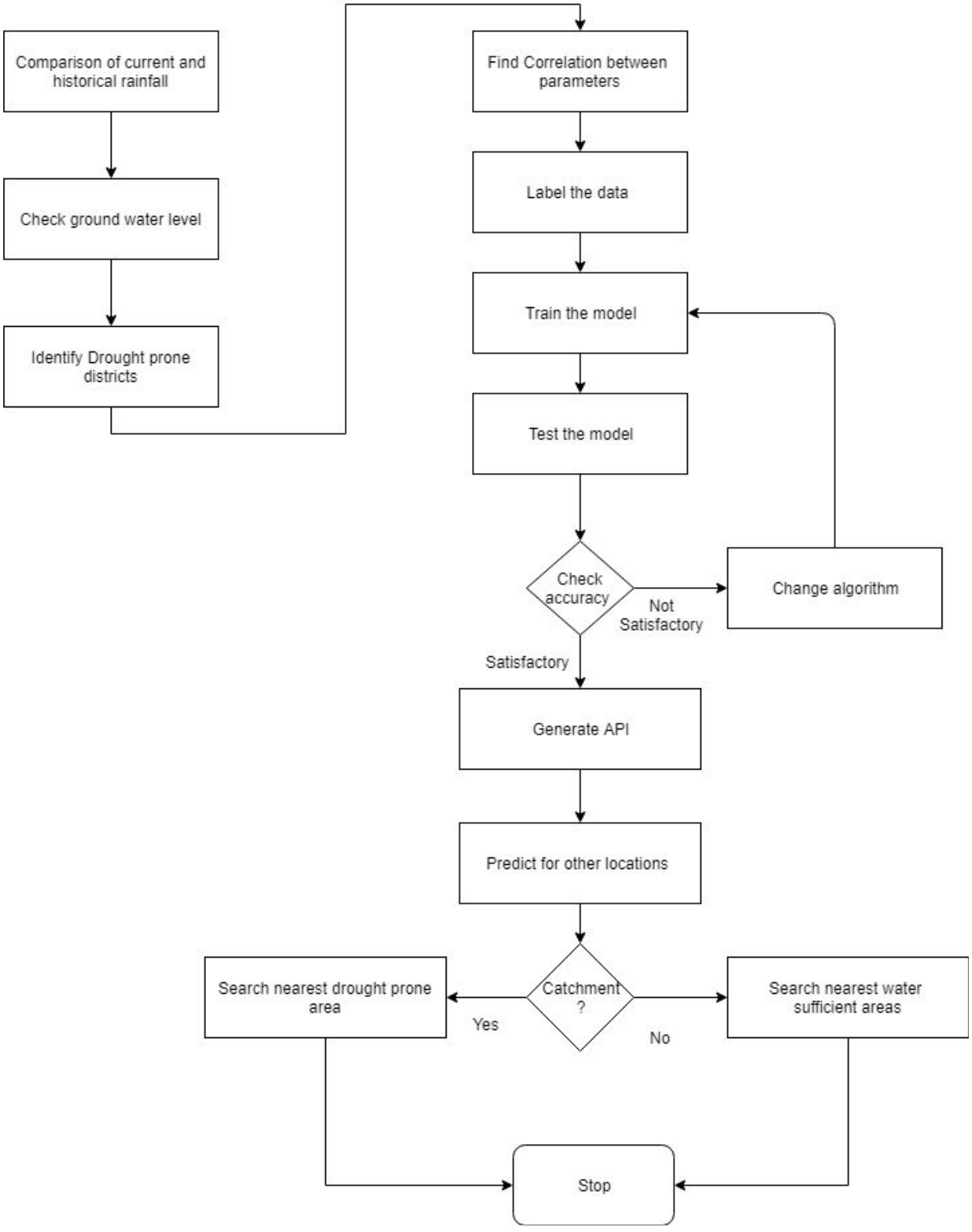


Figure 5 : Flowchart

4.4 PROJECT SCHEDULING & TRACKING USING GANTT CHART

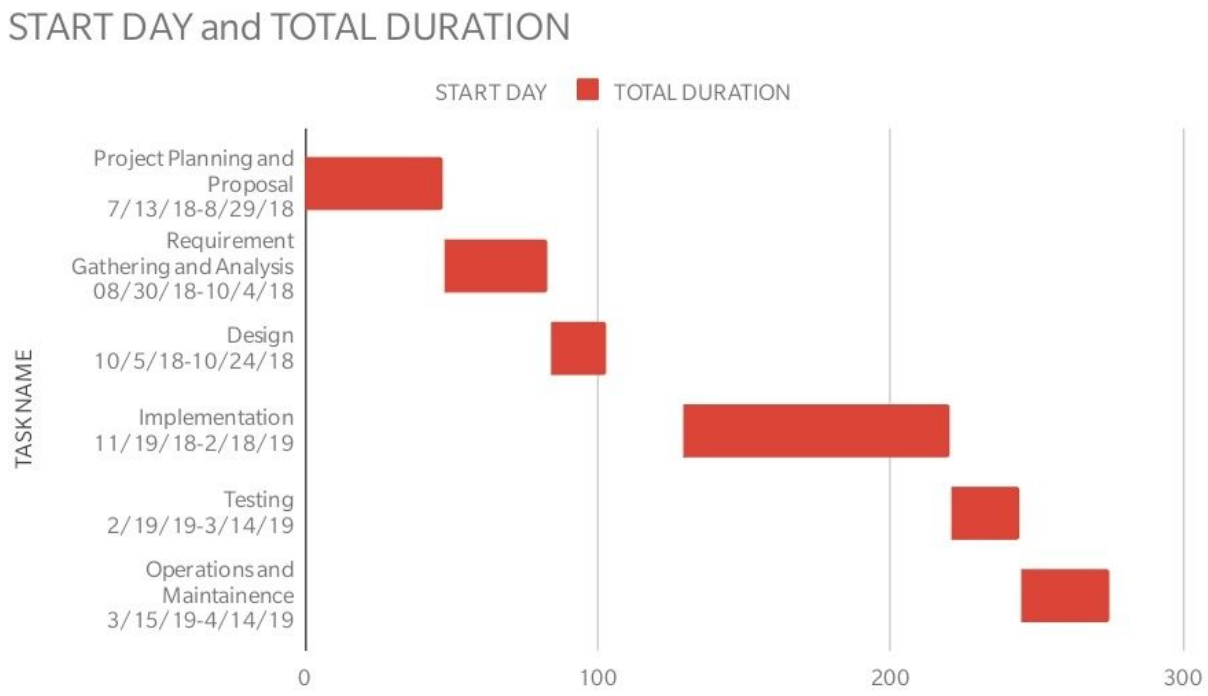


Figure 6 : Gantt Chart

CHAPTER 5
IMPLEMENTATION DETAILS

5.1. Algorithms considered for study

Algorithms used for feature extraction

1) **K-means clustering**

One of most used clustering algorithm is k-means clustering. It is simple and computationally faster than the

hierarchical clustering. And it can also work for large number of variable. But it produces different cluster result for

different number of clusters. So it is required to initialize the proper number of number of cluster, k. Again,

it is required to initialize the k number of centroid. Different value of initial centroid would result different cluster.

So selection of proper initial centroid is also an important task.

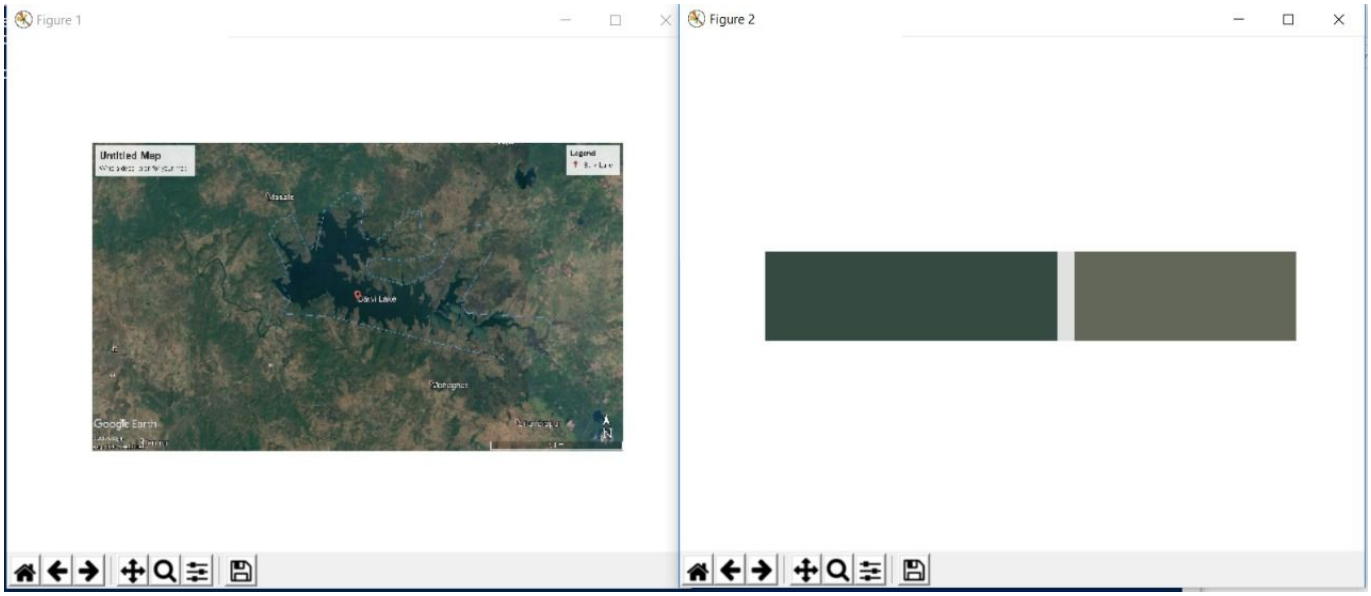


Figure 7 : K Means implementation

2) inRange function in opencv-python

The cv2.inRange function expects three arguments: the first is the image were we are going to perform color detection, the second is the lower limit of the color you want to detect, and the third argument is the upper limit of the color you want to detect.

After calling cv2.inRange, a binary mask is returned, where white pixels (255) represent pixels that fall into the upper and lower limit range and black pixels (0) do not.

1) Water body detection

Figure 8 : Satellite image before highlighting water body

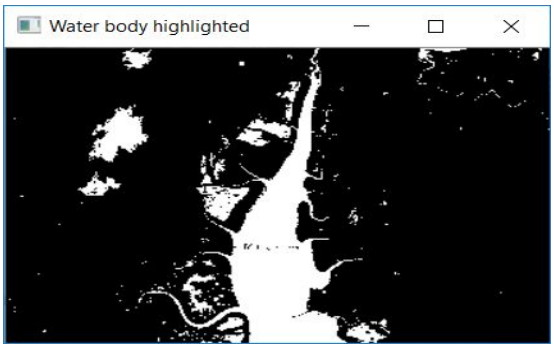


Figure 9 : Satellite image after highlighting water body

2) Vegetation detection

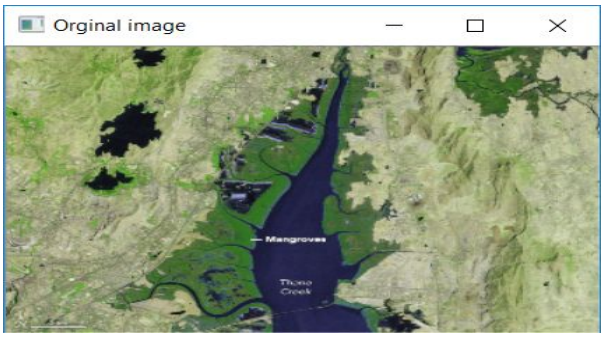


Figure 10 : Satellite image before highlighting vegetation

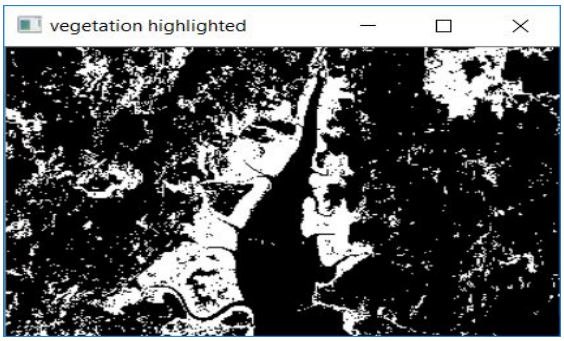


Figure 11 : Satellite image after highlighting vegetation

Figure 12 :Gathering elevation for given place

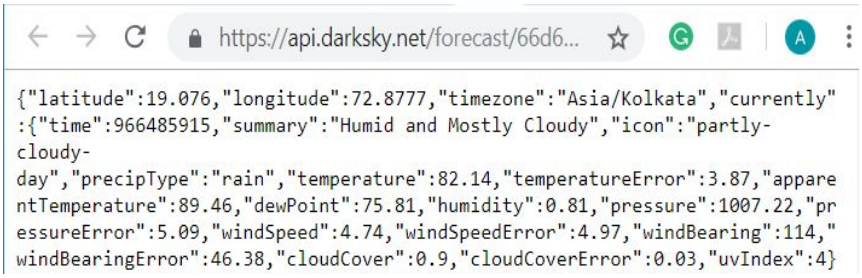


Figure 13 : Gathering weather-data for given place

1. The work on the proposed solution will show how we have exploited our nature and mother earth throughout these years. The solution will be helpful to retain the biodiversity. Our project with the help of Azure resources will be able to improve the quality of research and also to increase the catchment areas needed. Our data gathering team is facing issues with high processing of satellite images. Using GPU resources and Hadoop Architecture of Azure we will be able to process the raw input faster and store them efficiently.
2. These results would not only benefit our work, but also the work of thousands of other practitioners and researchers,who find it difficult to analyse the changes in the environment.The application will be open-source.We will disseminate our results close to coastal areas in India and across the larger network, which will in turn help to advance land monitoring and biodiversity conservation science across various Indian states and other places.

3) Two-Class Averaged Perceptron

The averaged perceptron method is an early and very simple version of a neural network. In this approach, inputs are classified into several possible outputs based on a linear function, and then combined with a set of weights that are derived from the feature vector—hence the name "perceptron."

The simpler perceptron models are suited to learning linearly separable patterns, whereas neural networks (especially deep neural networks) can model more complex class boundaries. However, perceptrons are faster, and because they process cases serially, perceptrons can be used with continuous training.

4) Two-Class Bayes Point Machine

The algorithm in this module uses a Bayesian approach to linear classification called the "Bayes Point Machine". This algorithm efficiently approximates the theoretically optimal Bayesian average of linear classifiers (in terms of generalization performance) by choosing one "average" classifier, the Bayes Point. Because the Bayes Point Machine is a Bayesian classification model, it is not prone to overfitting to the training data.

5) Two-Class Boosted Decision Tree

A boosted decision tree is an ensemble learning method in which the second tree corrects for the errors of the first tree, the third tree corrects for the errors of the first and second trees, and so forth. Predictions are based on the entire ensemble of trees together that makes the prediction.

Generally, when properly configured, boosted decision trees are the easiest methods with which to get top performance on a wide variety of machine learning tasks. However, they are also one of the more memory-intensive learners, and the current implementation holds everything in memory. Therefore, a boosted decision tree model might not be able to process the very large datasets that some linear learners can handle.

6) Two-Class Decision Forest

This decision forest algorithm is an ensemble learning method intended for classification tasks. Ensemble methods are based on the general principle that rather than relying on a single model, you can get better results and a more generalized model by creating multiple related models and combining them in some way. Generally, ensemble models provide better coverage and accuracy than single decision trees.

There are many ways to create individual models and combine them in an ensemble. This particular implementation of a decision forest works by building multiple decision trees and then voting on the most popular output class. Voting is one of the better-known methods for generating results in an ensemble model.

7) Two-Class Decision Jungle

Decision jungles are a recent extension to decision forests. A decision jungle consists of an ensemble of decision directed acyclic graphs (DAGs).

Decision jungles have the following advantages:

- By allowing tree branches to merge, a decision DAG typically has a lower memory footprint and better generalization performance than a decision tree, albeit at the cost of somewhat longer training time.
- Decision jungles are non-parametric models that can represent non-linear decision boundaries.
- They perform integrated feature selection and classification and are resilient in the presence of noisy features.

8) Two-Class Locally Deep Support Vector Machine

Support vector machines (SVMs) are an extremely popular and well-researched class of supervised learning models, which can be used in linear and non-linear classification tasks. Recent research has focused on ways to optimize these models to efficiently scale to larger training sets. In this implementation from Microsoft Research, the kernel function that is used for mapping data points to feature space is specifically designed to reduce the time needed for training while maintaining most of the classification accuracy.

This model is a supervised learning method, and therefore requires a tagged dataset, which includes a label column.

9) Two-Class Logistic Regression

Logistic regression is a well-known method in statistics that is used to predict the probability of an outcome, and is especially popular for classification tasks. The algorithm predicts the probability of occurrence of an event by fitting data to a logistic function.

10) Two-Class Neural Network

A neural network is a set of interconnected layers. The inputs are the first layer, and are connected to an output layer by an acyclic graph comprised of weighted edges and nodes. Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers. However, recent research has shown that deep neural networks (DNN) with many layers can be very effective in complex tasks such as image or speech recognition. The successive layers are used to model increasing levels of semantic depth.

The relationship between inputs and outputs is learned from training the neural network on the input data. The direction of the graph proceeds from the inputs through the hidden layer and to the output layer. All nodes in a layer are connected by the weighted edges to nodes in the next layer.

To compute the output of the network for a particular input, a value is calculated at each node in the hidden layers and in the output layer. The value is set by calculating the weighted sum of the values of the nodes from the previous layer. An activation function is then applied to that weighted sum.

11) Two-Class Support Vector Machine

Support vector machines are among the earliest of machine learning algorithms, and SVM models have been used in many applications, from information retrieval to text and image classification. SVMs can be used for both classification and regression tasks.

This SVM model is a supervised learning model that requires labeled data. In the training process, the algorithm analyzes input data and recognizes patterns in a multi-dimensional feature space called the hyperplane. All input examples are represented as points in this space, and are mapped to output categories in such a way that categories are divided by as wide and clear a gap as possible.

For prediction, the SVM algorithm assigns new examples into one category or the other, mapping them into that same space.

5.2. Comparative Analysis with the existing algorithms

Algorithm	Accuracy	True Positive	True Negative	False Positive	False Negative
Two-Class Averaged Perceptron	0.942	1176	6652	0	482
Two-Class Bayes Point Machine	0.906	901	6627	25	757
Two-Class Boosted Decision Tree	0.967	1605	6434	218	53
Two-Class Decision Forest	0.968	1588	6455	197	70
Two-Class Decision Jungle	0.972	1606	6473	179	52
Two-Class Locally Deep Support Vector Machine	0.953	1331	6592	60	327
Two-Class Logistic Regression	0.942	1176	6652	0	482
Two-Class Neural Network	0.899	1511	5957	695	147
Two-Class Support Vector Machine	0.942	1185	6646	6	473

Table 1 : Comparison of different algorithms considered for study

CHAPTER 6
TESTING

6.1 . Definition of testing

Testing is a verification process for quality assessment and improvement. Testing is basically done to find errors, faults in the system. The basic goal of software development process is to produce the software that has very few or no errors. In an effort to detect errors soon after they are introduced each phase ends with verification activity such as reviews. However most of these verification activities in the early phase of the software development are based on human evaluation and cannot detect all the errors. Testing plays an important role in quality assurance for the software. It is a dynamic method for the verification and validation, where the system to be tested is executed and the behavior of the system is observed.

6.2. Types of tests

6.2.1. Black box testing

Black-box testing is a method of software testing that examines the functionality of an application (e.g. what the software does) without peering into its internal structures or workings (see white-box testing). This method of test can be applied to virtually every level of software testing: unit, integration system and acceptance. It typically comprises most if not all higher level testing, but can also dominate unit testing as well.

6.2.2. White box testing

White-box testing (also known as clear box testing, glass box testing, transparent box testing, and structural testing) is a method of testing software that tests internal structures or workings of an application, as opposed to its functionality (i.e. black-box testing). In white-box testing an internal perspective of the system, as well as programming skills, are used to design test cases. The tester chooses inputs to exercise paths through the code and determine the appropriate outputs. This is analogous to testing nodes in a circuit, e.g. in-circuit testing (ICT).

6.2.3. Unit testing

Unit testing, also known as component testing refers to tests that verify the functionality of a specific section of code, usually at the function level. In an object-oriented environment, this is usually at the class level, and the minimal unit tests include the constructors and destructors.

These types of tests are usually written by developers as they work on code (white-box style), to ensure that the specific function is working as expected. One function might have multiple tests, to catch corner cases or other branches in the code. Unit testing alone cannot verify the functionality of a piece of software, but rather is used to ensure that the building blocks of the software work independently from each other.

Unit testing is a software development process that involves synchronized application of a broad spectrum of defect prevention and detection strategies in order to reduce software development risks, time, and costs. It is performed by the software developer or engineer during the construction phase of the software development lifecycle. Rather than replace traditional QA focuses, it augments it. Unit testing aims to eliminate construction errors before code is promoted to QA; this strategy is intended to increase the quality of the resulting software as well as the efficiency of the overall development and QA process.

Depending on the organization's expectations for software development, unit testing might include static code analysis, data flow analysis, metrics analysis, peer code reviews, code coverage analysis and other software verification practices.

6.2.4 Integration Testing

Integration testing is any type of software testing that seeks to verify the interfaces between components against a software design. Software components may be integrated in an iterative way or all together ("big bang"). Normally the former is considered a better practice since it allows interface issues to be located more quickly and fixed.

Integration testing works to expose defects in the interfaces and interaction between integrated components (modules).

6.2.5. Component interface testing

The practice of component interface testing can be used to check the handling of data passed between various units, or subsystem components, beyond full integration testing between those units. The data being passed can be considered as "message packets" and the range or data types can be checked, for data generated from one unit, and tested for validity before being passed into another unit. One option for interface testing is to keep a separate log file of data items being passed, often with a timestamp logged to allow analysis of thousands of cases of data passed between units for days or weeks. Tests can include checking the handling of some extreme data values while other interface variables are passed as normal values. Unusual data values in an interface can help explain unexpected performance in the next unit. Component interface testing is a variation of black-box testing, with the focus on the data values beyond just the related actions of a subsystem component.

6.2.6. System testing

System testing, or end-to-end testing, tests a completely integrated system to verify that it meets its requirements. For example, a system test might involve testing a login interface, then creating and editing an entry, plus sending or printing results, followed by summary processing or deletion (or archiving) of entries, then logoff.

In addition, the software testing should ensure that the program, as well as working as expected, does not also destroy or partially corrupt its operating environment or cause other processes within that environment to become inoperative (this includes not corrupting shared memory, not consuming or locking up excessive resources and leaving any parallel processes unharmed by its presence).

6.2.7. Alpha Testing

Alpha testing is a type of acceptance testing; performed to identify all possible issues/bugs before releasing the product to everyday users or public. The focus of this testing is to simulate real users by using black box and white box techniques.

The aim is to carry out the tasks that a typical user might perform. Alpha testing is carried out in a lab environment and usually the testers are internal employees of the organization. To put it as simple as possible, this kind of testing is called alpha only because it is done early on, near the end of the development of the software, and before beta testing.

6.2.8 Beta Testing

Beta Testing of a product is performed by "real users" of the software application in a "real environment" and can be considered as a form of external user acceptance testing.

Beta version of the software is released to a limited number of end-users of the product to obtain feedback on the product quality. Beta testing reduces product failure risks and provides increased quality of the product through customer validation.

It is the final test before shipping a product to the customers. Direct feedback from customers is a major advantage of Beta Testing. This testing helps to tests the product in real time environment.

6.3. Type of Testing considered with justification

Attributes used for training the model :

1. Rainfall
2. Groundwater
3. Soil Score
4. Elevation
5. Latitude
6. Longitude

Target Variable : Catchment

Possible values of target variable :

1 . 0 (means the given area is not suitable for catchment)

2. 1 (means the given area is suitable for catchment)

Algorithm used for training the model : Two Class Decision Jungle

Impact of attributes :

1. **Initial accuracy by considering all attributes for training :**

   Accuracy : 97.2%

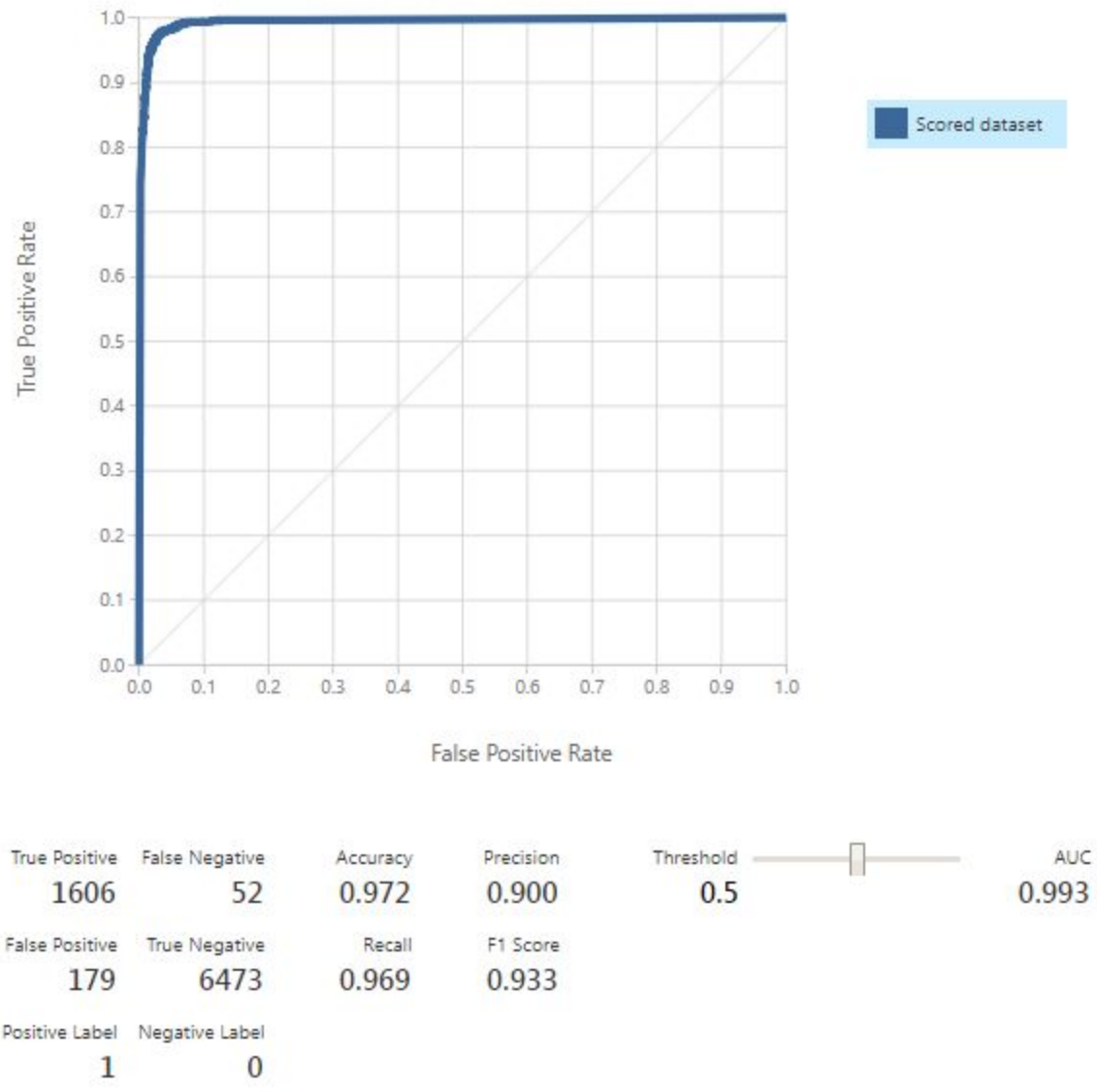


Figure 14 : Initial accuracy

2) **Effect of removing an attribute and keeping other attributes as they were :**

   1) **Removing attributes “Latitude” and “Longitude” :**

Accuracy : 96.6 %
Accuracy decreased by 0.6%

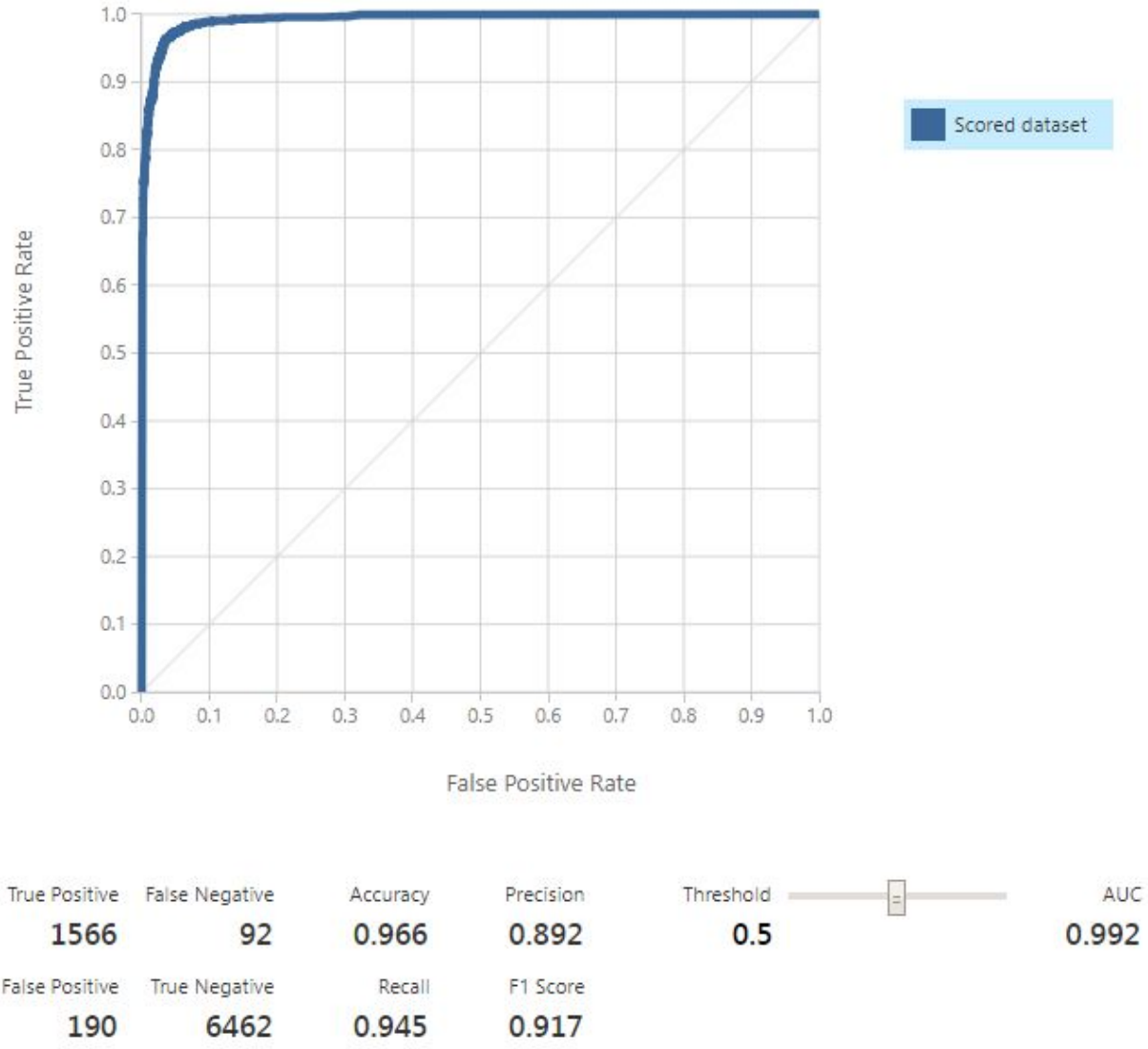
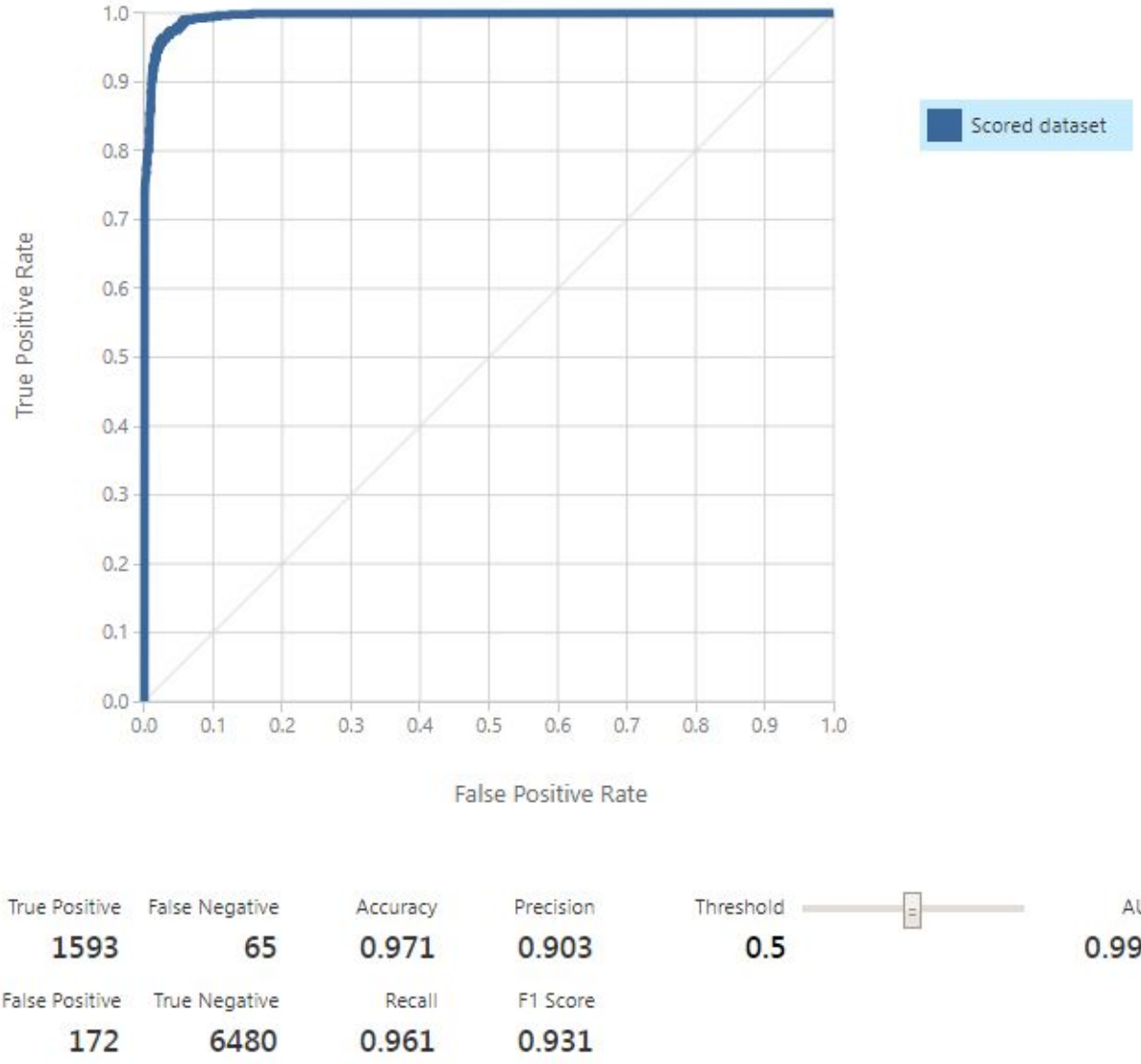


Figure 15 : Accuracy after removing “Latitude” and “Longitude”

2) Removing attribute “Elevation” :

Accuracy: 97.1%
Accuracy decreased by 0.1%



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
1593	65	0.971	0.903	0.5	0.995

False Positive	True Negative	Recall	F1 Score
172	6480	0.961	0.931

Figure 16 : Accuracy after removing “Elevation”

3) Removing attribute “Soil Score” :

Accuracy: 97.1%
Accuracy decreased by 0.1%

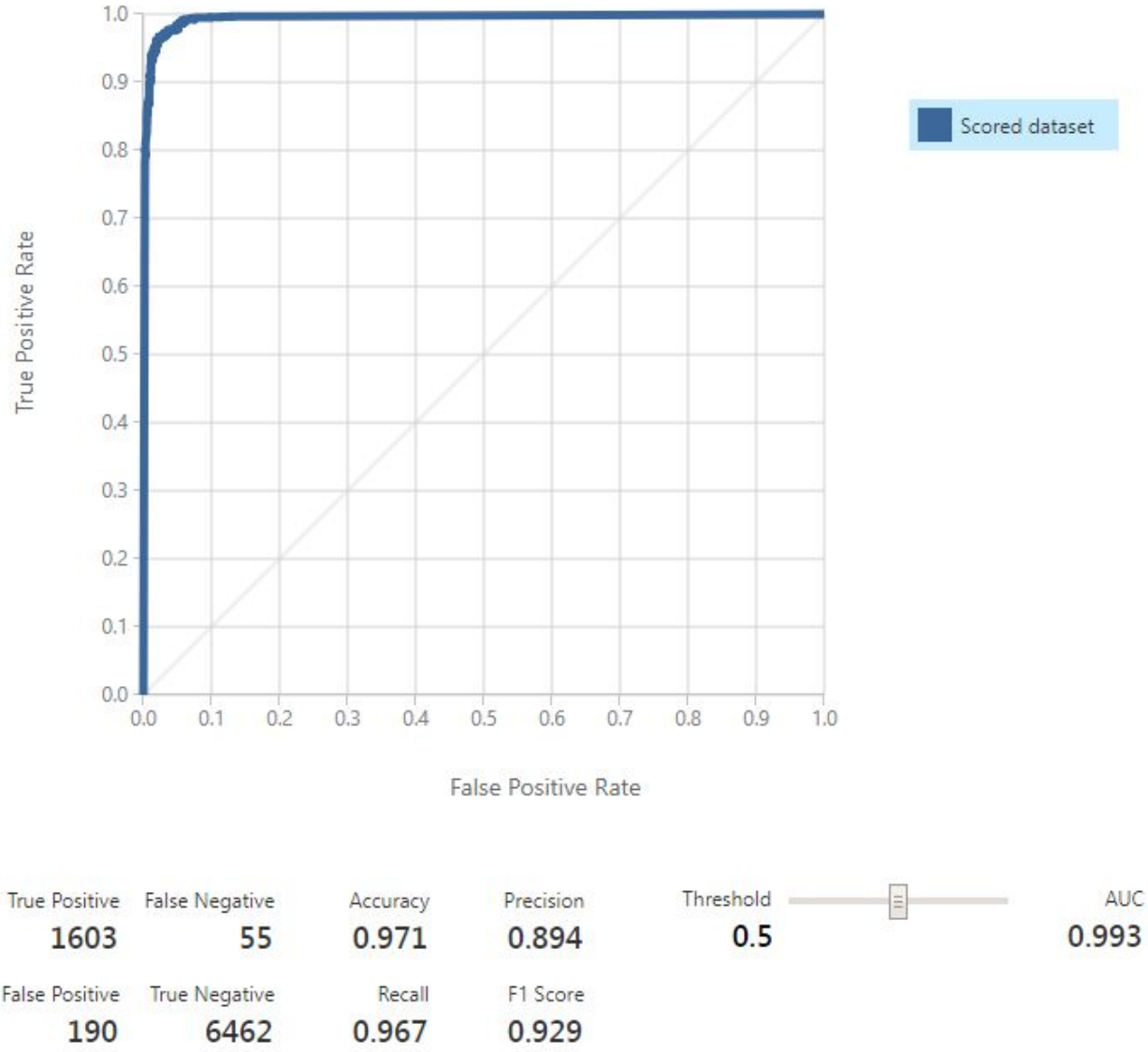


Figure 17 : Accuracy after removing “Soil Score”

4) Removing attribute “Groundwater” :

Accuracy: 96.9%
Accuracy decreased by 0.3%

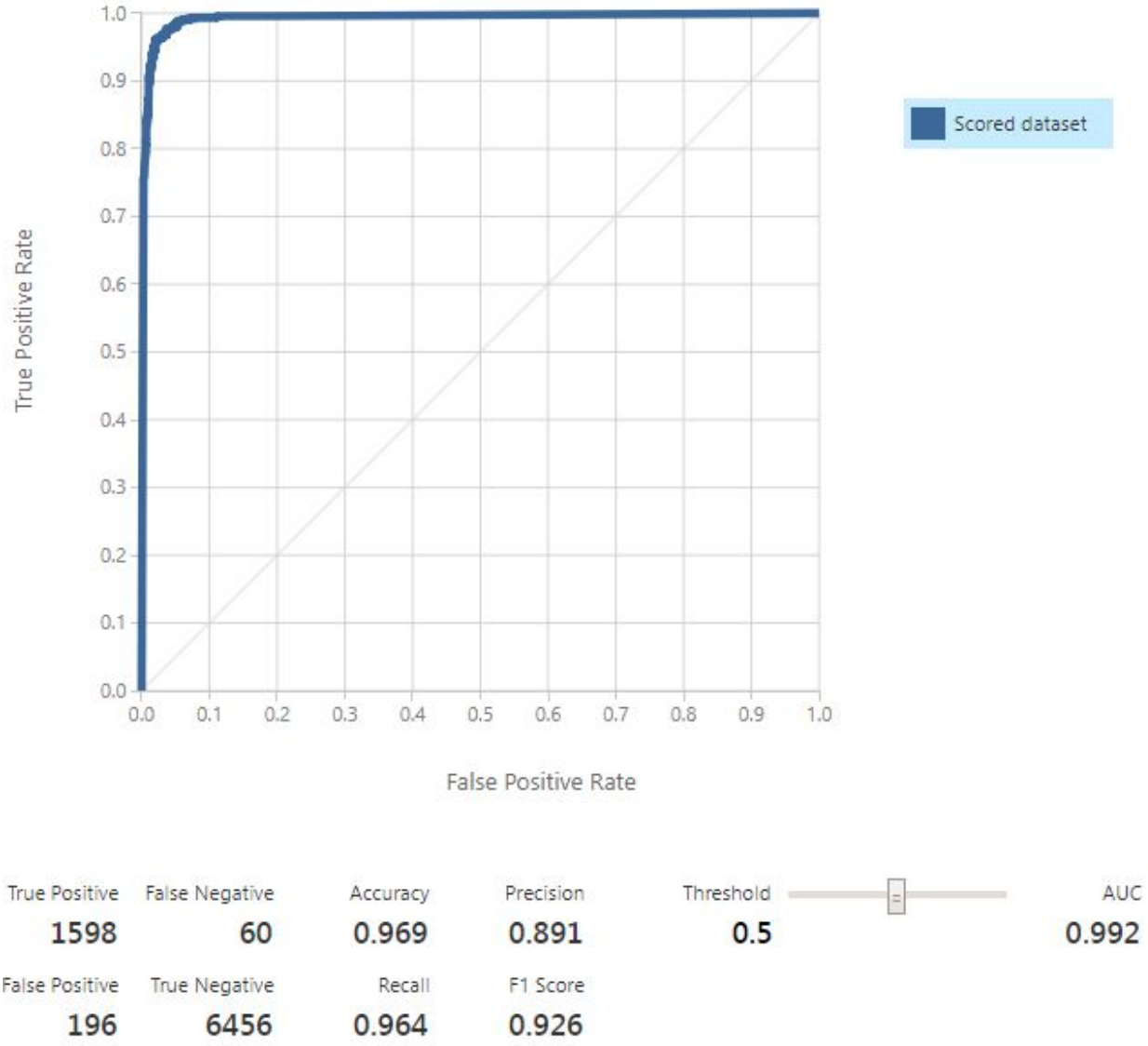


Figure 18 : Accuracy after removing “Groundwater”

5)Removing attribute “Rainfall” :

Accuracy: 95.3%
Accuracy decreased by 1.9%

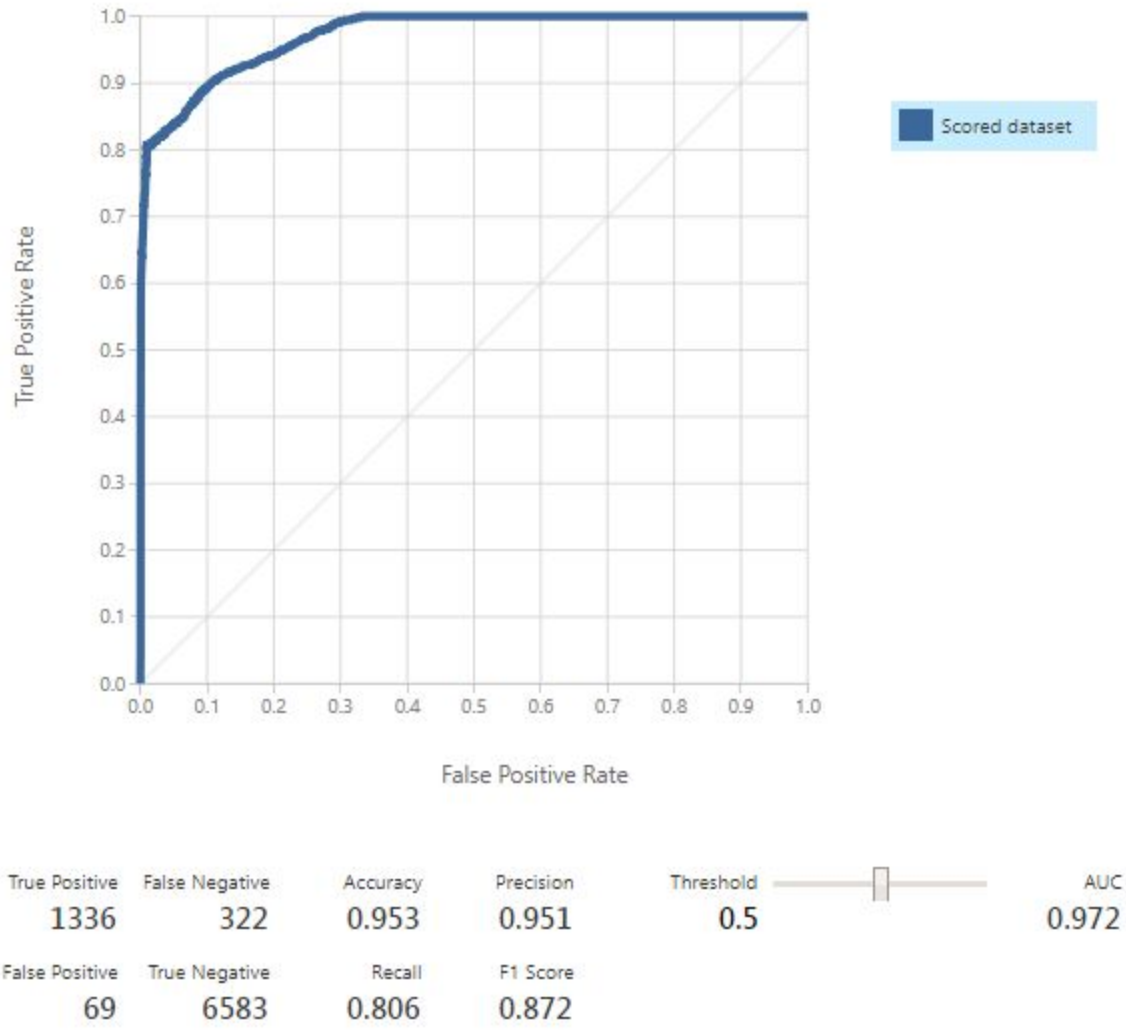


Figure 19 : Accuracy after removing “Rainfall”

Summary :

Algorithm used : Two class decision jungle

Removed attribute	Current Accuracy	Change in accuracy	Weight Assigned
Rainfall	95.30%	1.90%	4
Groundwater	96.90%	0.30%	3
Soil Score	97.10%	0.10%	2
Elevation	97.10%	0.10%	0.5

Table 2 : Impact analysis of attributes and comparison with weights assigned previously

6.4 Various test case scenarios considered

Training-testing splitting is 70%-30% in this case.

Algorithm	Training Accuracy	Testing Accuracy
Two-Class Averaged Perceptron	0.913	0.942
Two-Class Bayes Point Machine	0.900	0.906
Two-Class Boosted Decision Tree	0.988	0.967
Two-Class Decision Forest	0.996	0.968
Two-Class Decision Jungle	0.978	0.972
Two-Class Locally Deep Support Vector Machine	0.931	0.953
Two-Class Logistic Regression	0.913	0.942
Two-Class Neural Network	0.861	0.899
Two-Class Support Vector Machine	0.912	0.942

Table 3 : Comparison of training and testing accuracy of algorithms considered for study

6.5. Inference drawn from the test

Training accuracy is greater than testing accuracy in case of only three of the studied algorithms. In those algorithms, two-class decision forest algorithm is clearly overfitting because its training accuracy is 99.6%. Out of remaining two algorithms, two-class decision jungle has maximum testing accuracy. Hence our selection of the two-class decision jungle algorithm is justified.

CHAPTER 7
RESULT ANALYSIS

7.1. Modules under consideration

1. Identification of water deficient districts :

This module focuses on classifying the districts of Maharashtra as water sufficient or deficient based on rainwater and ground water level.

2. Indexing the data :

The data includes various parameters like rainfall , ground water , temperature, humidity which need to be indexed according to their importance.

3. Training the model:

Using various classification algorithms models need to be trained to identify which would best classify a location to be suitable for catchment or not.

4. Comparative study of various algorithms:

Evaluating the efficiency of various algorithms to determine which gives the best result

5. Giving a distribution plan :

After the trained machine learning model predicts whether a given location is suitable for catchment or not it looks for areas which are already self sufficient for water or need water and are in 100 km radius of the location to receive or supply water according to the prediction.

7.2. Screenshots of User Interface (UI)

1) Single input

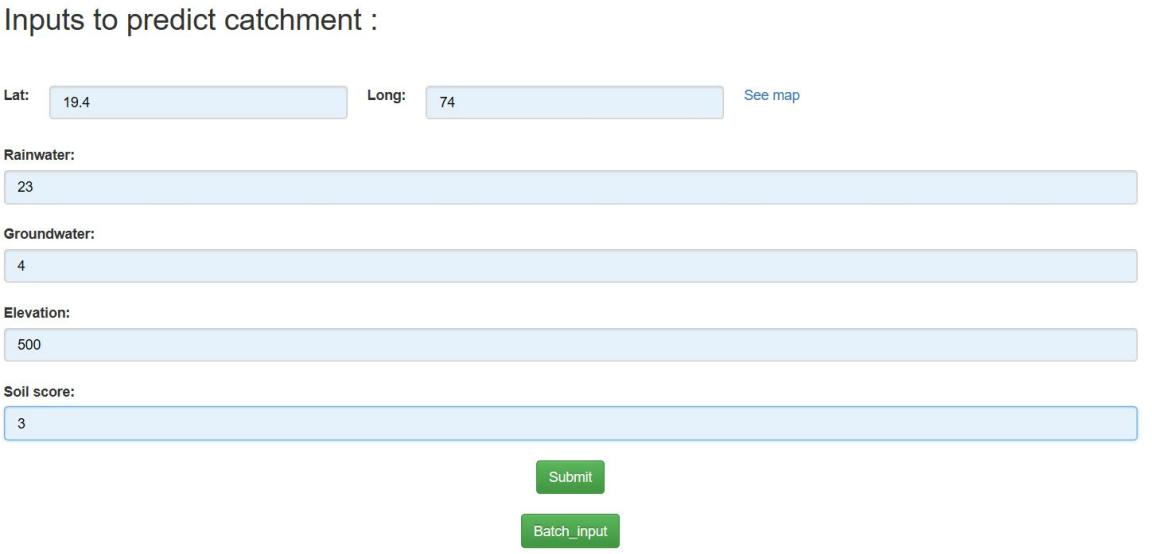


Figure 20 : Single input

2) Single input result

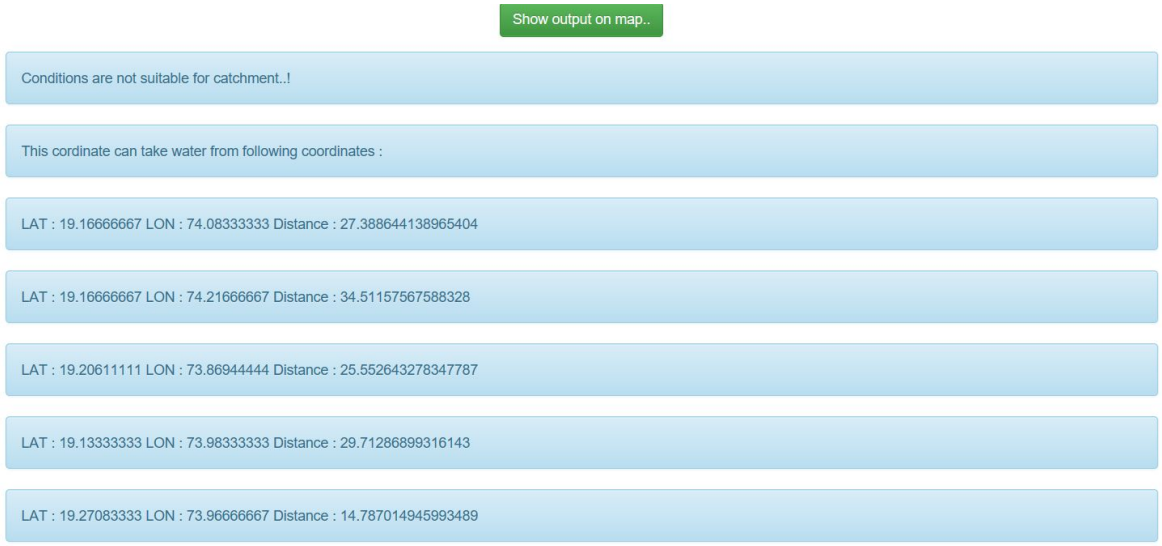


Figure 21 : Single input result

3)Single input result on map

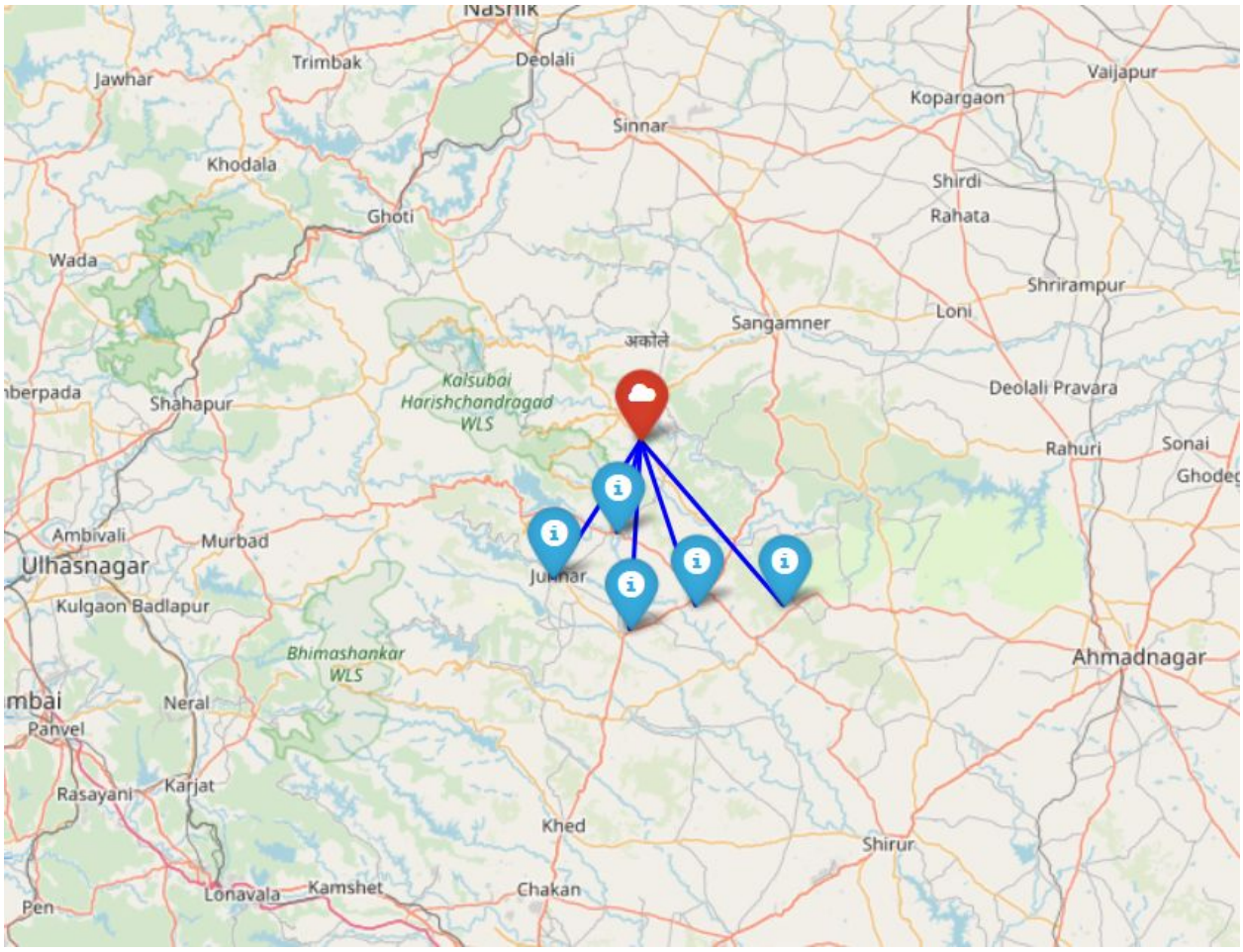


Figure 22 : Single input result on map

4) Batch input

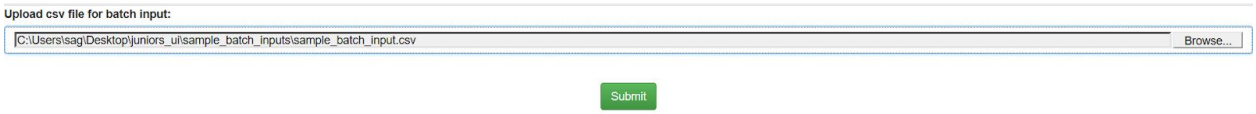


Figure 23 : Batch Input

5) Batch input result

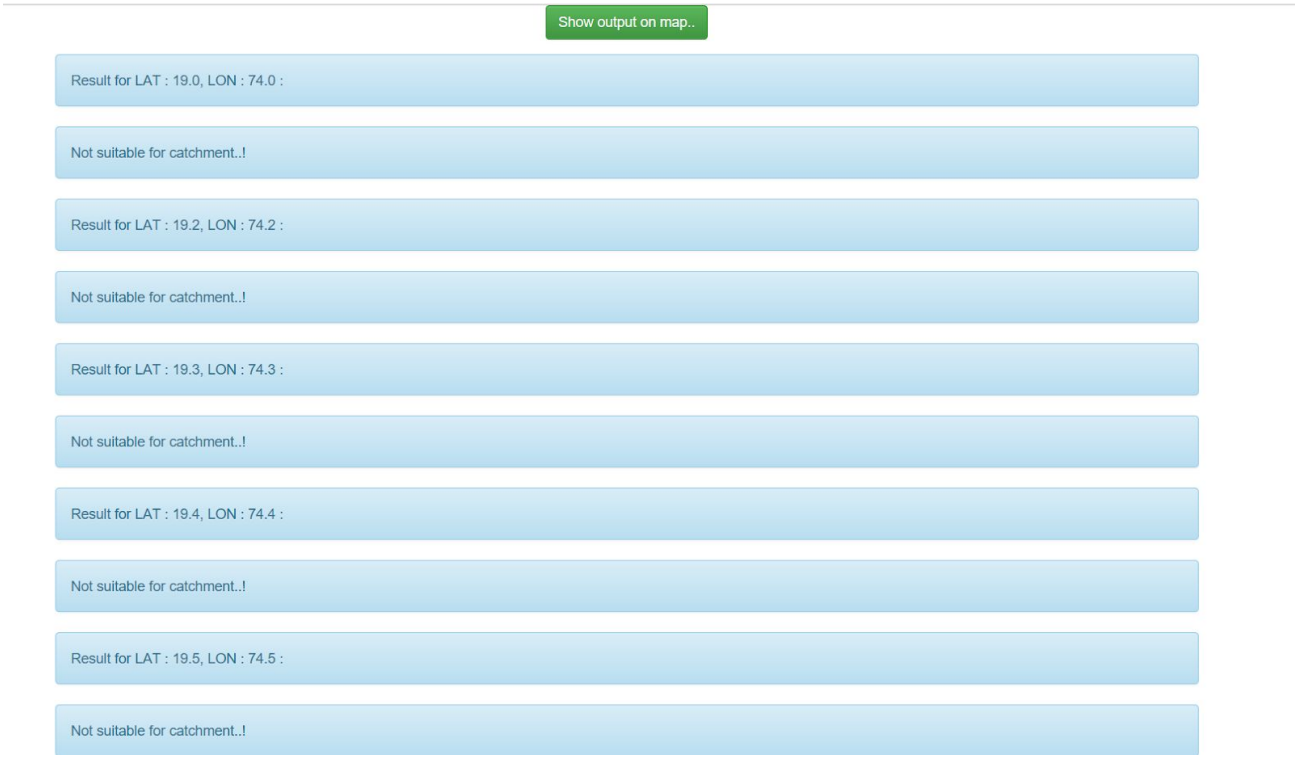


Figure 24 : Batch input result

6) Batch input result on map

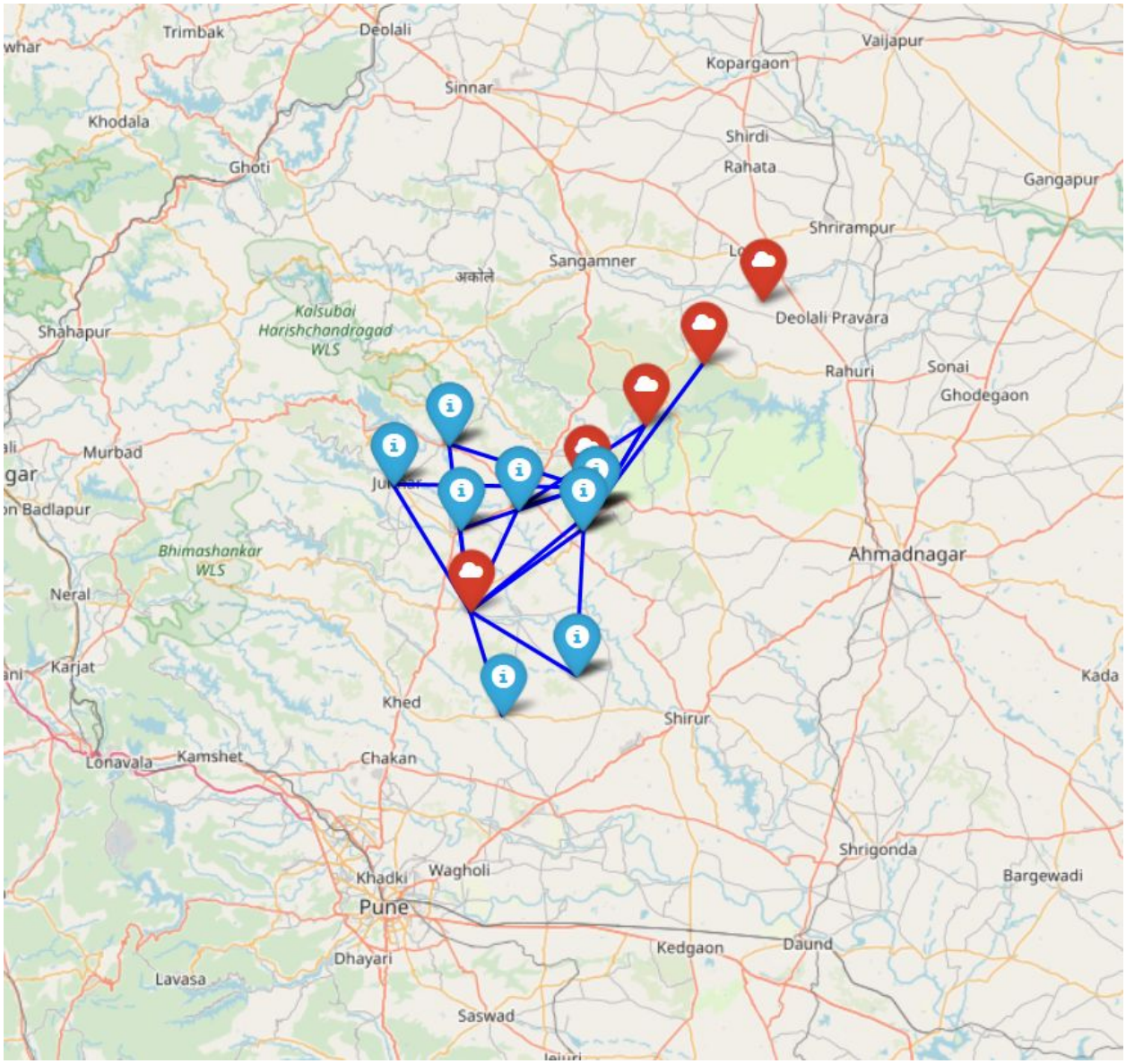


Figure 25 : Batch input result on map

7.3 Evaluation of the developed system

Accuracy : 97.2%
Precision : 90.0%
Recall : 96.9%
F1 Score : 93.3%

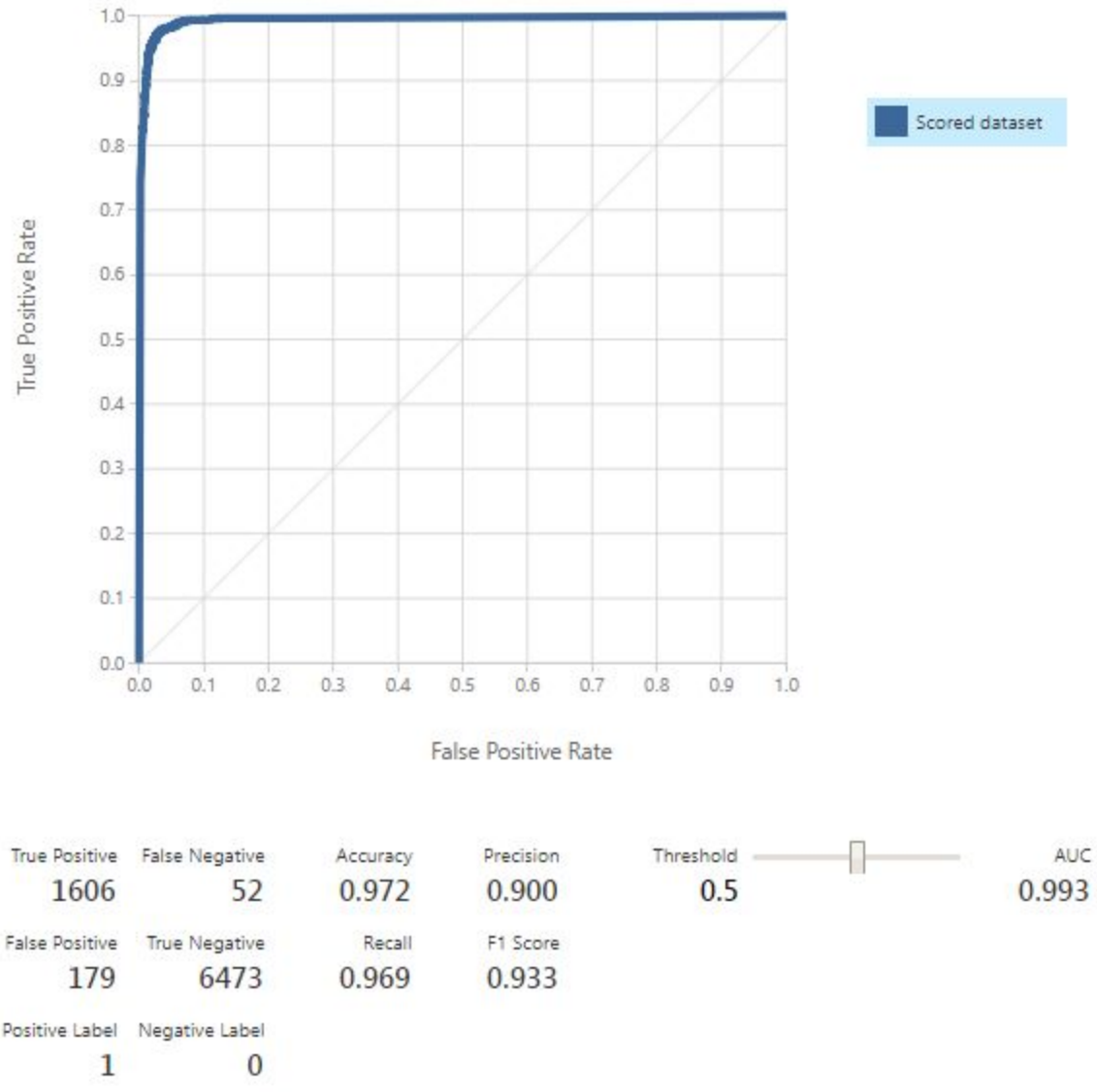


Figure 26 : Accuracy of the trained model

7.4 Graphical outputs of the various scenarios considered

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	DISTRICT	TEH_NAM	BLOCK_NAME	LAT	LON	SITE_NAME	YEAR_OBS	TOTAL_RAINFALL	RAINFALL	GRDWTR	SOIL_TYPE	Soil_Score	Elevation	Region_Score	Indexing1	catchmen	Indexing2	catchment2
2	Ahmadna	Ahmedna	Akola	19.5375	73.99	Malijhap	1996	965.08	15	4.0125	Black	5	649	1	402.5375	0	4206.858	0
3	Ahmadna	Ahmedna	Akola	19.525	73.925	Rajur	1996	965.08	14	3.23	Black	5	649	1	440.19	1	4204.51	0
4	Ahmadna	Bhingar	Jamkhed	18.73333	75.31667	Jamkhed	1996	965.08	24	4.6125	Black	5	650	1	404.8375	0	4209.158	0
5	Ahmadna	Bhingar	Jamkhed	18.58333	75.25	Nanaj-1	1996	965.08	14	2.325	Black	5	650	1	417.975	1	4202.295	0
6	Ahmadna	Deolali Pr	Karjat	18.64444	74.89667	Bhitkewadi	1996	965.08	19	5.01	Black	5	515	1	370.53	0	4142.85	0
7	Ahmadna	Deolali Pr	Karjat	18.51667	74.98333	Kolwadi	1996	965.08	22	8.925	Black	5	515	1	378.275	0	4154.595	0
8	Ahmadna	Deolali Pr	Karjat	18.63333	75.1	Pategaon	1996	965.08	21	1.85	Black	5	515	1	353.05	0	4133.37	0
9	Ahmadna	Deolali Pr	Karjat	18.43333	74.93333	Rassin	1996	965.08	20	5.7375	Black	5	515	1	352.7125	0	4145.033	0
10	Ahmadna	Deolali Pr	Karjat	18.63333	74.88333	Walvad	1996	965.08	17	2.125	Black	5	515	1	345.875	0	4134.195	0
11	Ahmadna	Ghulewad	Kopargaon	19.89444	74.35694	Kolpewadi	1996	965.08	18	2.505	Black	5	509	1	336.015	0	4132.335	0
12	Ahmadna	Ghulewad	Kopargaon	19.88333	74.475	Kopergaon-:	1996	965.08	16	2.175	Black	5	509	1	355.025	0	4131.345	0
13	Ahmadna	Ghulewad	Kopargaon	19.95	74.28333	Manjur	1996	965.08	21	9.76	Black	5	509	1	353.78	0	4154.1	0
14	Ahmadna	Ghulewad	Rahata	19.71667	74.48333	Rahata	1996	965.08	15	8.0225	Black	5	509	1	344.5675	0	4148.888	0
15	Ahmadna	Jamkhed	Nagar	19	74.63333	Chas	1996	965.08	14	2.125	Black	5	596	1	390.375	0	4174.695	0
16	Ahmadna	Jamkhed	Nagar	19.00139	74.91667	Chinchandi	1996	965.08	19	2.6575	Black	5	596	1	399.9725	0	4176.293	0
17	Ahmadna	Jamkhed	Nagar	18.9625	74.80694	Dahigaon	1996	965.08	21	3.45	Black	5	596	1	410.35	0	4178.67	0
18	Ahmadna	Jamkhed	Nagar	19.23333	74.66667	Dehre	1996	965.08	23	1.025	Black	5	596	1	367.075	0	4171.395	0
19	Ahmadna	Jamkhed	Nagar	19.10556	74.63056	Jakhangaon	1996	965.08	14	9.73	Black	5	596	1	405.19	0	4197.51	0
20	Ahmadna	Jamkhed	Nagar	19.2125	74.80556	Jeur	1996	965.08	17	7.715	Black	5	596	1	399.145	0	4191.465	0
21	Ahmadna	Jamkhed	Nagar	19.03333	74.83333	Takle-Kazi	1996	965.08	17	5.9375	Black	5	596	1	413.8125	0	4186.133	0
22	Ahmadna	Kopargaor	Nevasa	19.34722	74.88194	Ghodegaon	1996	965.08	22	3.105	Black	5	515	1	348.815	0	4137.135	0
23	Ahmadna	Kopargaor	Nevasa	19.53194	74.85417	Gonegaon	1996	965.08	18	10.15	Black	5	515	1	369.95	0	4158.27	0

Figure 27 : Dataset

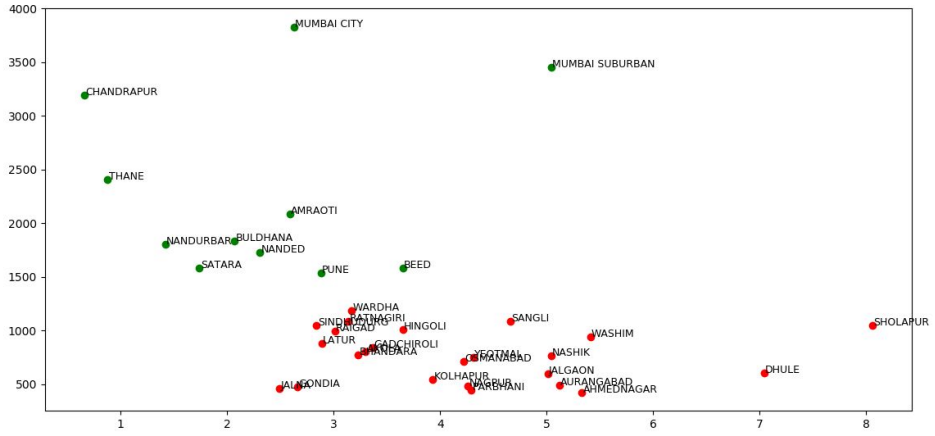


Figure 28 : Rainfall vs Monsoon Groundwater k-means clustering

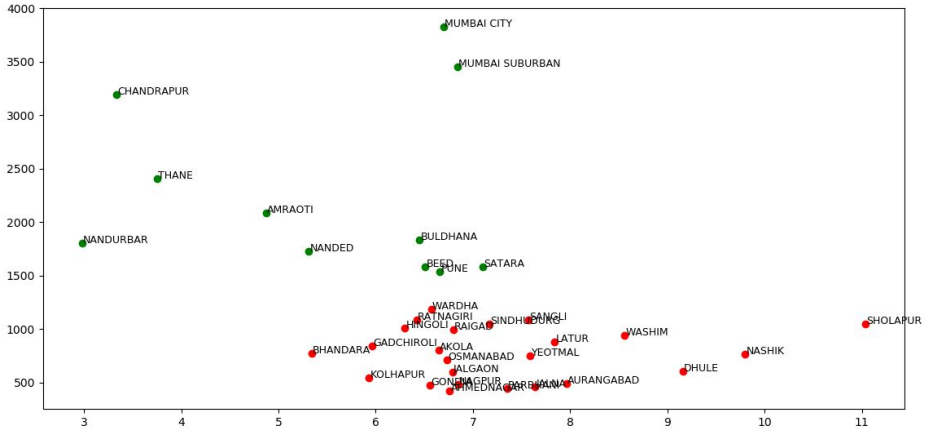


Figure 29 : Rainfall vs Pre-Monsoon Groundwater k-means clustering

CHAPTER 8
CONCLUSION

8.1 LIMITATIONS

- Due to unavailability of computing resources water percentage calculation using images was not feasible.
- Further the dataset used for training lacks few parameters due to the unavailability of open source data for these parameters. These include evaporation rate, vegetation change, urbanisation ,etc.
- The distribution plan has been represented on map. A detailed report for the same could not be generated due to lack of urbanization information.

8.2 CONCLUSION

This project focuses on the issues related to water scarcity in the rural parts of india, where there is limited water resources like lakes and rivers. The ground water level of these areas have also reduced drastically on account of various environmental issues like global warming, etc. The project aims at increasing the catchment areas that are hit by drought like conditions and thereby increasing the ground water level, which will eventually help in the long run. We aim at providing solution to make a region more sustainable. These areas not only lacks the rainfall but also the way the water is stored and supply is managed. These results would not only benefit our work, but also the work of thousands of other practitioners and researchers,who find it difficult to analyse the changes in the environment.The application will be open-source.We will disseminate our results close to coastal areas in India and across the larger network, which will in turn help to advance land monitoring and biodiversity conservation science across various Indian states and other places.

8.3 FUTURE SCOPE

- Further advancement in these models would allow us to predict the type of catchment that could be developed i.e. the predicted catchment would be a self- sustained one or would require human intervention.
- The effect of population of a region is not considered on the water requirements of the region to predict whether the amount of water available is sufficient for present population of the region. Hence , the effect of this parameter on our model could be made.

9. REFERENCES

9.1. Newspaper articles referred

1. 12 districts in Maharashtra stare at drought this year
2. Monitoring land use changes associated with urbanization: An object based image analysis approach

9.2. Research Papers Referred

1. Assia Kourgli ,Youcef Oukil , "Very High Resolution Satellite Images Filtering": Methods to filter or clean images
2. Na Li,Arnaud Martin,Remi Estival,"An automatic water detection approach based on Dempster-Shafer theory for multi-spectral images"
3. Zhang Ding, Ni Qi, Fang Dong, Li Jinhui, Yao Wei, Yuan Shenggui,"Water Bodies Identification From Mutispectral Images Using Gabor Filter, FCM and Canny Edge Detection Method "
4. T.Vignesh,K.K. Thyagharajan, ,"Application of Multispectral Remote Sensing Technology in Surface water body extraction "
5. Dipak R. Samal & Shirish S. Gedam , "Monitoring land use changes associated with urbanization: An object based image analysis approach"
6. Jiayuan Fan ; Tao Chen ; Shijian Lu , "Vegetation coverage detection from very high resolution satellite imagery"
7. Yudo Prasetyo, Setyo Ardi Gunawan and Zia Ul Maksum , "Determination of The Water Catchment Area in Semarang City Using a Combination of Object Based Image Analysis (OBIA) Classification, InSAR and Geographic Information System (GIS) Methods Based On a High-Resolution SPOT 6 Image and Radar Imagery
8. Y. Takefuji, K. C. Lee, and H. Aiso, "An artificial maximum neural network: A winner-take-all neuron model forcing the state of the system in a solution domain," Biol. Cybern., vol. 67, pp. 243–251, 1992.
9. K. C. Lee, N. Funabiki, and Y. Takefuji, "A parallel improvement algorithm for the bipartite subgraph problem," IEEE Trans. Neural Networks, vol. 3, pp. 139–145, Mar. 1992

9.3. Patents Referred

1. ZHENG ENHUI, FENG YIHUA, FU YAQIONG, CHEN LE, “Real-time detection method for drought status in detected area based on miniature unmanned plane”, CN20151185314 20150418

10. APPENDIX

10.1. Paper I

1. Paper Published

Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2018)
DVD Part Number:CFP18F70-DVD; ISBN:978-1-5386-4984-8

Water Catchment Control and Management

Richard Joseph
Assistant Professor, VESIT
Mumbai, Maharashtra, India
richard.joseph@ves.ac.in

Samujjwaal Dey
Student, VESIT
Mumbai, Maharashtra, India
2015samujjwaal.dey@ves.ac.in

Gurpreet Singh
Student, VESIT
Mumbai, Maharashtra, India
2015gurpreet.nagpal@ves.ac.in

Ashish Joshi
Student, VESIT
Mumbai, Maharashtra, India
2015ashish.joshi@ves.ac.in

***Abstract*—This paper aims at providing a solution to drought hit areas. The system proposed in this paper, takes into account various environmental factors like Rainfall, Temperature, Soil Type, Ground water, Humidity, Percentage of area covered by water body, Evaporation rate, Altitude, Percentage of area covered by Urbanisation, etc., to identify the regions that can conserve water for long durations and are best suitable to harvest the rainwater. The proposed system uses landsat satellite images and different API's to gather data from different sources. In this paper, we analyze the state of Maharashtra (India) to detect the drought prone areas and to figure out the catchment areas for them. We use clustering algorithms in Data Gathering Phase and neural network algorithms in the Prediction Phase to train the model to identify the right catchment areas after considering all the factors.**

Keywords— ANN, Hopfield Network, Catchment Areas, Hebb Learning Rule, Drought, Hadoop, CUDA Programming

I. Introduction

Over the past decade there has been a significant change in the state of water bodies. The shrinkage of the water bodies due to urbanization and changes in the climatic conditions has led to the scarcity of water and drought like conditions. Considering, the example of Marathwada region of Maharashtra over 41 percent of the villages in the region have reported 'average yield below 50 paise. This means the scarcity of basic resources such as food grain production and water is less than 50 percent of its actual capacity. Erratic showers are considered to be the main cause of less average yield .Our system proposes a solution to the above problem of scarcity of water by identifying potential areas for water catchment and optimum for distribution to rural areas.

In the proposed system the entire region is segmented into smaller sub-regions in our case we have considered the state of Maharashtra. The entire state of Maharashtra is segmented according to the longitude and latitude. Further the landsat and multispectral images of these smaller regions are collected using the Google Earth Pro. Further, once these images are collected various features are extracted from the images. From the landsat images the percentage of water and vegetation in a region are extracted over the timeline. Similarly, from the multispectral images percentage of urbanisation is extracted. This data is stored in a .CSV format along with other parameters like the rainfall, temperature, etc. Further, the data collected will be split into training data and testing data. Firstly, drought prone areas will be identified using supervised/unsupervised learning algorithms. Hopfield Network and Hebb Learning Rule for the activation function is used to identify areas suitable for catchment of water around these drought prone areas. Parameter like altitude, rainfall, temperature, soil type, etc would play a vital role to determine whether a region is suitable for catchment or not.

This system focuses on the issues related to water scarcity in the rural parts of India, where there is limited water resources like lakes and rivers. The ground water level of these areas has also reduced drastically on account of various environmental issues like global warming, pollution, etc. The project aims at increasing the catchment areas that are hit by drought like conditions and thereby increasing the ground water level, which will eventually help in the long run. We aim at providing solution to make a region more sustainable. These areas not only lack the rainfall but also the way the water is stored and supply is managed. The system would be a great help for governing bodies which are currently spending time on surveys to find areas suitable for catchment. Our system would automate the search process and would also provide an optimum distribution plan from the predicted catchment areas.

II. Literature Survey

The literature survey was done in two phases:

A. Feature Extraction

The system requires a number of features to be extracted primarily the water body detection and the vegetation detection from satellite images. Satellite images downloaded first need to cleaned. The existing cleaning methods include the use of three filters bilateral filter, anisotropic filter and the mean-shift filter. Out of these three the bilateral filter works best with the very high resolution images as it preserves the contours and provides the right amount of smoothening[1]. Further, for the detection of water bodies a comparative study between various spectral indices is done. These include NDWI, MNDWI, NDVI and RE_NDWI. The existing water body detection systems include combination of two models the spectral model and the supervised model. The spectral model to identify water makes use of specific property of water in spectral bands. As water has the strongest absorption in NIR channel, its NIR reflection is able to show a great difference compared to other land covers. The supervised model has two steps the classification step and the learning step. The above two models are combined using the Dempster-Shafer theory [2]. Another possible approach uses the Fuzzy c-means (FCM) algorithm and canny edge detection to mark a water body. In this approach the images are filtered using Gabor filter further, FCM is used for land cover and land use classification. Finally, using canny edge detection edges of a water body are marked [3]. Vegetation detection is carried out using the supervised patch-level scoring process and an unsupervised pixel-level classification. [5][6].

978-1-5386-4985-5/18/$31.00 ©2018 IEEE

B. Prediction of catchment areas

This section needs support of existing study done with the help of Artificial Neural Network (ANN) to train the system to predict the best possible catchment area. We use Hopfield network along with Hebb Learning Rule to predict the catchment areas.

Hopfield Networks: Hopfield network, a recurrent artificial neural network is a binary threshold unit. This means a Hopfield network unit takes only two different values to represent their state which are determined by whether the unit's input exceed the threshold or not. Usually 1 & -1 (rarely 1 & 0) are the values taken by the Hopfield network.

For a Hopfield network with N nodes Hebbian Learning updates the connectivity weights according to the formula:

$$w_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \epsilon_i^{\mu}\epsilon_j^{\mu}$$

.

Where, εxμ is the state of node x in pattern μ

In a Hopfield network the energy associated with each node is given by:

$$E_i = -\frac{1}{2} I_i s_i$$

Where 'Ii' is the weighted input sum of node i and 'si' is the state of 'i'.

Thus, the total energy of any given Hopfield network with N nodes is given as:

$$E = -\frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} s_i s_j + \sum_{i=1}^{n} \theta_i s_i$$

III. Methodology

The proposed system would work in following stages:

A. Data Gathering and Mapping

This module is related to gathering the data related to environmental factors from different sources. This data is collected w.r.t. Latitude and Longitude, to understand the environment of a particular location. In this module we gather data related to following parameters: Rainfall, Temperature, Soil Type, Ground water (present/absent), Humidity, Percentage of area covered by water body, Evaporation rate, Altitude, Percentage of area covered by Urbanisation. Data related to groundwater will be calculated with the help of vegetation cover and water bodies in the area along with temperature and humidity. The data gathering for parameters like temperature, humidity, soil type, rainfall and altitude can be collected with the help of API's. But parameters like water body cover, urban cover and vegetation cover needs Image Processing to be performed on Landsat-5 and Landsat-8 satellite Images which will be covered in the pre - processing section.

B. Data Pre - Processing and Feature Extraction

extracted from API's won't need much pre - processing and can be directly used as features. It will just need to be transformed into a suitable format for further analysis. Unsupervised learning along with spectral model will be used to identify the cover of water body, vegetation and urban regions. The process of extraction of features from the images needs the images to be of high resolution to catch every minute detail. For extracting the features of Maharashtra, approximately 7.5 lakh satellite image needs to be processed for a single timestamp and the system will consider 20 such timestamp and hence the process of feature extraction from the images is a very tedious task. Hence, this process will be done using Hadoop Architecture with CUDA programming to enhance the process of feature extraction. Hadoop environment will ensure faster processing with the help of many data nodes.

C. Identification of Drought Prone Areas

Feature extracted will now be used to train the neural network algorithm to identify the drought prone areas. In this section, we present how the Hopfield Network will train the model to identify the drought prone regions w.r.t. Latitudes and Longitudes.

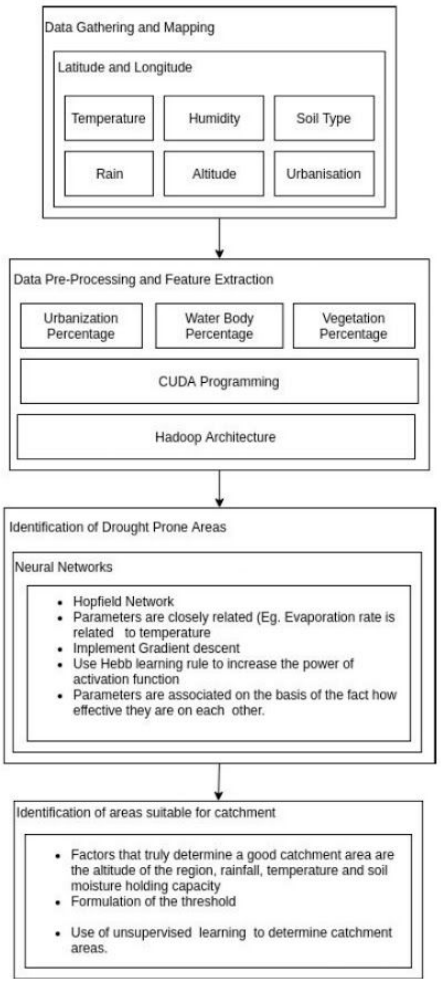


Figure 1 : Proposed System

We use neural networks as its goal approaches to combinatorial optimization to formulate the desired objective function being optimized, such that it can be viewed as a "natural" energy minimization problem. We have many parameters that are very closely related in this scenario for example the evaporation rate is directly proportional to temperature; moisture in soil is related to soil type, etc. All of these parameters with combinations will give different results and we need an optimized result that would help us detect the drought prone area. Hopfield networks implement a gradient descent method and with the help of Hebb Learning Rule we can increase the power of the activation function as the parameters are closely related. True advantage of using Hopfield-type neural networks to solve difficult optimization problems relates to speed considerations and hence we can consider various parameters and their association with each other to train the model [7].

To detect the drought prone areas we associate the parameters to each other on the basis of how effective they are on each other. Parameters like rainfall in the region, temperature in the region, soil type in consideration with water holding capacity of the soil, humidity will be strongly connected to each other as all will contribute to drought like conditions. Combinatorial optimization needs to be achieved with these parameters.

D. Identification of areas suitable for catchment

As we identify the areas that are hit by drought, our research to find a source of water starts exactly from the point where we found the hit for drought. We can help such areas in two ways. We can either help such areas to increase the ground water level or we can build some kind of reservoir for the area. All of these depend on many factors that will eventually help to harvest and conserve water for long duration of time. There are several factors that lead to drought like conditions. Few are natural and few are man-made.

Natural ways for drought to occur happens include parameters like Land and Water Temperature, Air Circulation and Weather Pattern (Humidity), Soil Moisture level and Soil Type.

Man made reasons for drought include population growth that increases the demand to be more than the supply, deforestation, constructions, global warming and agricultural needs. Agriculture needs include the type of crop that is grown and in which month it is grown and the amount of water it needs.

Factors that truly contribute for a suitable catchment area comprises of Altitude of the region, Rainfall, Temperature, Humidity, Soil Moisture holding Capacity. [8]As we used supervised learning for Identification of Drought Prone Areas, we have no particular measure to label a particular area to be fit for water catchment. So we formulate some threshold values for the different parameters of the environment that supports water harvesting and conservation. We can increase the water reserve for a region in two ways. Either we can improve the groundwater level or we can build reservoirs to store the water.

If we plan on increasing the ground water level then we search for areas with lower altitude, where the areas receives moderate amount of rainfall, where the soil water retention is higher and also where the temperature is not that high. But it's difficult to find such regions that might satisfy all the needs. For such reasons, methods like groundwater recharge, creation of wetlands and depression focus recharge is considered as an option to increase the groundwater level. So with the help of unsupervised learning and by using competitive learning rule, we can come out with combinations of parameters that might give us the best result for the region as all the combinations that are considered as output compete with each other and finally the winner takes all concept is implemented.[9][10] That's how we can choose the region with the output that gives the maximum value for a combination of parameter with the activation function.

IV. Results and Observation

The proposed system has observations of various phases like feature extraction, detected drought prone areas and predicted catchment areas

Feature extraction:

1) Water body detection

Figure 2 : Satellite image before highlighting water body

Figure 3 : Satellite image after highlighting water body

2) Vegetation detection

Figure 4 : Satellite image before highlighting vegetation

Figure 5 : Satellite image after highlighting vegetation

3) CSV file of water body percentage:

	A	B	C
1	Image Name	Water Percentage	
2	sriramsagar_1984.jpg	8.99	
3	sriramsagar_1985.jpg	9.1	
4	sriramsagar_1986.jpg	9.1	
5	sriramsagar_1987.jpg	9.1	
6	sriramsagar_1988.jpg	9.25	
7	sriramsagar_1989.jpg	9.24	
8	sriramsagar_1990.jpg	8.69	
9	sriramsagar_1991.jpg	11.54	
10	sriramsagar_1992.jpg	5.98	
11	sriramsagar_1993.jpg	5.41	
12	sriramsagar_1994.jpg	4.82	
13	sriramsagar_1995.jpg	6.11	
14	sriramsagar_1996.jpg	10.74	
15	sriramsagar_1997.jpg	8.71	
16	sriramsagar_1998.jpg	2.54	
17	sriramsagar_1999.jpg	6.83	
18	sriramsagar_2000.jpg	10.79	
19	sriramsagar_2001.jpg	3.54	
20	sriramsagar_2002.jpg	6.87	
21	sriramsagar_2003.jpg	3.89	
22	sriramsagar_2004.jpg	2.47	
23	sriramsagar_2005.jpg	1.98	
24	sriramsagar_2006.jpg	3.95	
25	sriramsagar_2007.jpg	6.37	
26	sriramsagar_2008.jpg	5.49	
27	sriramsagar_2009.jpg	3.98	
28	sriramsagar_2010.jpg	2.83	

Figure 6 : Results obtained from water body highlighting are saved in a .csv file

4) Gathering elevation for given place:

Figure 7 : Result provided by open-elevation API for given longitude and latitude

5) Gathering weather-data for given place:

Figure 8 : Result provided by darksky.net API for given longitude and latitude

6) Region affected by Drought:

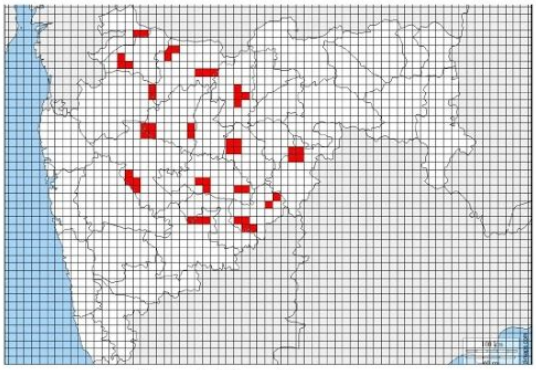


Figure 9 : Drought prone areas

7) Region identified as Catchment Areas:

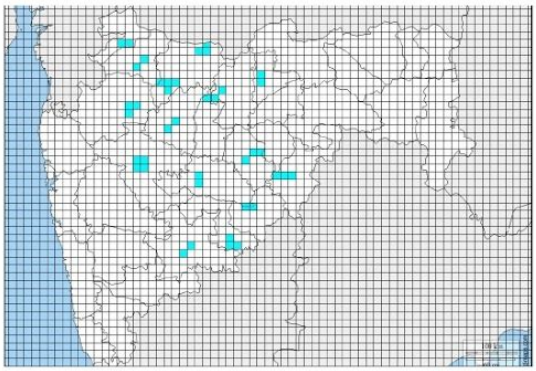


Figure 10 : Catchment areas predicted

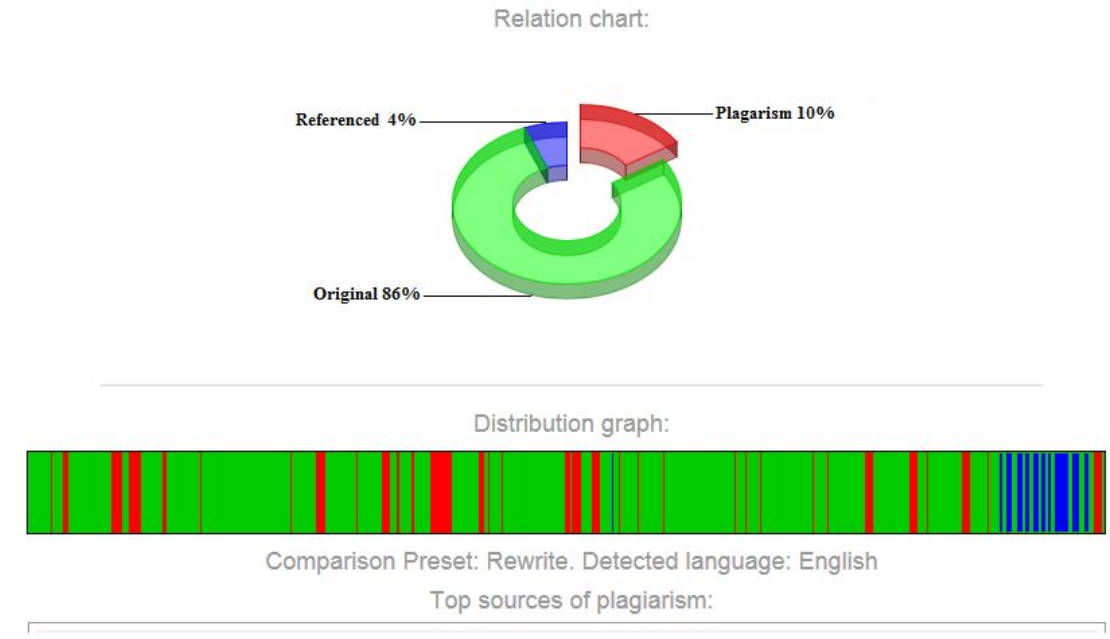
V. Conclusion

The changing rainfall patterns and rising global average temperatures are affecting the water holding capacity of soil and the rate of evaporation of water which has led to diminishing water bodies across many states in India, especially in the state of Maharashtra. Landsat 8 and Landsat 5 remote sensed satellite images have been used to identify the water bodies whose water surface area has reduced through the last three decades. Climatic factors like annual rainfall, soil type and altitude above sea level, humidity and area of vegetation have been used to identify actual and potential drought affected regions in Maharashtra. Hopfield networks, which implement gradient descent method and with the help of Hebb Learning Rule, has been found to be the optimal methodology to identify the drought regions. Further, with the help of unsupervised learning and by using competitive learning rule, we have come out with the right permutation of environmental parameters that have given us the best result for the region which are potential catchment areas of water. All the combinations that are given as output compete with each other and finally the winner takes all concept is implemented to identify the most suitable and optimal water catchment area near a drought affected region.

VI. References

1. Assia Kourgli ,Youcef Oukil, "Very High Resolution Satellite Images Filtering": Methods to filter or clean images
2. Na Li,Arnaud Martin,Remi Estival, "An automatic water detection approach based on Dempster-Shafer theory for multi-spectral images"
3. Zhang Ding, Ni Qi, Fang Dong, Li Jinhui, Yao Wei, Yuan Shenggui,"Water Bodies Identification From Mutispectral Images Using Gabor Filter, FCM and Canny Edge Detection Method "
4. T.Vignesh,K.K. Thyagharajan, ,"Application of Multispectral Remote Sensing Technology in Surface water body extraction "
5. Dipak R. Samal & Shirish S. Gedam , "Monitoring land use changes associated with urbanization: An object based image analysis approach"
6. Jiayuan Fan ; Tao Chen ; Shijian Lu , "Vegetation coverage detection from very high resolution satellite imagery"
7. Norbert A Agana, Abdollah Homaifar, "A Deep Learning Based Approach for Long-Term Drought Prediction"
8. Yudo Prasetyo, Setyo Ardi Gunawan and Zia Ul Maksum , "Determination of The Water Catchment Area in Semarang City Using a Combination of Object Based Image Analysis (OBIA) Classification, InSAR and Geographic Information System (GIS) Methods Based On a High-Resolution SPOT 6 Image and Radar Imagery"
9. Y. Takefuji, K. C. Lee, and H. Aiso, "An artificial maximum neural network: A winner-take-all neuron model forcing the state of the system in a solution domain," Biol. Cybern., vol. 67, pp. 243–251, 1992.
10. K. C. Lee, N. Funabiki, and Y. Takefuji, "A parallel improvement algorithm for the bipartite subgraph problem," IEEE Trans. Neural Networks, vol. 3, pp. 139–145, Mar. 1992.

2. Plagiarism report and Certificates of the paper published



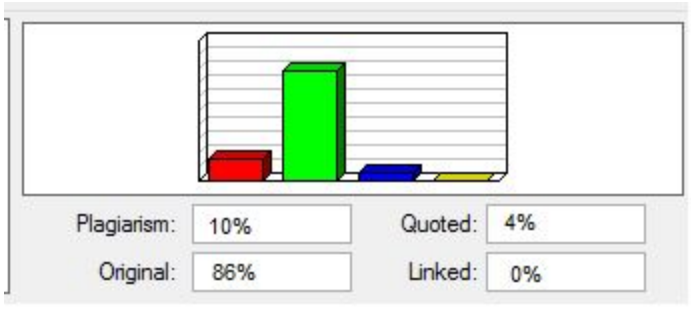


Figure 30: Plagarism Report

CERTIFICATE OF PARTICIPATION

This is to certify

Richard Joseph

has presented the paper entitled

Water Catchment Control and Management

in the 3rd International Conference on Inventive Computation Technologies (ICICT - 2018), organized by RVS Technical Campus during November 15-16, 2018 at Coimbatore, India.

IEEE Xplore Part Number:CFP18F70-ART

IEEE Xplore ISBN: 978-1-5386-4985-5

Session Chair Organizing Secretary Conference Chair

CERTIFICATE OF PARTICIPATION

This is to certify

Samujjwaal Dey

has presented the paper entitled

Water Catchment Control and Management

in the 3rd International Conference on Inventive Computation Technologies (ICICT - 2018), organized by RVS Technical Campus during November 15-16, 2018 at Coimbatore, India.

IEEE Xplore Part Number:CFP18F70-ART

IEEE Xplore ISBN: 978-1-5386-4985-5

Session Chair Organizing Secretary Conference Chair

CERTIFICATE OF PARTICIPATION

This is to certify

Ashish Joshi

has presented the paper entitled

Water Catchment Control and Management

in the 3rd International Conference on Inventive Computation Technologies (ICICT - 2018), organized by RVS Technical Campus during November 15-16, 2018 at Coimbatore, India.

IEEE Xplore Part Number:CFP18F70-ART

IEEE Xplore ISBN: 978-1-5386-4985-5

Session Chair Organizing Secretary Conference Chair

CERTIFICATE OF PARTICIPATION

This is to certify

Gurpreet Singh

has presented the paper entitled

Water Catchment Control and Management

in the 3rd International Conference on Inventive Computation Technologies (ICICT - 2018), organized by RVS Technical Campus during November 15-16, 2018 at Coimbatore, India.

IEEE Xplore Part Number:CFP18F70-ART

IEEE Xplore ISBN: 978-1-5386-4985-5

Session Chair Organizing Secretary Conference Chair

Figure 31 : Certificate of Paper Publication

3. Project review sheets

Inhouse/ Industry:

Class: D17 A/B/C

Group No.: 14

Project Evaluation Sheet 2018 - 19

Title of Project: Water Supply Management with catchment control using HPC

Group Members(sign): Samujjwaal Dey, Ashish Joshi, Gurpreet Singh

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	3	3	2	3	2	2	2	2	2	1	2	3	5	36

Comments: Very good concept, Algorithm is not clear, Need to study in detail

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	3	3	2	3	2	2	2	2	2	1	2	3	5	36

Comments: good concept, needed detailed study about collection of dataset and algorithm.

Date: __th September, 2018

Name & Signature Reviewer2

Inhouse/ Industry:

Class: D17 A/B/C

Group No.: 14

Project Evaluation Sheet 2018 - 19

Title of Project: Water Supply Management with catchment Control using HPC

Group Members(sign): Samujjwaal Dey (15), Ashish Joshi (23), Gurpreet Singh (39)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	3	4	2	2	2	1	1	2	2	3	4	35

Comments: Good project, Dataset Ready but Algorithms used each stage is not demonstrated.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	3	4	2	2	2	1	1	2	2	3	4	35

Comments: Good project, Data extraction clear but processing algorithms need to be studied in depth

Date: 15th October, 2018

Name & Signature Reviewer2

Inhouse/ Industry:

Class: D17 ~~A/B/C~~
Group No.: 14

Project Evaluation Sheet 2018 - 19

Title of Project: Water Supply management with catchment control

Group Members(sign): Samijjwaal Dey (15) Ashish Joshi (23) ~~Ashutosh~~ Gurpreet Singh (39)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	**Total Marks** (50)
Review of Project Stage 1	~~4~~	4	4	2	4	2	2	2	2	3	2	2	4	4	41

Comments: Running behind the schedule, ~~focus~~ focus more on implementation.

Yugchhaya Joshi
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	**Total Marks** (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2		2	3	2	4	4	41

Comments: integration of modules need to be done. behind the schedule

Date: 9th Feb, 2019

Name & Signature Reviewer2

Inhouse/ Industry:

Class: D17 ~~A/B/C~~
Group No.: 14

Project Evaluation Sheet 2018 - 19

Title of Project: Water Supply management & Catchment system using HPC

Group Members(sign): Gurpreet Singh (39) Ashish Joshi (23) ~~Ashutosh~~ Samijjwaal Dey (15)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	~~4~~	3	3	2	~~4~~	2	2	2	2	2	3	3	~~4~~	3	39

Comments: Running behind the schedule.

Yugchhaya Joshi
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	**Total Marks** (50)
Review of Project Stage 1	4	3	3	2	4	2	2	2	2	2	3	3	4	3	40

Comments: Need to be more specific in output

Date: 6th March, 2019

Anuja N
Name & Signature Reviewer2