Journal of
Water Resource Engineering and Management

ISSN: 2349-4336
Volume 11, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoWREM

https://journals.stmjournals.com/jowrem

Research JoWREM

Smart Water Harvester

Namratha Km[1], Lekha Achuth[2,*]

Abstract

Smart water harvester is a wiser use of Data Science in the optimization of rainwater harvesting, taking into account the forecast of precipitation and ideal catchment areas, and basically image processing using machine learning. In that respect, the system, via predictive algorithms like Random Forests, predicts the amount of rainfall by taking into consideration historical and real-time data, while Digital Elevation Models (DEM) and visualization methodologies of images make geographical area and rooftop surfaces analyze to choose the best water collection surface area. Together with the identification of optimal locations for catchment areas, the system helps in assessing the amount of rainwater that can be collected, including ways of efficiently constructing and managing water storage systems. This tool is particularly useful in urban areas, as catchment rooftop rainfall is usually expensive and cumbersome to undertake without proper guidance. Therefore, the smart water harvester not only reduces these obstacles but also has an added advantage in encouraging the conservation of water supplies, since non-drinking water consumption is incentivized through possible tax credits. Meanwhile, it harnesses methods of data science, such as multivariate regression and watershed algorithms, for high-tech information on rainwater catchment. This helps decision-making at individual homeowners, business enterprises, and municipalities in terms of the installation of centralized rainwater storage systems that reduce dependency on the traditional water supplies and improve resource management. This is done with an overall view of realizing sustainability in water use practices and promoting widespread rainwater harvesting as a better practice of supporting water conservation.

Keywords: Catchment area, image visualization, random forest, digital evaluation model, data science

INTRODUCTION

Smart Water Harvester is a data science predictionwhich has the capability of estimating the amount of rain- fall and using image visualization, to estimate the catchmentarea. As a result, the total amount of water may be accumulated and the best location to build the catchment area can bedetermined. Consequently, non-drinking water use and associated tax advantages gain. The prediction's overall goal is to save the water that has been collected through rainwater gathering. This encourages regular people to start efforts to collect rainwater for their homes or other structures. The total catchment area brings in the different ways in which the user can collect the rainwater over the roof by image visualization and hence helps to analyze the same. At the moment, building water storage tanks separately for each building or apartment leads in substantial investment expenses for individuals or groups when it comes to rooftop rainwater collecting. There is no process or application available to determine the advantages of such facilities, how well these setups can utilize shared storage tanks, or the minimum capacity neededfor these shared tanks.

***Author for Correspondence**
Lekha Achuth
E-mail: lekha@pes.edu

[1]Student, Department of Computer Application, PES University, Bengaluru, Karnataka, India
[2]Professor, Department of Computer Application, PES University, Bengaluru, Karnataka, India

Received Date: August 22, 2024
Accepted Date: September 03, 2024
Published Date: September 10, 2024

Citation: Namratha Km, Lekha Achuth. Smart Water Harvester. Journal of Water Resource Engineering and Management. 2024; 11(3): 1–7p.

LITERATURE SURVEY

Rainfall-induced collapses along mountain highways are common catastrophes that result in severe fatalities among humans. Therefore, the capacity to conduct a proper landslip review is going to be highly helpful for governmental organizations and land architects practicing utilization to create suitable approaches to mitigation. Here, we present an artificial learning approach based on a random forest classifier (RFC) and a GIS-based dataset for the geographic forecast of flooding-induced landslides along highland routes. The RFC is used as a method for supervised learning to generalize the category limit that divides the input data of ten landslide factor conditioning (slope, aspect, relief intensity, top shape, topographic wetness measure, separate to paths, distance to waters, lithology, keep to errors, and rain) into two distinct class labels: 'landslide' and 'non-landslide'. The proposed model provides a better level of prediction accuracy, as shown by findings from experiments with a Cross validation procedure and sensitivity assessment on the RFC model parameters, with an area under the trajectory of 0.92. Other benchmarking techniques, including as discriminant analysis, logarithmic regression, synthetic neural networks (ANN), relevance machine learning and support vector machines, are greatly outperformed by the RFC. We strongly propose the RFC as a very effective tool for geographic modeling of rainfall-induced landslides based on our experimental results and comparison research [1].

The combination of properties found in random forests (RF) systems makes them ideal for use in remote sensing of precipitation. When dealing with complicated non-linear phenomena like precipitation, it is crucial to be able to capture non-linear associations of structures between predictors and response. This work focuses on attributing rainfall amounts to depositing cloud areas in conjunction with extratropical storms in midlatitude, encompassing both stormy and advective-stratiform precipitated cloud areas, due to the shortcomings of existing visual rainfall retrievals. As a result, the rainfall levels are allocated to rain zones that have been previously defined and categorized in accordance with the processes that lead to precipitation. Knowledge on the cloud top height is included using water vapor-IR variations and IR sky top temperature as prediction factors [2]. In a variety of geo-environments, landslides are one of the geological dangers that might occur. In this work, we assess the performance of two cutting edge machine learning models, namely Decision tree algorithms and a random forest algorithm, to estimate the occurrences of major rainfall-triggered landslides in volcanic areas, east Asia [3]. Assuming inundation has occurred up to a certain level h, each already-found catchment basin specifically, those watershed areas which associated lowest has an altitude lower than or equal to h—should have its own label. The elevation h plus one pixel will then be calculated. These images are placed in a first-in-first-out queued if not less than one of their neighbors is labelled. Starting from these pixels, the queue structure enables to extend the labeled catchment basins including them by computing geodesic influence zones. The remaining unlabeled stage h plus one pixels are then considered new minima and allocated new tags. Those pixels with an identical geometrical distance to distinct catchment regions are labelled as watershed pixels in the end product, whereas all the pixels belonging to the identical catchment region have another label. Thus an ecological change is understandable and comes pretty near to the basic idea of a floodplain. But given that there is a second process of organizing, it is slightly more difficult [4]. Upon reviewing the prior investigation and polling [5]. Many waters from rain collecting devices aren't built to conserve as much rain as they might. Rising urbanization necessitates the usage of municipal gathering of water, a technique that makes use of a combined freshwater catchment for the neighboring society in addition to the implementation of tedious laws and regulations. Due to the complexity of the process, determining the feasibility of this kind of a framework requires extensive physical scrutiny. Using aerial photograph categorization, height estimate, precipitation forecasting, and optimal barrel shipment, this research proposes a technology built around computer vision, machine learning, and other techniques that optimizes every step of the procedure. The appliance can therefore assess the viability of a project and forecast the point at which it becomes profitable using the determined criteria.

DISCUSSION

The system mainly consists of two major modules. Phase 1: Rainwater prediction 1) Multivariate linear regression (MLR) 2) Random Forest (RF) Phase 2: Catchment area Detection: 1) Canny Edge Algorithm 2) Watershed 3) Digital Elevation (DEM). The overall process flow will start with the loading of data from text file and converting to the pandas csv file the loaded data is being preprocessed by filling the drop values and label encoding the categorical variables the data is next being divided into test and train dataset the testing dataset. Using the testing dataset, the model is built the built model is being tested on the testing data the accuracy and the mean error value is calculated and the desired result model is taken into consideration. The rainfall prediction is per- formed using this model. The CSV file is loaded into the data frame the NA values are treated and the Station field is label encoded the model is created from Multivariate Linear regression Random Forest Regressor. The model that produces the most accurate results will be saved in a pickle file. Pickle is a is a module that may be used to serialize Python instance constructions, which is the process of turning an item in memory into a stream of byte values that can be saved as a digital file on disc. Later references to the same item won't be serialized again since Pickle maintains track of the things it has previously serialized. One issue that has continuously plagued data analytics is the detection and cleaning of dirty data, and its negligence may lead to flawed analytics and unreliable decisions. In recent years both industry and academia seem to have given increased attention to issues regarding data cleaning, concentrating on new abstractions, user-interface features, scalable methods, covering, and statistical techniques.

Therefore, the capacity to conduct a reliable landslip appraisal will be highly helpful for federal agencies and practicing land utilization managers to create suitable remedies. Through a number of illustrated instances, we will discuss the cutting-edge approaches while also highlighting their shortcomings. We also address new work that incorporates such techniques into a mathematical projection The structure even though historically these techniques are separate from statistical techniques like outlier identification. evaluating the implications of data washing on mathematical modelling and utilizing neural networks to increase the efficacy as well as precision of data purification [6]. Considering it needs both a recognition of the information organization and the calculation of the degree of noise in order to eliminate anomalies and save the ”best” information, information cleaning is typically more challenging that screening. A crucial component of data preparation is a highly effective digital information better. The latest solution for web-based information screening despite the usage of model-based processes was put out. To sort the input on the web, they used fractal detection on a dynamic window of collaborative measurement, and to balance aberrations as well as process changes, they used a limited impulse action median fusion filter. Although what is filtered might eliminate a significant amount of the underlying input's content, the offered screen concentrates mostly on data filtering for multiscale process data tainted with anomalies rather than data preparation. This is essential to get the data ready to ensure that the improved filter-cleaner just eliminate anomalies while maintaining all other accurate data facts, making it a key need. Being a sort of filter, there is, nevertheless, a little chance of excluding “beneficial anomalies”.

METHODOLOGY

The consumer connected for the visual plane will allow for computer interaction. The company is tied to this display layer. The company layer is going to communicate with the capabilities offered by the program's information manufacturing, rainfall harvesting, picture being processed, and watershed identification in the layer upon which it is possible to build the app's functionality. Information from the Knowledge Acquisition tier is supplied into the provider layer. Knowledge collected by the Indian Metrological Survey of India, which is immediately available for input into the simulations, is included in the forecast. Additionally, the data recuperation includes multiple panoramas of various cities and locales that are supplied into the database in order to educate the knowledge. The data is being taken from the Indian meteorological survey and they are the real-time data of Bangalore city.

Random Forest Algorithm

Breiman was the first to design random forests (RFs), which are extremely versatile and effective ensemble classifiers constructed using decision trees. lately there has been an increase in interest in RF, and it has been used in a variety of contexts to address categorization issues. However, there aren't many instances where RFs are used to categorize rainfall situations In Besut station, on the east coast of Peninsular Malaysia, rainfall incidence was recently proposed as the sole such use of RFs. The number of variables being measured (mtry) and the quantity of trees (ntree) are the two calibration parameters for RFs [7]. RF is an incorporated machine learning approach that builds many unrelated decision trees using the method known as random resampling bootstrap methodology and the node stochastic classification strategy, then votes to determine the final classification outcomes. Features having complicated relationships can be analyzed via RF. The RF technique provides a quicker learning rate and improved resistance to distortion and data deletion. It is possible for one to use the variable significance measurement as a tool for choosing features for the given data [8]. The Clustered Random Forest technique is applied to the detection of intruders in a network. This works by segmenting a given network into different character systems for manners of communication, all of which can be analyzed in detail. It examines every node in the network and creates log files that are given as input for the proposed technique. In the most current research on this subject, RWH zones are chosen using geographic data mining (GIS), multicriteria analysis, and hydrological models. However, there are certain studies that identify potential zones taking socioeconomic factors into consideration, typically in confined areas [9]. Time series modeling is being applied on the data to get the previous year rainfall data. In the past, time-series modeling has been a significant area of academic study, with applications in fields like climate modeling. For a specific entity iii at time ttt, time-series forecasting models predict future values of a target variable yi,ty_{i,t}yi,t For a certain entity i at time t, time-series forecasting models anticipate future values of a goal yi,t. Each one may be viewed simultaneously and constitutes an ordered collection of temporal data, such as observations from several meteorological stations in climatology or vital signs from various patients in medicine. Due to its complexity and enduring uses, such as disaster predicting along with tracking pollutant level concentrations, among others, rainfall forecasting has recently earned the uttermost scientific significance. Therefore, methods that combine time-series data with Machine Learning algorithms are being investigated as a substitute to get over these problems [10]. This paper presents a detailed discussion on the major challenges of computer architecture and system design that could be mitigated by using ML methodologies and also discusses the popular ML strategies to handle these issues. Though the focus is essentially on computer architecture, the view is expanded to cover data centers also as big warehouse computing systems. We briefly cover some related areas including code generation and compilers, and investigate how ML can substantially affect and change design automation. We also discuss some future opportunities and directions and anticipate that the contribution of including ML in computer architecture and systems will continue to rise and have significant impacts [11]. Machine learning has now turned a strong compel for unleashing the power of data-driven enterprises to drive innovation, ensure efficiency, and advance sustainability. However, most practical applications of machine learning disappoint for various reasons. Few projects of machine learning ever succeeded, while many proofs-of-concept never turned into large-scale deployments. This is not surprising from a research point of view, since, so far, the ML community has focused on creating ML models rather than on (a) transforming them into production ready ML solutions and on (b) developing the necessary integration and coordination of the various complex system components and infrastructure, together with defining the roles for automating and managing ML systems in practical environments [12].

Digital Elevation

A Digital Elevation Model develops the bare terrestrial surface in a grid system that is referenced to a vertical datum. The process of removing non-ground features, such as bridges and roads, produces a refined digital elevation model. Constructed objects of human origin Examples include power lines, structures, and towers as well as natural features, such as trees and all other forms of vegetation are consistently removed from the DEM. Digital Elevation Models (DEMs) are essential products of aerial photography and photographing and one of the main layers of topographic databases at the regional, local, and global levels of the geographic data infrastructure. It is frequently necessary to combine information from multiple investigations in order to provide a holistic viewpoint and direct future study. In reaction this research offers an in-depth investigation of DEM fusion, including a comprehensive review of the initial processing efficiency of operation, methodologies and applications. Still-unresolved issues and open concerns are recognized through debate and comparing and contrasting, and new study possibilities are suggested [13].

Watershed Algorithm

In watershed segmentation, an image is viewed as a topographic landscape of bright crests and dark valleys. The height of this landscape is usually defined from the gray values of the pixels or their magnitude of gradient. This 3D representation is utilized to divide an image by the watershed transform into catchment basins. A catchment basin consists of all points from where the steepest descent leads to one specific local minimum. Watersheds are the divisions that separate these basins. The watershed transform is a complete segmentation of the image that assigns each pixel to a particular region or a watershed.

IMPLEMENTATION

Data

Smart water prediction data is collected from the Indian Meteorological Department for the prediction of rainwater and it has 400 records and 27 stations in Bangalore city. It contains the different station data. The images of the building's rooftop is collected from the Earth explorer. The data will be using time series to get the previous year and the next previous year data set. In this way the data can reflect the different months and its consecutive months data in the dataset Figure 1.

RESULTS

The Random Forest Regressor Model is developed with the desired attributes of estimators and random states. With more trees included in the algorithm, an Algorithm for Random Forests becomes more accurate and capable of addressing problems. Another classification algorithm called Random Forest employs several decision trees on services for low input information to boost the dataset's predicted accuracy. The Random Forest model is the most effective of the three models. The random forest shows a similar trend with the training and testing set model on the real data set Figure 2 and 3.

1	Station	Date	Rainfall
2	Anekal	01-01-2010	3
3	Anekal	01-02-2010	8
4	Anekal	01-03-2010	0
5	Anekal	01-04-2010	49.2
6	Anekal	01-05-2010	164
7	Anekal	01-06-2010	47
8	Anekal	01-07-2010	111
9	Anekal	01-08-2010	101.6
10	Anekal	01-09-2010	51.6
11	Anekal	01-10-2010	82.8
12	Anekal	01-11-2010	163.4
13	Anekal	02-12-2010	5.8
14	Anekal	01-01-2011	0
15	Anekal	01-02-2011	5.8
16	Anekal	01-03-2011	7.2
17	Anekal	01-04-2011	143.2
18	Anekal	01-05-2011	132.5
19	Anekal	01-06-2011	55.8
20	Anekal	01-07-2011	62.6
21	Anekal	01-08-2011	231
22	Anekal	01-09-2011	66.4
23	Anekal	01-10-2011	249
24	Anekal	01-11-2011	92.4
25	Anekal	01-12-2011	6
26	Anekal	01-01-2012	0
27	Anekal	01-02-2012	12

Figure 1. Model process flow.

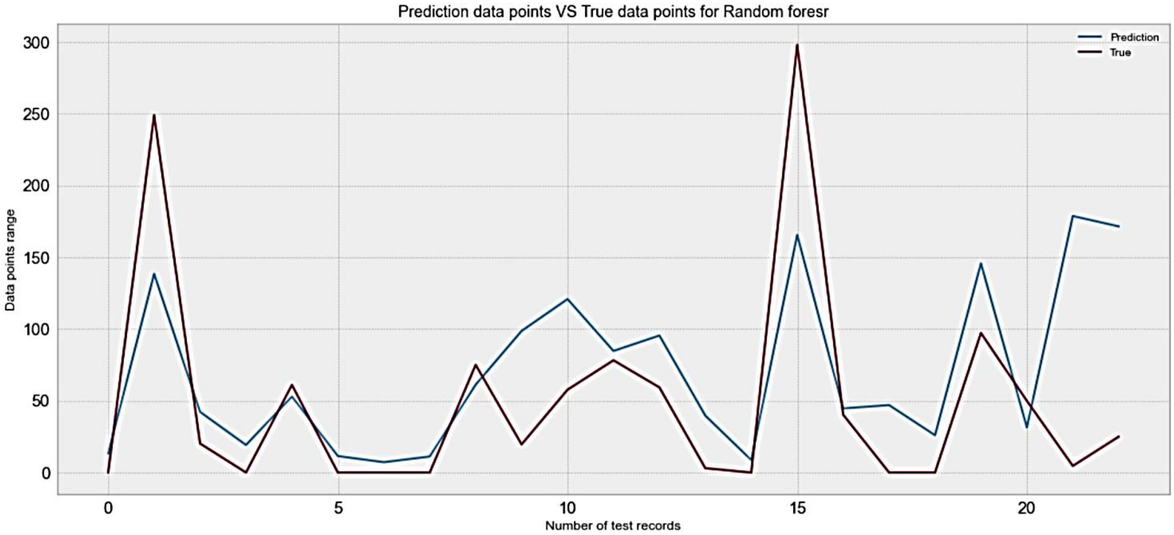

Figure 2. Graph of random forest.

Test case ID	Test Case Scenario	Algorithm	Mean Absolute Error	Score	Selected model
TC01	Smart water harvester Model	Multivariate linear regression	5998.90	12	NO
		Random Forest regression	854.24	77	Yes
		XGBoost	6077.90	40.76	NO

Figure 3. Comparison of different Algorithm.

CONCLUSION

The total water the user can save in the building can be predicted from this application by using the random forest model. This enables the preparation and full utilization of water that has been saved. The total catchment area are being detected using the digital elevation model and the watershed algorithm. The Smart water harvester helps to predict rainfall in a given area and helps in the detection of the total catchment areas. The total water which can be stored in the building can be predicted using highly learned and powerful machine learning models like the random forest model and watershed algorithm. The random forest model will have an accuracy of 854.24 mean absolute error and a score of 77 percent. Optimize the placement of centralized rainwater collection tanks and determine the quantity of water that might be collected using image and dwelling data. This is an application where in users can understand the mount of rainwater they can save from rainfall in the given area of the city. Rainfall forecasting is a challenging task to predict factors associated with rainfall such as wind, humidity, and temperature. Basically, rainfall forecasting task is usually performed using supervised learning techniques. Since there are many different supervised learning techniques, different performances could be gained from them. In addition rainfall data could be formed in different forms including long-terms (e.g. months) and short-terms (e.g. daily). Therefore, selecting an appropriate technique for a specific duration of rainfall is a crucial task.

REFERENCES

1. Viet-Hung Dang, Tien Bui Dieu, Xuan-Linh Tran, and Nhat- Duc Hoang. Enhancing the accuracy of rainfall-induced land- slide prediction along mountain roads with a gis-based ran- dom forest classifier. *Bulletin of Engineering Geology and the Environment*, 78:2835–2849, 2019.
2. Meike Ku¨hnlein, Tim Appelhans, Boris Thies, and Thomas Nauss. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — a random forests- based approach applied to msg seviri. *Remote Sensing of En- vironment*, 141:129–143, 2014.

3. Jie Dou, Ali P. Yunus, Dieu Tien Bui, Abdelaziz Merghadi, Mehebub Sahana, Zhongfan Zhu, Chi-Wen Chen, Khabat Khosravi, Yong Yang, and Binh Thai Pham. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the izu-oshima volcanic island, japan. *Science of The Total Environ- ment*, 662:332–346, 2019.
4. Han Sun, Jingyu Yang, and Mingwu Ren. A fast watershed algorithm based on chain code and its application in image segmentation. *Pattern Recognition Letters*, 26(9):1266–1274, 2005.
5. Chalachew Muluken Liyew and Haileyesus Amsaya Melese. Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8:1–11, 2021.
6. Xu Chu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. Data cleaning: Overview and emerging challenges. In Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, page 2201–2206, New York, NY, USA, 2016. Association for Computing Machinery.
7. Quoc Bao Pham, Tao-Chang Yang, Chen-Min Kuo, Hung-Wei Tseng, and Pao-Shan Yu. Combing random forest and least square support vector regression for improving extreme rain- fall downscaling. *Water*, 11(3), 2019.
8. Dongxiao Niu, Keke Wang, Lijie Sun, Jing Wu, and Xiaomin Xu. Short-term photovoltaic power generation forecasting based on random forest feature selection and ceemd: A case study. *Applied Soft Computing*, 93:106389, 2020.
9. Amirhossein Shadmehri Toosi, Erfan Ghasemi Tousi, Seyed Ali Ghassemi, Ali Cheshomi, and Sina Alaghmand. A multi-criteria decision analysis approach towards efficient rainwater harvesting. *Journal of Hydrology*, 582:124501, 2020.
10. Ari Yair Barrera-Animas, Lukumon O. Oyedele, Muhammad Bilal, Taofeek Dolapo Akinosho, Juan Manuel Davila Delgado, and Lukman Adewale Akanbi. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7:100204, 2022.
11. Nan Wu and Yuan Xie. A survey of machine learning for computer architecture and systems. *ACM Comput. Surv.*, 55(3), feb 2022.
12. Dominik Kreuzberger, Niklas Ku¨hl, and Sebastian Hirschl. Machine learning operations (mlops): Overview, definition and architecture. *IEEE Access*, 11:31866–31879, 2023.
13. Chukwuma J. Okolie and Julian L. Smit. A systematic re- view and meta-analysis of digital elevation model (dem) fusion: pre-processing, methods and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:1–29, 2022.

