Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 28

CHANGE DETECTION AND PREDICTION OF LAKE WATER USING MACHINE LEARNING TECHNIQUES

Harshini K A
Student
Department of CSE,
Sathyabama Institute of Science and Technology.
kaharshini@gmail.com

Hamsavarthini E
Student
Department of CSE,
Sathyabama Institute of Science and Technology.
hamsavarthini16@gmail.com

Dr. P. Sardar Maran
Assistant Professor
Department of CSE,
Sathyabama Institute of Science and Technology.
psmaran@gmail.com

Dr. B. Prabhu Dass Batvari
Assistant Professor,
Scientist C,
Sathyabama Institute of Science and Technology.
batvarib@gmail.com

Abstract-Regression analysis, which predicts changes in the river area over time, may be used to forecast the river area using machine learning. Using historical river area data collected at various times in time, regression models may be trained to predict future river areas. Regression analysis's potential to forecast changes in river area over time using machine learning is tested in the proposed study. In order to forecast river area based on the data at hand, we employed linear regression, hill regression, and random forest regression models. Relevant information must be gathered over time, such as the river area and other variables that may impact the river region, such as climate, dry land, etc., in order to apply this method to real-world scenarios. usage and water's velocity. Future changes in the stream may be predicted using the regression model after it has been trained using the data. Machine learning may be used to predict changes in river basins for a variety of purposes, including urban planning, water resource management, and environmental protection. Stakeholders may prepare for and react to anticipated effects in their immediate environment, such as floods or droughts, by forecasting changes in the watershed. In the project, data is stored in a CSV file in square kilometers (square kilometers) time (years). The project's goal is to provide examples of how to apply regression models. predicting changes in ecological or environmental factors using data that is already available. The Rettri, Korattur, Ambattur, and Puzhal rivers are the ones whose data were chosen.

Keywords: Machine Learning, Prediction, Linear Regression, Random forest, Ridge Regression.

I. INTRODUCTION

Rivers are an essential natural resource that help with transportation, recreation, and many other facets of daily life. Nevertheless, a river's area may fluctuate through time as a result of a variety of variables, including climatic changes, changes in land use, and human activities. It is essential for managing water resources,

developing cities, and monitoring the environment to be able to predict how river areas will change over time. A potent technology that may be used to forecast changes in river areas is machine learning. may be used to develop a model that forecasts the position of a river using historical information. Regression models may be used to discover trends and patterns in data and predict possible changes in the river region.

The usage of this method may be used to create preventative steps to deal with anticipated problems caused by a river's changing area, such as managing water resources, creating flood control plans, and preparing for changes in land use. Stakeholders can make better management choices for this crucial resource to support human activities and safeguard the environment by utilizing machine learning to forecast changes in river area.

Machine learning has recently grown in prominence in the field of environmental modeling and management due to its capacity to handle enormous datasets and spot patterns that human analysts might not immediately recognize. One of the most popular forms of machine learning models for environmental prediction tasks, such as river area, is regression analysis.

In this code snippet, we forecast changes in river area over time using three distinct regression models: linear regression, ridge regression, and random forest regression. The historical data used to train the models is supplied from a CSV file with the year and location of the river. The trained models are then used to estimate future changes in the river region. We then evaluate the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 28

three models' performances by calculating their mean squared errors.

In order to visualize the outcomes of the model predictions and spot any patterns or trends in the data, scatter plots and line plots are used to depict the results of the model predictions. Consequently, we create a.pkl file from the trained random forest regression model that can be utilized for.

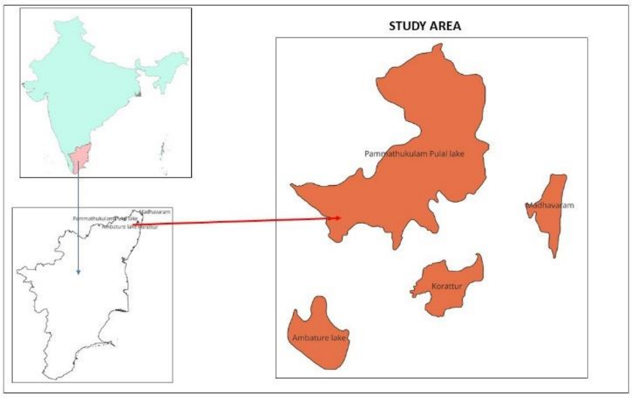


Fig 1.1 Study Area

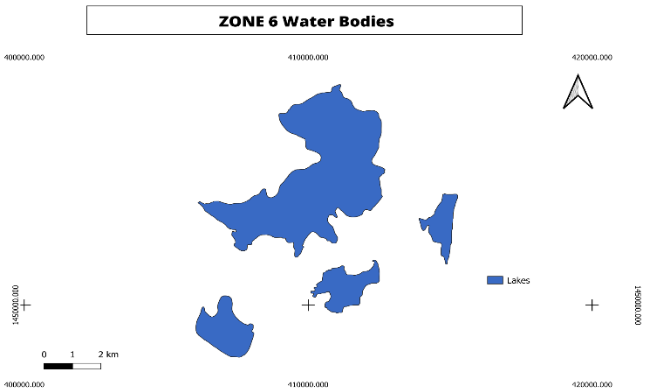


Fig 1.2 Topography

II. LITERATURE SURVEY

[1] One study by Singh et al. (2016) investigated the relationship between land use and river area in the Gomti river basin in India. The study used Landsat images and GIS techniques to extract land use information and estimate the river area. The results showed that the river area was significantly affected by land use changes, particularly the expansion of urban areas and agricultural land. The study concluded that land use planning is essential to maintain the ecological integrity of the river basin.

[2]Another study by Zhang et al. (2019) used a machine learning approach to predict river area based on meteorological data. To anticipate river area in China's Yellow River basin, the researchers used random forest, support vector regression, and artificial neural network models. With an accuracy of 0.86, the random forest model performed the best. The study suggested that machine learning models could be a useful tool for river area prediction.

[3]Similarly, a study by Zhu et al. (2017) used a support vector regression model to predict river area based on hydrological data. The study used data from the Qiantang River basin in China and found that the support vector regression model had a higher accuracy than traditional Multiple linear regression and ANN are examples of predictive methods. The study suggested that machine learning models could be a valuable tool for river area prediction in areas with complex hydrological conditions.

[4] Mueller, N. et al. (2016) For obtaining information on surface water, researchers studied 25 years of Landsat images across Australia. The study instrument, known as WOfS (Water Observations from Space), assisted in better identifying and managing Australia's surface water resources. The maps created by WOfS enabled a new method of retrieving information. The combination of massive amounts of rising data linked to superfast processors and multispectral satellite information in a traditional matrix format had permitted the establishment of a single analysis that can be performed consistently across data. The technique required around 2 1013 observations and 184,500 scenes over 27 years. The continental map of surface water was created in this manner.

[5] Another investigation conducted by Zhang et al. (2018), the authors investigated the relationship between river area and rainfall in the Yangtze River basin in China. The study used a random forest model to predict the river area based on rainfall data The random forest model has a high accuracy of 0.87, according to the results., indicating that rainfall was a significant factor in determining river area in the basin.

[6]Another study was conducted by Cheng et al. (2019) to predict the river width using a hybrid machine learning approach. The river width was predicted using

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 28

the SVM and the DNN .The findings demonstrated that the hybrid technique could properly forecast river width., especially in the areas where the river width was less than 100 meters.

[7]In another study, Lian et al. (2020) used the deep learning technique to predict the river area of the Yangtze River in China. The study used data from 2003 to 2018 and the outcome demonstrated that the deep learning technique capable of making precise predictions of river area, especially in the upstream region.

[8]Furthermore, a study by Chen et al. (2019) used a deep learning technique to predict the water level of the Yangtze River in China. The study used data from 2009 to 2017 The results demonstrated that the deep learning approach was capable of properly predicting water level, especially during flood seasons.

III. EXISTING SYSTEM

(i) Planimetric method: Using this approach, the river is divided into numerous divisions, the width of each segment is determined, and the area is calculated using straightforward geometric calculations. Although this approach is straightforward and reasonably priced, it does not take into consideration variations in river depth or slope, which might have an impact on the accuracy of the results.

(ii)Cross-sectional method: This approach entails taking many measurements of the river's width and depth along a cross-section in order to determine its area using integral calculus. Compared to the planimetric approach, this method can produce more precise findings, but it necessitates more exact measurements and computations.

(iii)Remote sensing method: The surface area of the river is mapped using satellite data or aerial photos. Although this approach can produce extremely precise findings, it needs specific tools and knowledge to evaluate the data.

(iv)Computer modeling method: This approach requires simulating the river's water flow using computer models, then calculating the area based on the simulated data. This methodology can produce quite precise findings, but it necessitates a thorough understanding of hydrology and computer modeling.

Disadvantage of Existing System

Planimetric method: This approach implies that the river's width is constant throughout.

Cross-sectional method: This procedure necessitates exact measurements of the river's depth and breadth at several locations along a cross-section, which can be time-consuming and expensive.

Remote sensing method: This approach can be costly and time-consuming as it takes specific tools and knowledge to understand the data.

IV. DATASET

For our analysis, we employed a dataset of satellite imagery. The collection consists of images from the Landsat OLI satellite with high resolution.

In this project, the datasets from the USGS Earth Explorer website were used to study the changes in water bodies progressively.

The study’s goal was to analyze the changes in the water bodies in a particular area over the last four - five decades. The data was collected for each decade, starting from 1978 to 2023. The area under study was delineated using QGIS, a popular Geographic Information System (GIS) software.

The first step in the study was to prepare the map of the study area. This was done using QGIS, which allowed the us to identify the boundaries of the study area accurately. Once the study area was delineated, the water bodies within this area were identified and delineated. This process involved separating the water bodies from the landmass using the satellite imagery available on the USGS Earth Explorer website.

Once the water bodies were identified and delineated, we processed the data to convert it into shape files. Shape files are a type of GIS data format that can be used to store geographic data. They are used to represent points, lines, and polygons on a map.

The next step in the study was to calculate the variation in the size of the water bodies over each decade. The data was analyzed to determine how the size of the water bodies had changed over time. Then we calculated

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 28

the variation in square kilometers for each decade, which allowed us to quantify the changes that had occurred.

The quality of the data used to build predictive models depends on that data, hence effective data gathering techniques are essential to creating high-performing models. Rettri, Korattur, Ambattur, and Puzhal

V. PROPOSED SYSTEM

Fig 5.1 Block diagram

Data collection – rettri, korattur,ambattur,puzhal

Data preprocessing **-** The data set's training and testing are then applied to them. The dataset in this case is machine-tested and machine-trained.

The dataset is then fed into a collection of machine learning algorithms, where a number of them are contrasted to see which produces the greatest accuracy results. The final result is shown in a GUI interface and the automated machine learning methods are applied. - rettri, korattur, ambattur, puzhal future area prediction.

Algorithms:
Regression algorithms is a traditional statistical approach that is often utilized for regression tasks, which involves predicting a continuous numerical value.

In the case of river area prediction, we have a dataset that includes information about the year and the river area for several rivers (Rettri, Puzhal, Korattur, and Ambattur).

The objective is to create a model that can detect lake area. for a given year based on the historical data.

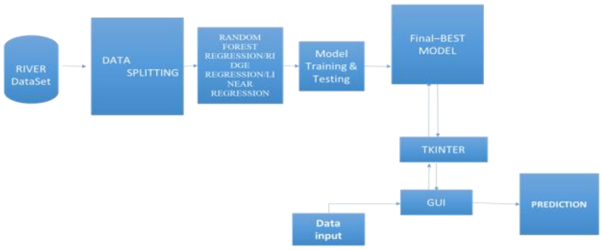


Fig 5.2 system architecture

RANDOM FOREST REGRESSION

Random forest is a suitable algorithm for this task because it can handle non- linear correlations between the input (year) and output variables(area).

It performs by generating several decision trees and then combining their forecasts to get a final prediction. Each decision tree is constructed using a random portion of the input variables, which reduces overfitting and improves model accuracy. overfitting and increase the model's accuracy.

LINEAR REGRESSION

Linear regression is an appropriate approach for this assignment since it can describe the linear connection between the input variables (year) and the output variable (river area).

It functions by fitting a straight line to the data that optimizes the sum of the squared errors among expected and actual values.

However, it is worth noting that linear regression assumes that the relationship between the input and output Variables are linear, which might not constantly be the case in real-world scenarios. More complicated algorithms are used in such cases. such as random forest regression may provide better accuracy.

RIDGE MODEL REGRESSION

Ridge regression is a preprocessing step that adds a penalty component to the cost function of linear regression. This penalty factor prevents overfitting by reducing the values towards 0, which can be advantageous when working with datasets with a significant number of features or strongly correlated features.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 28

Using Ridge regression can help to mitigate the effects of this correlation and prevent overfitting.

Additionally, Ridge regression can also help to improve the stability of the model by reducing the variance of the coefficient estimates. This can be extremely beneficial when working with datasets that include a high degree of noise or when there are missing values in the dataset.

VI. OUTCOMES

- We have developed a machine learning-based model technique that calculates intense water levels using historical data.
- The estimated performance of three regression models was investigated (Linear, Random Forest, and Ridge Regression).
- Random Forest Regression (RFR), which has a low error index, was chosen as the machine learning method. For the best performance in estimating water levels, we used the RFR algorithm in our system.

VII. CONCLUSIONS

As a result, this study uses this strategy to identify the body of water.

We use annual river data from 1978 to 2020 for this purpose.

The identification performance for each research region is compared using Linear Regression, the Random Forest technique, and Ridge regression.

The accuracy indices and local comparison are used to evaluate the classification results and used to predict the future.

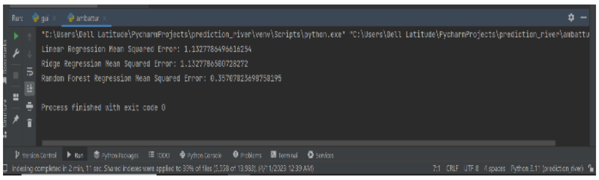


Fig 7.1 Mean Square Error

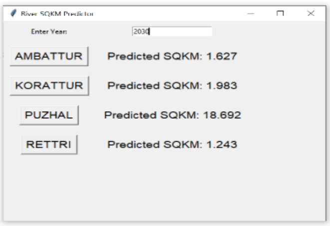
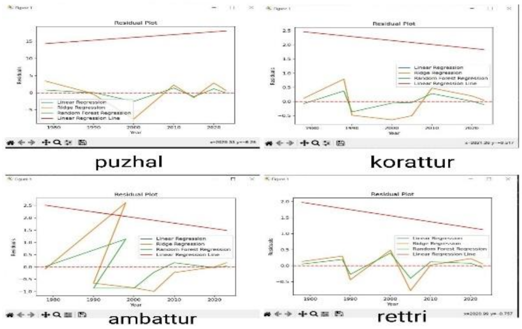


Fig 7.2 Predicted sq. km



7.3 Residual Plots

REFERENCES

[1] Erbek, F.S., Ozkan, C., and Taberner, 2004.

Maximum Likelihood Classification Method vs. Supervised Artificial Neural Network Comparison

Algorithms for land-use activities. International Journal of Remote Sensing, 25(9), 1733-1748.

[2] Y.M. Guo, Y. Liu, A. Oerlemans, S. Y. Lao, S. Y. Wu, and M. S. Lew, 2016. Review of deep learning for visual comprehension.

Neurocomputing,187,pp.27-48. Bovik, A.C., Passalacqua, P., and F. Isikdogan, 2017. Deep learning surface water mapping. Selected Topics in Applied Earth Observations and Remote Sensing, Ieee Journal, 10(11), pp. 17–28. Ji,

[3] L. Y., X. R. Geng, K. Sun, Y. C. Zhao, and P. Gong, 2015. Water Mapping Target Detection Method Using Landsat 8 OLI/TIRS Imagery.

7(2), pp. 794–817 in Water. He, C.Y., Cao, X., Chen, J., Peng, S., Sun, F.D., and Gong, 2014. Liao, A.P., Chen, L.J., Chen, J., global mapping of land water using high-resolution remote sensing. Earth Sciences, Science China, 57(10), pp. 2305-2316. A 2016 study by Xie, H., Luo, X, Xu, X, Pan, H.Y., and Tong, X.H.

[4] OLI Landsat 8 image evaluation for uncontrolled inland water extraction.

37(8), 1826–1844, International Journal of Remote Sensing. Xu, H.Q., 2006. Enhancing open water features in remotely sensed photography by modifying the normalized difference water index (NDWI).

[5] The International Journal of Remote Sensing, 27(14), pages 3025–3033.

Ikeshima, D., Trigg, M.A., and Yamazaki, 2015. creation of a global map of water bodies measuring 90 meters using multiple Landsat time series. Environment and Remote Sensing, 171, 337–351