Machine-learining-BASED RIVERBANK Erosion Prediction USING RIVER Channel observations from Historical Satellite DATA

Tran Thanh Dan1\*, Hiroyuki Miyazaki1, Rohit Khati1, Megharaj Neupane1, Yuasa Takashi2, and Takahito Suzuki2

1GLODAL, Inc., Japan

CrossGate Building 7th Floor, 1-101-1 Sakuragi-cho Nakaku Yokohama-shi Kanagawa 231-0062 Japan

Email: ttdan@glodal-inc.com; miyazaki@glodal-inc.com, rhtkhati@gmail.com, neupanemegh888@gmail.com

2Pacific Consultants Co., Ltd., Japan

3-22, Kanda-Nishikicho, Chiyoda-ku, Tokyo 101-8462, Japan

Email: takashi.yuasa@tk.pacific.co.jp, takahito.suzuki@tk.pacific.co.jp

**KEY WORDS:** riverbank erosion, machine learning, satellite data, regression model, prediction errors

**ABSTRACT:** This study aimed to develop a cutting-edge machine-learning regression model to predict riverbank erosion using historical satellite data. The model utilized various features, such as 90 m slope, topographic positioning index (TPI), angle differences between line segments, and many other features, as independent variables. The dependent variable is the riverbank erosion distance, which is to be predicted for multiple years. For model development, yearly data from 2000 to 2010, with one-year intervals, were selected and the random forest (RF) regression model was chosen to predict the riverbank erosion distance. The performance of the model was evaluated using the R2 metric and RMSE, with a value of 0.89 and 513.61, respectively. In contrast to previous research, which included additional hydrological parameters, our main focus was on satellite-based data for riverbank erosion prediction. While integrating this hydrological information could potentially improve the prediction of erosion, but its incorporation presented challenges in terms of data availability and modelling complexities. To address prediction errors and explore their underlying causes, we plan to acquire hydrological datasets from ground stations and riverbank protection works in future work. By integrating this additional information, we believe that our study will significantly enhance the accuracy of the riverbank erosion prediction model.

# INTRODUCTION

In Bangladesh, riverbank erosion poses a significant problem, particularly along the Jamuna River. This river, one of the world’s largest, continuously encroaches upon the surrounding land, resulting in substantial damage and displacement. Annually, approximately 8,700 hectares of land succumb to erosion, affecting the lives of approximately 200,000 individuals who lose their homes and agricultural plots (Alam, 2017). The issue escalates during the monsoon or rainy season. Furthermore, global climate change has led to rising sea levels, exacerbating erosion in Bangladesh, particularly during the monsoon season when water discharge from upstream to downstream is significantly elevated (Khan et al., 2014). Khan et al. (2014) estimated that approximately 607 km2 of riverbank on both sides of the river eroded between 1973 and 2014. Additionally, Rashid et al. (2021) documented the continuous shift of the river’s thalweg, significantly impacting the livelihoods and socio-economic conditions of the affected population due to extensive land erosion.

Predicting riverbank erosion is a critical component of comprehending and mitigating the dynamic changes in riverine landscapes. Historically, achieving precise predictions has been challenging, necessitating advanced methodologies capable of harnessing the power of remote sensing data and machine learning techniques. In this study, we embark on the development of a cutting-edge machine-learning model designed to predict riverbank erosion using historical satellite data. Our primary objective is to forecast riverbank erosion distances spanning multiple years, providing valuable insights into the evolving dynamics of riverbanks over time.

# METHODOLOGY

The methodology for this R&D project aims to improve the accuracy of the riverbank erosion prediction model by incorporating historical satellite imagery and AI-based analysis. The research will be conducted in several phases, including data collection, data analysis, and prediction model development.

A diagram of a process

Description automatically generated

Figure 1. Workflow of the proposed method for riverbank erosion prediction.

## Study area

The Jamuna River is the second-largest river in the country, flowing through Tibet, China, India, and Bangladesh. The Jamuna River is downstream of the Brahmaputra River, which was formed after a series of earthquakes and floods that occurred between 1782 and 1787. The width of the river varies from 3 km to 20 km with an average width of about 10 km. During the rainy season, the river expands significantly, reaching widths of up to 5 km in some areas. Many channels are submerged during this time, resulting in a single, wider channel. Riverbank erosion is a perennial problem and a major critical hazard in Bangladesh, affecting people living along the rivers. The Jamuna River has a complicated network of sandbars and channels, and shifting riverbank lines due to erosion and depositional activities is a critical issue. A study by Rabbi et al. (2013) found that riverbank erosion has increasingly affected people in the riverine zone of the Jamuna River. The erosion has a significant impact on population change and human settlements along the riverbank, resulting in the displacement of people.

A map of the land

Description automatically generated

Figure 2. The study area for this research, as well as the area of interest, is outlined by the red polygon.

## Data processing

The target of this research is to perform river erosion distance prediction using AI models. To create a machine learning regression model, we need datasets. The training datasets include features (independent variables) and target datasets (dependent variable that is to be predicted). In our case, the riverbank erosion distance is the dependent variable therefore this needs to be calculated for the period of years. In this study, we have selected the years with one-year intervals for the model development. The methodology of calculation of the dependent variable (the erosion distance) is discussed below.

**Calculation of erosion distance**

To calculate erosion distances along the complex braided Jamuna River, several steps were undertaken (Figure 3):

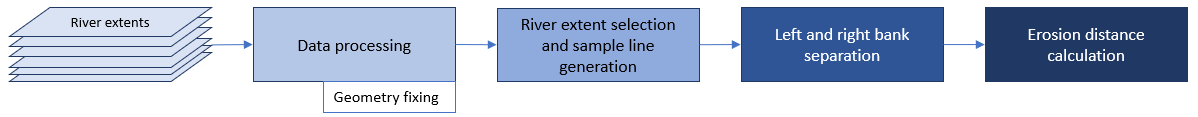


Figure 3. Workflow of erosion distance calculation

Geometry fixing: the first step was to fix these geometries using the Fix Geometries tool in QGIS. This process aimed to create a valid representation of the geometry while retaining all input vertices.

River extent selection: there were a lot of water body extents within and outside of the main river extent. So, we selected the main extent of the Jamuna River. This was done by selecting the polygon with the largest area. Then, we converted this main polygon extent of the Jamuna River into lines. This conversion was achieved by using the boundary function from the geopandas function.

The problem with these lines is that they are pixelated lines not straight lines. While calculating the distance, these lines will produce errors and not give the exact distance between any two lines. Thus, we need to smooth the pixelated lines as shown in Figure 4.

To sample the river extent at equal intervals we exploded the smoothed lines. This algorithm takes a linear layer and creates a new one in which each line is replaced by a set of lines representing the segments in the original line as shown in Figure 5.

|  |  |
| --- | --- |
| Figure 4. Smoothing the pixelated lines. | Figure 5. A segment of the exploded line. |

Left and right bank separation: the river borderlines and river centerlines datasets were joined using the join\_nearest function. The principle guiding this process was based on the x-coordinates of the river borderlines in relation to the centerlines. If the x-coordinate of a river borderline was less than the x-coordinate of the centerline, it was categorized as a left borderline; otherwise, it was classified as a right borderline.

Erosion distance calculation: to calculate the erosion distance two years river boundary data was taken, for example 2010 and 2011. The erosion distance can be positive or negative. It is determined based on the location of x coordinates of a particular sampling line. The condition for positive and negative erosion distance is based on whether the river boundary is left or right.

For the right side:

* If x coordinate of 2010 > x coordinate of 2011

Then distance is negative

For the left side:

* If x coordinate of 2010 < x coordinate of 2011

Then distance is negative.

**Generation of independent variables**

The process for independent variables generation for modeling riverbank erosion is shown in Figure 6. This workflow involves six input parameters:

* DEM 90m
* NIR band
* Red band
* River centerline with a category column indicating river width (this shapefile should undergo correction using the “fix geometries” tool)
* River shape
* Slope map with a 90m resolution

These inputs are provided in both raster and vector formats. For intermediate outputs, all resulting files are in the shapefile format. These intermediate files include slope data at 90m resolution, slope data, aspect data, TPI (Topographic Position Index) data, and azimuth data. The final output is saved in \*. gpkg (GeoPackage) format.

A screenshot of a computer screen

Description automatically generated

Figure 6: Workflow of independent variables generation

## Machine learning models

**Random forest**

Random forest, as its name implies, consists of many individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction (see figure below).

ダイアグラム が含まれている画像

自動的に生成された説明

Figure 7. Concept of Random Forest (Yiu, 2019)

**Artificial neural network**

Artificial neural networks are one of the deep learning algorithms that simulate the workings of neurons in the human brain. Artificial Neural Networks consist of the Input layer, Hidden layers, Output layer. The hidden layer can be more than one in number. Each layer consists of a number of neurons. Each layer will be having an Activation Function associated with each of the neurons. The activation function is the function that is responsible for introducing non-linearity in the relationship.

A diagram of a machine learning

Description automatically generated

Figure 8 ANN neural network (Fernández-Cabán et al., 2018)

**1-dimensional CNN (1D-CNN)**

The convolutional neural network can also be used for regression analysis. In the diagram below it shows how the CNN model works to predict the linear output from the dataset provided.

A close-up of a paper

Description automatically generated

Figure 9. Concept of 1-D CNN model

# EXPERIMENTS AND RESULTS

## Model training and comparative results

We have created a graphical modeler in QGIS to facilitate the automated generation of input training data for model training. In our endeavor to build a robust model, we generated training data spanning the period from 2000 to 2010. To ensure the accuracy of the model and generalizability, we utilized data from 2000 to 2009 for the training set, and 2010 for the testing set. Furthermore, we implemented a procedure to remove data points associated with riverbank protection works by using the construction status and construction year. This process helps us enhance the precision and relevance of our model’s inputs.

As a result of the performance of the models, the RF model did better than the other two. The RF model did well with a high R2 score and a small RMSE.

Table 1. Model comparison using the mid-term dataset, from 2000 to 2010.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model name** | **R2 score** | **MAE** | **RMSE** |
| RF | 0.899 |  | 513.608 |
| ANN | -0.452 |  | 994.653 |
| 1D-CNN |  | 808.538 |  |

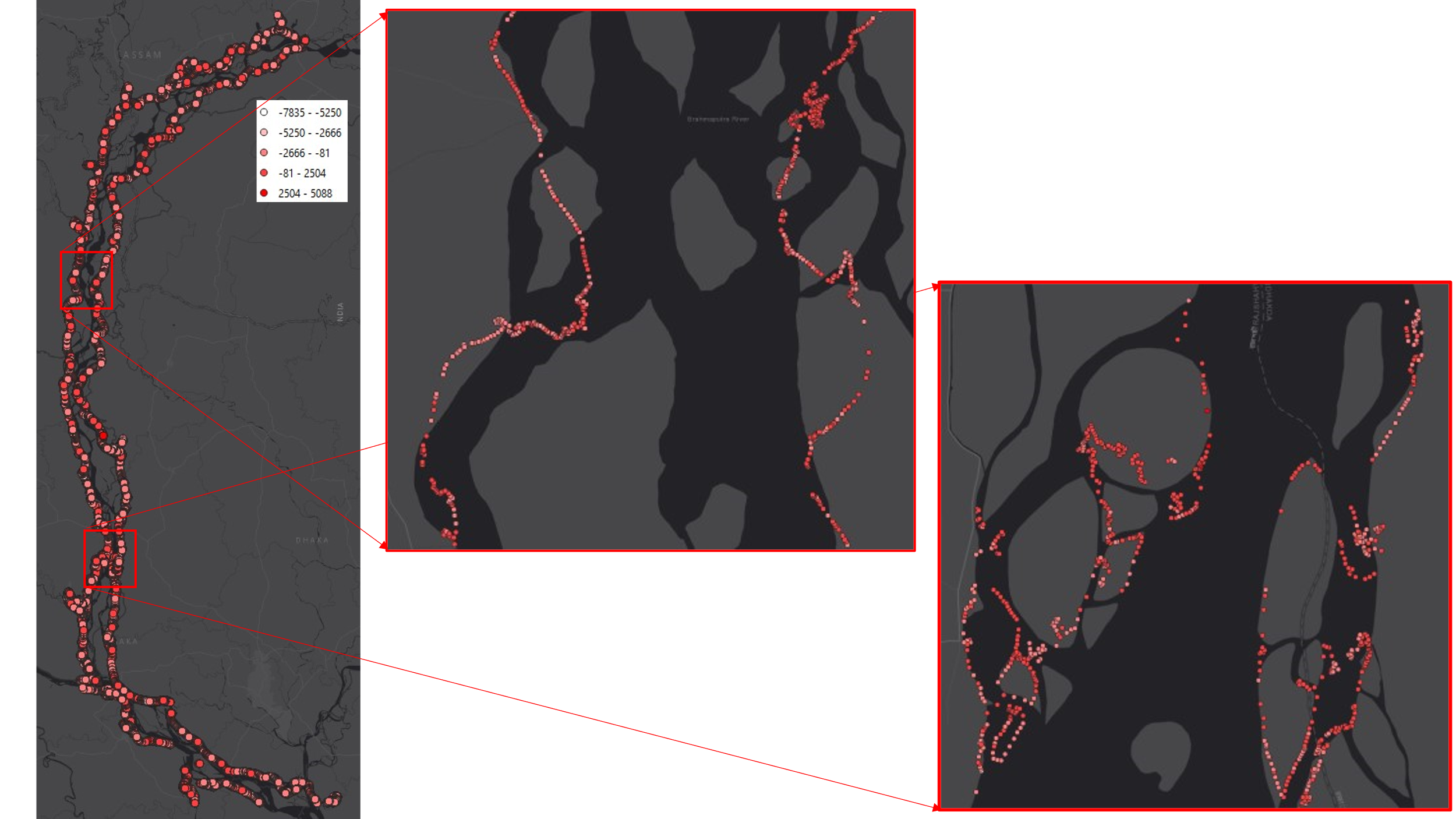


Figure 10. Spatial distribution erosion prediction result in 2010.

We made an intriguing observation: the chosen model frequently underestimates the severity of erosion, indicating suboptimal predictive performance in this regard. Importantly, it is worth noting that the model’s errors are not random; they follow a discernible pattern.

## Limitations of the proposed methodology

We utilized the river extent dataset extracted from Landsat imagery, specifically a composite captured during the dry season. However, this dataset does not include riverbank information. Upon closer examination, we observed an intriguing pattern: erosion tended to occur at greater distances when the river’s size decreased. Notably, this phenomenon was particularly pronounced in the years 2002 and 2008, when the river contracted, and erosion events occurred at more distant locations simultaneously.

And also, we encountered challenges related to the accuracy of erosion distance calculations, particularly when connecting rivers and river tributaries was involved. These factors introduced inconsistencies in the data and led to inaccuracies in the erosion distance calculations.

Moreover, we recognized that riverbank changes throughout the year could introduce errors in erosion calculations. To mitigate this issue, we opted to extract the riverbank by merging quarterly river extent data for the entire year. This approach helped address the limitations associated with dynamic riverbank changes during the year and enhanced the precision of our erosion calculations.

# CONCLUSION AND FUTURE WORKS

In this study, we explored and compared various machine learning models, including Random Forest, ANN regression, and 1D CNN regression models, to predict riverbank erosion using remote sensing data as predictive features. Among these models, the Random Forest regression model consistently demonstrated superior performance, as highlighted in Table 1.

As our primary objective is to develop a satellite-based prediction method for riverbank erosion, we acknowledge that some studies, such as Deng et al. (2022), have successfully incorporated hydrological parameters to enhance predictive accuracy. While our focus remains on utilizing remote sensing data exclusively, we recognize the value of hydrological datasets from sampled surveys and ground stations to investigate the causes of prediction errors further. Additionally, our analysis revealed the availability of a dataset containing information about construction sites, including their location, construction year, destruction year, and current status. This dataset presents an opportunity to analyze prediction errors that may be attributed to man-made constructions, a factor not considered by our model. Future research will involve a comparative analysis of this construction site data and our prediction errors. Furthermore, we acknowledge the potential utility of SAR data for monitoring waterbody shapes, even during the rainy seasons, due to its frequent interval monitoring capabilities. This opens avenues for more comprehensive and timely assessments of riverbank dynamics.

We remain committed to improving predictive accuracy and understanding the complex factors influencing riverbank erosion through continued research and data integration.

**References from Journals**:

Alam, G. M. M. Livelihood Cycle and Vulnerability of Rural Households to Climate Change and Hazards in Bangladesh. Environmental Management 59 (2017): 777–791. <https://doi.org/10.1007/s00267-017-0826-3>.

Bangladesh Government Portal. http://bbs.portal.gov.bd.

Deng, B.; Xiong, K.; Huang, Z.; Jiang, C.; Liu, J.; Luo, W.; Xiang, Y. Monitoring and Predicting Channel Morphology of the Tongtian River, Headwater of the Yangtze River Using Landsat Images and Lightweight Neural Network. Remote Sens. 2022, 14, 3107. https://doi.org/10.3390/rs14133107.

Fernández-Cabán, Pedro, Forrest Masters, and Brian Phillips. Predicting Roof Pressures on a Low-Rise Structure From Freestream Turbulence Using Artificial Neural Networks. Frontiers in Built Environment 4 (2018): 10.3389/fbuil.2018.00068.

Khan, I., M. Ahammad, and S. Sarker. A Study on River Bank Erosion of Jamuna River Using GIS and Remote Sensing Technology. International Journal of Engineering Development and Research 2, no. 4 (2014): 3365-3371.

Rabbi, H., A. S. M. Saifullah, M. Sheikh, S. Sarker, M. H. Md, and A. C. Bhowmick. Recent Study on Riverbank Erosion and Its Impacts on Land Displaced People in Sirajgonj Riverine Area of Bangladesh. World Journal of Applied Environmental Chemistry 2, no. 2 (2013): 36–43. ISSN: 2277-8055. www.environmentaljournals.org.

Rashid, M.B., M.A. Habib, R. Khan, and A.R.M.T. Islam. Land Transform and Its Consequences Due to the Route Change of the Brahmaputra River in Bangladesh. International Journal of River Basin Management 21, no. 1 (2023): 113-125.

Understanding Random Forest. Towards Data Science. https://towardsdatascience.com/understanding-random-forest-58381e0602d2.