ISSN 2321 3361 © 2022 IJESC

Research Article Volume 12 Issue No.5

Extraction of River Network from Satellite Images

Dr. Kakoli Banerjee[1], Dr. Deepti Aggarwal[2], Girish Kumar B C[3], Animesh Tripathi[4], Kratarth Gour[5], MayankMehra[6]
Head of the Department[1], Assistant Professor[2,3], student[4,5,6]
Department of computer science and Engineering
JSS Academy of Technical Education, Noida, India

Abstract:
In recent years, the extraction of water bodies from satellite data has received a lot of attention. Several strategies for demarcating water bodies from distinct satellite imagery with different spatial, spectral, and temporal features have been developed. The extraction method we used was U-net with TensorFlow, which is a deep learning model. The finest delineation of water bodies of various sizes comes through visual interpretation of satellite data, but it takes time, especially when working with high-resolution data. In modern times, the use of satellite images for mapping the natural resources such as forests and water bodies has become more prevalent. Both Forest and water resources have heavy usage, therefore consistent monitoring is crucial for long-term management. The mapping of the water bodies is done in the spatiotemporal domain to examine and evaluate the degree and rate of their deprivation and desertion, which plays a vital role in the global carbon cycle and climatic changes. Geospatial tools are proving to be useful for such effect assessments and conservation measures implementation.

Keywords: Water Body Delineation, Deep Learning, Satellite Imagery, Spatiotemporal Domain

1. Introduction:

Rivers are a vital natural resource that has a huge impact on the worldwide carbon cycle and climate crisis [1]. As a result, rivers are a crucial element of the environment, playing a critical part in human, plant, and animal existence as well as in financial growth of the nation. Water from the rivers is used for a variety of purposes, including farming, drinking, and hydroelectric developments. As a result, it's critical to look after the rivers and detect any regions of contamination in the river. Rivers, on the other hand, have complex geometries with various branches and orientations, making river extraction from satellite data a difficult task. [3]. Because of their capacity to register imageries regardless of climate situations, synthetic aperture radar (SAR) images are increasingly being employed for water monitoring [5]. Optimal thresholding approaches are used in various techniques for extracting water from SAR pictures [8, 13, 11]. These basic procedures, on the other hand, produced a lot of incorrect categorization findings, especially when the water bodies are tiny. For example, in [10], there were five diverse thresholding approaches that were applied to three satellite photos in order to establish the most successful image segmentation strategy.The Mean method, Edge Maximization Technique, P-tile approach, and Histogram Dependent Technique (HDT), are examples of threshold methods (EMT). According to Kalyankar and Al-Amri [10], HDT and EMT procedures produced great outcomes. HDT is an approach based on histogram, in which the histogram is dependent on the value of the threshold.
The threshold value is used to identify the area of the image where the feature of interest blends into the background from the remainder of the image. The EMT approach is applied when there is a significant difference in illumination between the feature of interest and the background. Gong et al. [30] presented a fuzzy c-means algorithm and the Otsu threshold method-based automatic adaptive threshold segmentation method for extracting tidal streams.
As previously said, research on water management has gained traction as ecological concerns, such as increasing populations and degradation of water resources, have grown. Water is a vital life-sustaining resource, so protection of natural water supplies should be a high priority task [4]. In this context, water bodies extraction and detection from satellite data is useful for a variety of design and evolution activities, including coastal mapping, river erosion mapping, and water resource management. Scientific study in the field of water resources management and planning has looked at a variety of satellite data to better understand the geographical, spectral, and temporal properties of land, with a focus onwater bodies [2]. The use of satellite remote sensing photos was not the sole method for detecting water resources; an Unmanned Aerial Vehicle (UAV) was also used [24, 9]. For example, "on the GeoEye-1 satellite image and mosaic image obtained by a drone on the coast of San Vito Lo Capo, Randazzo et al. [7] developed an image processing framework for coastal extraction as well as shallow water depth." Several academics have used numerous machine learning technologies and high-resolution satellite pictures to examine water bodies [31, 35, 36].
Given the significance of water bodies extraction from satellite images, the main purpose of this paper is to derive a method for the automatic extraction of water bodies from Landsat satellite images, which comprises of multiple image processing phases. The phases include image enhancement, image augmentation, model creation and testing.

2. Dataset:

The dataset used for this research is the collection of images of water bodies captured by the Sentinel-2 satellite uploaded on Kaggle. Each image is accompanied by a black and white mask, where white represents water and black represents something other than water. Masks created by computing the NWDI (Normalized Water Difference Index) are routinely employed to detect and measure vegetation in satellite pictures, although water bodies are detected using higher thresholds. There were two different directories containing 2269 images each.

3. Related Work:

Several papers have been published that describe several approaches for segmenting features in satellite images. Several artefacts have been effectively discovered and removed, despite the fact that rivers have not been extensively studied. Dhanachandra et al. [19] successfully recovered rivers from satellite pictures and the two approaches were investigated. The first involved employing a color histogram methodology to extract rivers, as well as a "hill-climbing algorithm and the k-means clustering method". The second involved the usage of grayscale picture thresholding and morphological degradation. Based on both methodologies, the results were reasonable. During the extraction of landscapes from low-resolution satellite images, Syrris et al. [28] emphasized the necessity of picture augmentation and contrast correction. The authors created a case study that shows how enhancing techniques like linear and decorrelation stretching can be used. They showed how low-resolution satellite photos might bechanged or "fixed" to improve feature extraction.

To identify the surface water in Landsat 8 satellite photos, Jiang et al. [25] used a multilayer Perceptron Artificial Neural Network (ANN). To extract lakes from satellite data, a study described in [20] combined ANN and a threshold method with a set of mathematical morphological operations. Water bodies were extracted from high-resolution remote sensing pictures obtained from Google earth imaging using a segmentation approach based on Convolutional Neural Networks (CNN) [18]. [15] used a deep learning approach to interactively extract water bodies, farmlands, woods, and other non-artificial features from high-resolution satellite photos. Their method was tested on two datasets with a variety of item types and complex situations [15].

"An automated Lake and Reservoir Extraction Process" was presented by Meng et al. [25]. (LREP). This approach used 154 Gaofen-2 photos from Zigong, Xianning, and Liaoyuan to build a "Modified Two-Mode (MTM) method" for detectng water bodies from distorted imageries. Remote sensing photos were utilized for detecting water bodies with the help of an upgraded deep convolutional encoder-decoder network that was published in [20] through conditional random fields and superpixel segmentation.

Rishikeshan and Ramesh [21] proposed "a flexible Mathematical Morphological (MM)-driven approach" for recovering features of water bodies from a variety of satellite data with varying spatial resolutions. In China, a large no of lakes was extracted from the Landsat 8 land images using an operational water body extraction approach. For attaining an automatic dynamic threshold, this approach looks for the least threshold range values instead of requiring to calculate the histogram peaks [8].

Furthermore, for the identification of pollution in river water, a range of machine learning and image processing algorithms have been used. [33] conducted an analysis of many satellite sensors, image processing technologies, and numerous classification approaches for classifying river water. Navarro et al. [22] proposed an unsupervised method for extracting inland water bodies that incorporated the "local Moran index of spatial association with morphological processes". Dereli and Tercan [23] devised an approach for detecting changes in shoreline of Turkey's Lake Salda from 1975 to 2019. From 2008 to 2016, a segmentation method based on the stepwise thresholding methodology was used on SAR pictures in Ontario, Canada for mapping and tracking the changes in surface water area [13].

In [24], a method for recognizing water from Landsat photos was reported that used a multiscale CNN and Google Earth Engine (GEE). [26] suggested a method for extracting rivers from SAR imagery that incorporated global salience characteristics, and a "multi-feature fusion method based on principal component analysis, and an Active Contour Model (ACM)". The accuracy of a novel automated water detection and picture fusion method was proposed in [27] that was used to detect the actual changes in water bodies on a larger scale. "Automatic water extraction index, GEE platform, and Landsat 8 OLI data of New Zealand for the years 2014, 2015, 2016, 2017 and 2018" was combined to evaluate the accuracy of this novel water body identification technique.

4. Methodology:

The project methodology is divided into the following steps:

A. Data Collection
B. Data Exploration
C. Data Augmentation
D. Model Creation
E. Model Testing

A. Data Collection:

The Sentinel-2 satellite collected a collection of photos of aquatic bodies. Each image includes a black and white mask, with white representing water and black representing anything other than water. The masks were created by calculating the NWDI (Normalized Water Difference Index), which is commonly used in satellite photos to detect and measure vegetation, but a higher threshold was employed to detect water bodies. Two different directories contain 2269 images each where one represents original satellite images and the other masks generated images.

B. Data Exploration:

Data exploration is the processing of exploring the data in order to analyze its accuracy. This step helps us in understandingthe things that are needed to be done in order to get the accurate results. In data exploration, we choose some images randomly from the dataset and analyze the original and their masked images. Since we need some mathematical backing to make surewhether we can rely on the dataset we plotted a graph to measure the quality of the dataset.

Below you can see the result of the images.

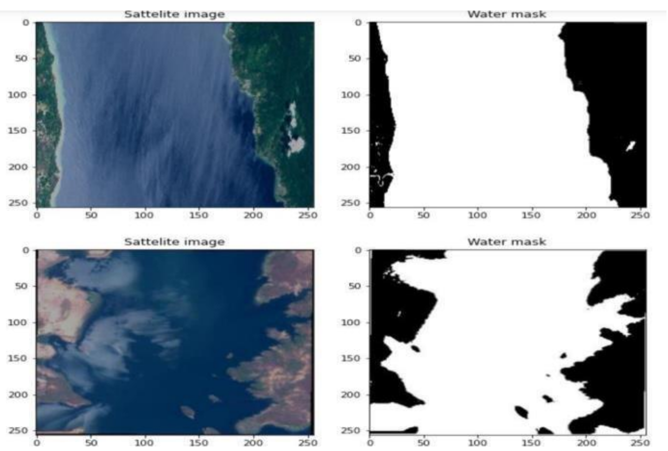

Fig. 1 Visualization of satellite images and its masked region

The masked images are quite accurate but we cannot depend upon them to make sure whether we can rely on the data or not so we will measure the data quality.

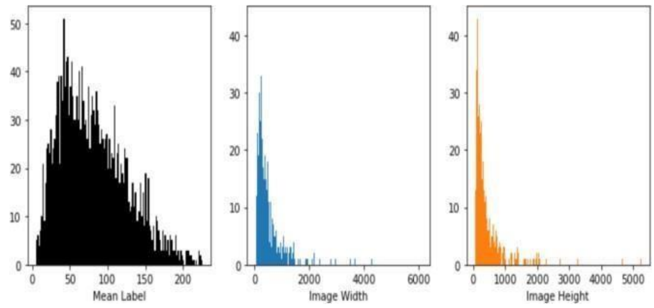

Fig 2. Graphical representation of the data quality

By looking at the graph we can conclude that there are no completely black or white images present in the masked dataset, hence there is no need for us to perform filtering. These images seem to be labeled right in many cases. Hence, we are good with the dataset.

C. Data Augmentation

The quantity and diversity of training data determine the success of most machine learning models and deep learning models in particular. However, one of the most prevalent problems in applying machine learning in the organization is a lack of data. This is due to the fact that gathering such information can be costly and time-consuming in many circumstances.

Here in our case, we have just 2269 masked and 2269 satellite image datasets. As we are using a Deep learning algorithm to train our model, it is insufficient. Hence, we are using the data augmentation technique to increase the number of datasets for our project. Let's start with the definition:

“Data augmentation is a set of techniques for producing additional data points from current data in order to artificially increase the amount of data available”. Examples of this are making diffident adjustments to the original data or employing deep learning models in order to produce fresh data points.

1) Why is it important now?

Machine learning applications are quickly broadening and increasing, particularly in the deep learning arena. Techniques for data augmentation could be beneficial in addressing the problems that the artificial intelligence industry is facing. The results and performance of deep learning models can be improved by training the datasets on new and varied cases which are produced by data augmentation. A deep learning model performs more accurately when the dataset is sufficiently large.

2) How is it done?

Making basic changes to visual data is popular for data augmentation. In addition, new synthetic data is created using generative adversarial networks (GANs). The following are examples of traditional image processing activities for data augmentation:

1. Padding
2. Random Rotating
3. Re-scaling,
4. Vertical and Horizontal flipping
5. Translation (image is moved along X, Y direction)
6. Cropping
7. Zooming
8. Darkening & Brightening/Color modification
9. Gray Scaling
10. Changing Contrast
11. Adding Noise
12. Random Erasing
13. How is it done in our project?

3. Proposed Model

For our project, we are using rotation techniques to increase the size of the dataset. Using this method, the dataset can be increased from 2x to 4x. We are using rotation angles of 90, 180, 270. Hence the dataset will be increased by 4 times.

```
# import the Python Image
from PIL import Image

for f in filenames:
    # rotating & saving images
    s_image=Image.open(images_dir+'/'+f)
    (s_image.rotate(90)).save('dataset/in/'+'90'+f)
    (s_image.rotate(180)).save('dataset/in/'+'180'+f)
    (s_image.rotate(270)).save('dataset/in/'+'270'+f)

    #rotating masked images
    m_image=Image.open(masks_dir+f)
    (s_image.rotate(90)).save('dataset/out/'+'90'+f)
    (s_image.rotate(180)).save('dataset/out/'+'180'+f)
    (s_image.rotate(270)).save('dataset/out/'+'270'+f)

dirname, _, filenames = next(os.walk(images_dir))
len(filenames)*2

18152
```

Fig. 3 Augmentation Process

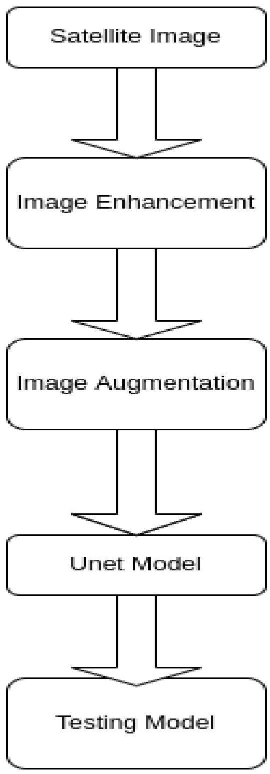

Fig. 4 Model Flowchart

D. Model Creation

As the number of datasets becomes 18.2k and the model is totally based on CNN. Therefore, for the better accuracy of the model, a complex model is need to be created.

In the training model, we have divided it into three different parts: Encoder, Centre, Decoder. The image segmentation tasks can be very well performed using the U-net model. The implementation has a lower number of channels due to memory restrictions. It also may be simple enough for binary classification.

Below can be seen the layers and the parameter of the U-net model.

```
model.summary()

Model: "model_2"

 Layer (type)                   Output Shape          Param #     Connected to
==================================================================================================
 input_3 (InputLayer)           [(None, None, None,   0           []
                                 3)]

 conv2d_36 (Conv2D)             (None, None, None,    896         ['input_3[0][0]']
                                32)

 conv2d_37 (Conv2D)             (None, None, None,    9248        ['conv2d_36[0][0]']
                                32)

 max_pooling2d_6 (MaxPooling2D)  (None, None, None,    0          ['conv2d_37[0][0]']
                                32)

 conv2d_38 (Conv2D)             (None, None, None,    18496       ['max_pooling2d_6[0][0]']
                                64)

 conv2d_39 (Conv2D)             (None, None, None,    36928       ['conv2d_38[0][0]']
                                64)

 max_pooling2d_7 (MaxPooling2D)  (None, None, None,    0          ['conv2d_39[0][0]']
                                64)

 conv2d_40 (Conv2D)             (None, None, None,    73856       ['max_pooling2d_7[0][0]']
                                128)

 conv2d_41 (Conv2D)             (None, None, None,    147584      ['conv2d_40[0][0]']
                                128)

 max_pooling2d_8 (MaxPooling2D)  (None, None, None,    0          ['conv2d_41[0][0]']
                                128)

 conv2d_42 (Conv2D)             (None, None, None,    295168      ['max_pooling2d_8[0][0]']
                                256)

 conv2d_43 (Conv2D)             (None, None, None,    590080      ['conv2d_42[0][0]']
```

Fig. 4 U-net Layers

```
==================================
Total params: 1,925,601
Trainable params: 1,925,601
Non-trainable params: 0
```

Fig. 5 U-net Parameters

E. Model Testing

The Model was tested on a dataset of images taken fromthe Sentinel-2 satellite.

Below can be seen some random pictures of satellite images whose water network have been predicted.

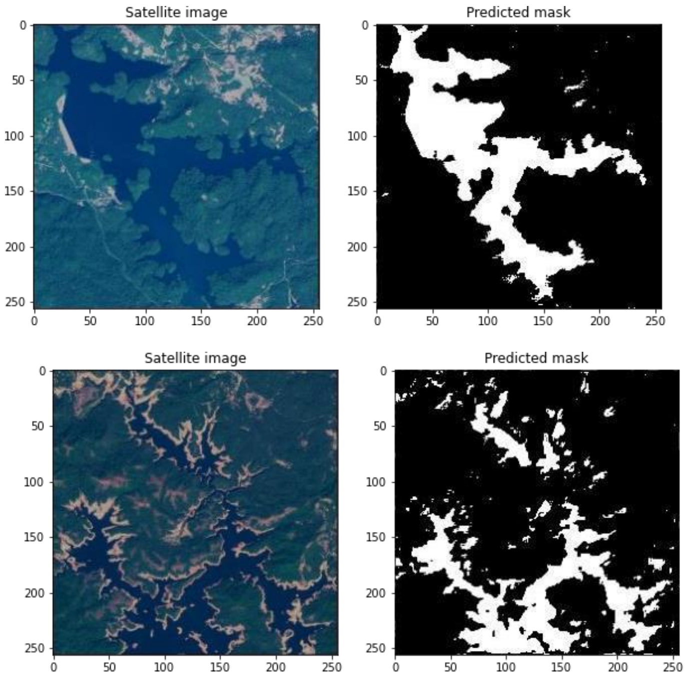

Fig. 6 Satellite Images and their Predicted Mask

5. Accuracy:

In this paper, we presented our preliminary model for detecting rivers and watersheds from satellite images, which was based on image processing techniques. A collection of photos from the Sentinel-2 satellite were used to evaluate the methodology. A new level of segmentation was used to better recognise rivers and watersheds. With our suggested model, we were able to achieve a high level of accuracy, which is much greater than existing U-net and TensorFlow-based models currently available.

6. Conclusion and Future Scope:

In this paper, we presented our proposal for extracting river networks from satellite photos using U-net and TensorFlow deep learning models along with image processing methodology in this publication. A dataset of photos from Kaggle was used to test the methodology. This research could pave the way for more research into developing water resource management, which is crucial for generations to come. The volume and quality of water bodies have been seen to be rapidly declining. Presently, the government must prepare for protecting the quality of water bodies in order to prevent further deterioration. Apart from simply accumulating, the government need secure central warehouses for preserving compressed data about water body indexes [32, 12, 34]. These warehouses can come handy while testing various machine- learning algorithms for not only extracting river networks but also for forecasting contamination and diseases caused due tothat [29, 14].

7. References

[1] Stagl J, Mayr E, Koch H, Hattermann FF, Huang S. Effects of climate change on the hydrological cycle in central and eastern Europe. In managing protected areas in central and eastern Europe under climate change 2014 (pp. 31-43). Springer, Dordrecht

[2] Zhaohui Z, Prinet V, Songde MA. Water body extraction from multi-source satellite images. In IGARSS 2003. International geoscience and remote sensing symposium. proceedings (IEEE Cat. No. 03CH37477) 2003 (pp. 3970-2). IEEE.

[3] Rizvi IA, Mohan BK, Bhatia PR. Automatic object extraction using object-based image classification technique from high resolution remotely sensed images. In proceedings of the international conference and workshop on emerging trends in technology 2010 (pp. 623-8).

[4] Shah V, Choudhary A, Tewari K. River extraction from satellite image. International Journal of Computer Science Issues. 2011; 8(4):386-91.

[5] Hakdaoui S, Emran A. Extraction of water information based on sar radar and optical image processing: case of flood disaster in southern Morocco. In geospatial technology 2020 (pp. 15-29). Springer, Cham.

[6] Wang X, Xie H. A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management.

[7] Randazzo G, Barreca G, Cascio M, Crupi A, Fontana M, Gregorio F, et al. Analysis of very high spatial resolution images for automatic shoreline extraction and satellite-derived bathymetry mapping. Geosciences. 2020; 10(5):172.

[8] Zhang F, Li J, Zhang B, Shen Q, Ye H, Wang S, Lu Z. A simple automated dynamic threshold extraction method

for the classification of large water bodies from landsat-8 OLI water index images. International Journal of Remote Sensing. 2018; 39(11):3429-51.
[9] Tymków P, Jóźków G, Walicka A, Karpina M, Borkowski A. Identification of water body extent based on remote sensing data collected with unmanned aerial vehicle. Water. 2019; 11(2).
[10] Al-Amri SS, Kalyankar NV. Image segmentation by using threshold techniques. arXiv preprint arXiv:1005.4020. 2010.
[11] Zhang W, Hu B, Brown GS. Automatic surface water mapping using polarimetric SAR data for long-term change detection. Water. 2020; 12(3).
[12] Banerjee, K., & Prasad, R. A. (2014, October). A new technique in reference-based DNA sequence compression algorithm: Enabling partial decompression. In AIP Conference Proceedings (Vol. 1618, No. 1, pp. 799-802). American Institute of Physics.
[13] Hu S, Qin J, Ren J, Zhao H, Ren J, Hong H. Automatic extraction of water inundation areas using sentinel-1 data for large plain areas. Remote Sensing. 2020; 12(2).
[14] Sharma, T., Banerjee, K., Mathur, S., & Bali, V. (2020). Stress analysis using machine learning techniques. International Journal of Advanced Science and Technology, 29(3), 14654-14665.
[15] Li K, Hu X, Jiang H, Shu Z, Zhang M. Attention guided multi- scale segmentation neural network for interactive extraction of region objects from high resolution satellite imagery. Remote Sensing.2020;12(5):789.
[16] Meng L, Zhang Z, Zhang W, Ye J, Wu C, Chen D, et al. An automatic extraction method for lakes and reservoirs using satellite images. IEEE Access. 2019; 7:62443-56.
[17] Feng W, Sui H, Huang W, Xu C, An K. Water body extraction from very high-resolution remote sensing imagery using deep U- Net and a superpixel-based conditional random field model. IEEE Geoscience and Remote Sensing Letters. 2018; 16(4):618-22.
[18] Miao Z, Fu K, Sun H, Sun X, Yan M. Automatic water-body segmentation from high-resolution satellite images via deep networks. IEEE Geoscience and Remote Sensing Letters. 2018; 15(4):602-6.
[19] Dhanachandra N, Manglem K, Chanu YJ. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science. 2015; 54:764-71.
[20] Rishikeshan CA, Ramesh H. An ANN supported mathematical morphology based algorithm for lakes extraction from satellite images. ISH Journal of Hydraulic Engineering. 2018; 24(2):222-9.
[21] Rishikeshan CA, Ramesh H. An automated mathematical morphology driven algorithm for water body extraction from remotely sensed images. ISPRS Journal of Photogrammetry and Remote Sensing. 2018; 146:11-21.
[22] Navarro JC, Salazar-Garibay A, Téllez-Quiñones A, Orozco-del- Castillo M, López-Caloca AA. Inland water body extraction in complex reliefs from Sentinel-1 satellite data. Journal of Applied Remote Sensing. 2019; 13(1).
[23] Dereli MA, Tercan E. Assessment of shoreline changes using historical satellite images and geospatial analysis along the lake salda in turkey. Earth Science Informatics. 2020:1-10.
[24] Wang Y, Li Z, Zeng C, Xia GS, Shen H. An urban water extraction method combining deep learning and google earth engine. Journal of selected topics in applied earth observations and remote sensing. 2020; 13:768-81.
[25] Jiang W, He G, Long T, Ni Y, Liu H, Peng Y, Lv K, Wang G. Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. Remote Sensing. 2018; 10(5).
[26] Liu Y, Zhang P, He Y, Peng Z. River detection based on feature fusion from synthetic aperture radar images. Journal of Applied Remote Sensing. 2020; 14(1):016505.
[27] Nguyen UN, Pham LT, Dang TD. An automatic water detection approach using Landsat 8 OLI and google earth engine cloud computing to map lakes and reservoirs in New Zealand. Environmental monitoring and assessment. 2019; 191(4):235.
[28] Syrris V, Ferri S, Ehrlich D, Pesaresi M. Image enhancement and feature extraction based on low resolution satellite data. Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015; 8(5):1986-95.
[29] Yadav, N., Banerjee, K., & Bali, V. (2020). A survey on fatigue detection of workers using machine learning. International Journal of E-Health and Medical Communications (IJEHMC), 11(3), 1-8.
[30] Gong Z, Wang Q, Guan H, Zhou D, Zhang L, Jing R, et al. Extracting tidal creek features in a heterogeneous background using Sentinel-2 imagery: a case study in the Yellow River Delta, China. International Journal of Remote Sensing. 2020; 41(10):3653-76.
[31] Banerjee, K., Kumar, M. S., & Tilak, L. N. (2021). Delineation of potential groundwater zones using Analytical hierarchy process (AHP) for Gautham Buddh Nagar District, Uttar Pradesh, India. Materials Today: Proceedings, 44, 4976-4983.
[32] K. Banerjee and R. A. Prasad, "Reference based inter chromosomal similarity-based DNA sequence compression algorithm," 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 234-238, doi: 10.1109/CCAA.2017.8229806.
[33] Patel HJ, Dabhi VK, Prajapati HB. River water pollution analysis using high resolution satellite images: a survey. In 5th international conference on advanced computing & communication systems (ICACCS) 2019 (pp. 520-5). IEEE.
[34] Banerjee, K., & Bali, V. (2020). Design and Development of Bioinformatics Feature Based DNA Sequence Data Compression Algorithm. EAI Endorsed Trans. Pervasive Health Technol., 5(20), e5.
[35] Banerjee K., Santhosh Kumar M.B., Tilak L.N., Vashistha S. (2021) Analysis of Groundwater Quality Using GIS-Based Water Quality Index in Noida, Gautam Buddh Nagar, Uttar Pradesh (UP), India. In: Choudhary A., Agrawal A.P., Logeswaran R., Unhelkar B. (eds) Applications of Artificial Intelligence and Machine Learning. Lecture Notes in Electrical Engineering, vol 778. Springer, Singapore. https://doi.org/10.1007/978-981-16-3067-5_14
[36] Banerjee K, Bali V, Nawaz N, Bali S, Mathur S, Mishra RK, Rani S. A Machine-Learning Approach for Prediction of Water Contamination Using Latitude, Longitude, and Elevation. Water. 2022; 14(5):728. https://doi.org/10.3390/w14050728.