International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Water Quality using Deep Learning with VGG19 Approach

Donthula Mamatha[1,] Dr. Halavath Balaji[2] and Dr. Sreedhar Bhukya[3]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: Water quality evaluation is essential to environmental management and monitoring. Traditional methods often rely on manual inspection and It may be expensive and taking time. Deep learning has become a potent method for picture categorization in recent years. Iinvolving an evaluation of the water purity. This research VGG 19 a Convolutional water purity.Our suggestion is a channel-specifications attention gate integrates a channel-wise attention gate to focus on relevant features in the input images. Our approach includes preprocessing steps such as background elimination, removal of non-essential features, image enhancement, and noise removal to improve classification accuracy. We experimented with a collection of water surface achieving an accuracy of 97.5%, out performing previous approaches. The results demonstrate the effectiveness of deep learning models in water quality classification and highlight the importance of preprocessing techniques in improving classification performance.

Keywords: *Convolution Neural Network, VGG19, Deep Learning*

1. Introduction

Water quality is a critical factor in maintaining environmental health as well as guaranteeing access to pure, safe drinking water. Traditional methods of water quality assessment often involve time-consuming and expensive laboratory tests. However, recent advancements deep learning, especially with regard to computer vision have shown promise automating processs of water quality assessment through image analysiss.

Deep learning models were effectively implemented, including convolution neural networks (CNNs).various images classification job.In this study, we focus on the application of the VGG 19 architecture, a widely used CNN, for water quality classification.

We propose a hierarchical attention model that incorporates a channel-wise attention gate the models capacity to concentrate on pertinent features in water surface images. Pre-processing contributes significantly to the deep learning models' ability to classify water quality. Techniques such as background elimination, removal of non-essential features, image enhancement, and noise removal are essential for improving the quality of input data and enhancing the model's performance. In this paper, we present our approach to water quality classification using deep learning and VGG 19. We describe the preprocessing steps and the architecture of our hierarchical attention model. We also give the outcome of experiments conducted on a collection of water surface images, demonstrating the effectiveness of our approach in achieving high classification accuracy.

[1]Student of M.Tech, Department of Computer Science and Engineering ,
[2]Professors, Department of Computer Science and Engineering ,
[3]Sreenidhi Institute of Science and Technology, Hyderabad,Telangana, India.
Email: mamathait76@gmail.com, halavathbalaji@sreenidhi.edu.in , sreedharb@sreenidhi.edu.in

2. Literature Review

Deep Learning-Integrated Quality of Water Monitor and Event Identification in the Internet of Things [1] paper provides an extensive analysis of the use of deep learning methods IOT founded water quality monitoring actual water quality parameter classification is achieved by the authors' proposed system, which combines recurrent and Convolutional neural networks (RNNs)The goal for the research is to create an integrated Internet of Things platform for the evaluation of water quality, with a focus on the possibility for early event detection and prompt reactions to episodes of water contamination.

Deep Neural Networks for Water Quality Monitoring [2] this paper discussed on deep neural network applications for water quality monitoring is examined in this research. In order to forecast water quality metrics using past data, the authors suggest using a deep learning algorithm. To increase forecast precision and evaluate accuracy for water quality, the study shows how deep learning can capture intricate patterns in time series data related to water quality.

A Meta-Analysis and Survey of Deep Learning in Remote Sensing [3] this research offers a Deep learning meta-analysisapplications about remote sensingwhich is highly important to water quality assessment even if it is not specifically focused on water quality. It examines several deep learning models and how well they work in remote sensing applications, providing insights into the possibilities of combining deep learning with satellite and remote sensing data for the monitoring and categorization of water quality.

Water Quality Monitoring Using Remote Sensing and Machine Learning: A Review [4] the application of deep learning and remote sensing to water quality monitoring is covered in this review paper. An overview of the most recent methods for assessing water quality using data from satellites and aircraft is given. In order to categorize water quality and identify pollution events, the article investigates the possibility of integrating remote sensing data with cutting-edge machine learning algorithms.

An Approach to Automatic Water Quality Assessment Using Deep Learning [5,6] this work proposes a deep learning technique for

automated evaluation of water quality. To classify the quality of water samples based on pictures of bodies of water, the authors employ a Convolutional neural network (CNN).[7-10] In this regard, the study highlights the benefits of deep learning and it potential for real-time, image-based categorization of water quality [11-12].

3. Materials and Methods

Here is a sample materials and methods section for a water quality classification using deep learning study.

Materials

3.1. 1.Data Collection:

In this paper we use water images as dataset and collected from Google online images which have three categories of five categories. Pixel values form images are taken as input and labels are used as output and each folder has 50 images. Which are used for training and images are standardized to a fixed image size 224 × 224 pixels

3.1.2. Data pre-processing:

Background elimination: Removed backgrounds to focus on the water surface.

Elimination of non-essential objects: Removed non-essential objects like boats, people, or animals from the images.

Image enhancement: Enhanced the contrast and brightness of the images to improve visibility.

Noise removal: Applied noise reduction techniques to improve image quality

3.1.3. Data splitting:

The stages that involve Using photos that are captured, the following steps are involved in categorizing the water quality: These are: Dataset training, validation,testing the collection of data, or the gathering of specimens of water from various locations and water surface photographs, is the initial step in the project. Google earth photos and smart phone photos may be among those. The project model is trained using the gathered data and the water sample data. In the experiment's data pre-processing section, gathering data and dataset training are completed. A well-trained dataset is necessary to achieve excellent accuracy. The dataset's validation establishes the correctness of the provided water images samples, which were gathered from various locations and classified into two folders'he trained dataset, which also has two subfolders with photographs of safe and harmful water, is compared to the two validation dataset folders.

3.2. CNN Algorithm

In computer vision, For deep learning, convolution neural networks (CNNs) are a popular choice. The field of artificial intelligence, referred to as "computer vision," enables computers to interpret and process visual information such as photo graphs. Popular kind of deep neural network for interpreting visual images is a Convolutional neural network, or CNN. It is made to recognize the geographical component hierarchy from input photos naturally and flexible.

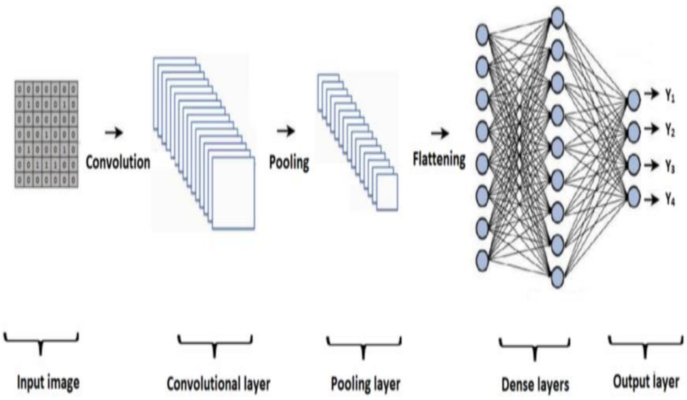

Fig 1: CNN Architecture

Here's an explanation of the key components of a typical CNN architecture:

3.2.1. Input Layer:

After accepting the input image(s), that layer sends them to the following layer. Usually, a matrix of dimensions is used to display the input picture, with every dimension denoting a different feature in that picture. (e.g., width, height, color channels). An input image of water surface, typically with dimensions such as 224 x 224 pixels and 3 color channels (RGB)

3.2.2. Convolutional Layer:

Applying an array of filters, also called kernels, to a given picture or image is what the Convolutional layer does. A few features, including edges, surfaces, or forms, are extracted from an input by every filter. The filter used moves across the provided image(s) via the convolution process, computing the product of dots among its filter weights and the correct input pixel numbers. For every filter, this procedure creates a feature map that captures several faces of the incoming data. To extract pattern space from the input image, apply many Convolutional layers with tiny filter sizes (e.g., 3x3). These layers may be stacked to expand the neural system and enable is to pick up additional aspects.

3.2.3. Activation Function:

Following the steps of convolution, feature maps are subjected element-by-element to an activation function (e.g., ReLU, or Fixed Linear Unit). By adding non-linearity, the model can recognize complex structures in the information. Following every Convolutional layer, add irregularities within the structure by applying an irregular activation function, such as ReLU.

3.2.4. Pooling Layer: The pooling layer down samples generates maps of features using the convolution layer, which are down sampled via the layer for pooling. In doing so, essential data is preserved while the geographic dimensions of the map of features are decreased. Typical and maximum pooling are two popular pooling procedures.

3.2.5. Dense layer: After multiple layers of convolution and pooling, the neural network's excellent processing is carried out using fully linked layers. Totally linked neurons may recognize complex structures in information since they are coupled to every activation from the level above them.

3.2.6. Output Layer: The CNN's ultimate output is generated by the output layer. The job at hand determines how many neurons belong to the output layer. The final layer usually outputs probable outcomes for every class using a soft max activation function for classification tasks. In conclusion, Convolutional and pooling layers in CNNs are made to take characteristics out of the input pictures: layers that are completely linked are then added for deeper reasoning. CNNs may accomplish state-of-the-art outcomes in several computer vision applications object recognition, picture grouping and image categorization, thanks to their design.

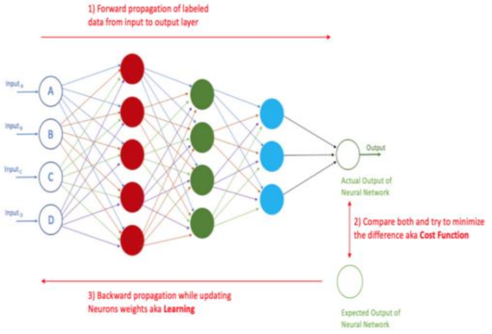

Fig 2: Topology of a feed-forward model with two hidden layers

In order to establish the ideal weights and activate just the most potent and categorize neurons for categorization, forward and backward propagation iteratively go through all of the training data in the network.

3.3.1. Epochs:

One pass constitutes an epoch. The entire dataset through the neural network, during training, the model goes through many epochs to learn from the data.

3.3.2. Forward Propagation:

In each epoch, forward propagation calculates the predicted output of the network for each input sample. The loss function compares these predictions to the actual labels, measuring the discrepancy between them.

3.3.3. Backward Propagation:

After forward propagation, backward propagation determines the range of the reduction function for the weight. An optimization method updates the weights based on this gradient-like descent. Back propagation helps the network learn from its mistakes and improve its predictions.

3.3.4. Optimization:

The optimization process adjusts the weights and biases of the networkto reduce the loss function as much as possible. The aforementioned procedure is iterated several times until the model settles on an ideal combination of weights and biases.

3.3.5. Learning rate:

As the model sees more examples and goes through more epochs, it learns to make better predictions, leading to a decrease in the loss measure.

3.3.6. Cost Function:

The cost function (or loss function) calculates the average loss across all training samples. It offers an indicator of the general performance of the vehicle. Overall, this iterative process of forward and backward propagation, combined with optimization, permits continuous performance improvement of the neural network by learning from information.

Sr. No.	Hyperparameters	Value
1.	Learning Rate	0.001
2.	Batch Size	32
3.	Epochs	20
4.	Activation Function	ReLu
5.	Optimizer	Adam
6.	Metric	Accuracy
7.	Loss Function	SparseCategoricalFocalLoss(gamma=2) SparseCategoricalCrossentropy

Table1: Shown the training samples data

3.4. VGG19 MODEL

Every training sample's mean loss is determined by the cost function, also called the loss function. It provides an indication of the vehicle's reliability. In general, through learning from data, the neural network's performance may be continuously improved through the recurrent forward and backward propagation process and optimization.Eight - step setup for my top-performing VGG19 model is shown below. Equipped with the already-trained elements and an acute understanding of the form, color, and pattern that make up a picture, VGG19is a sophisticated CNN with training on millions of different pictures and challenging classification problems, VGG19 is incredibly deep. I just froze VGG19'sthe layers, and finally built a simple two-layer network on top of it to accomplish my categorization objective of identifying among images using and without branches. I did train VGG19 any further.

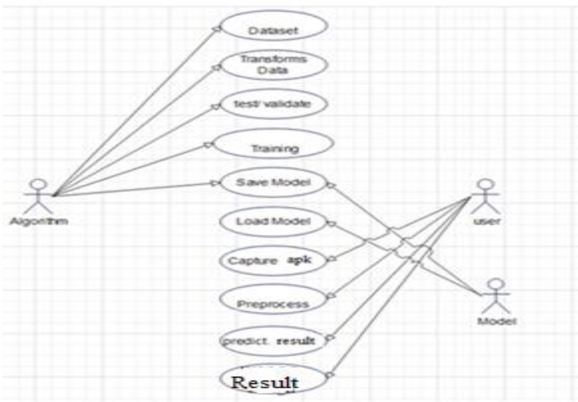

Fig 3: Shown use case diagram

1. Load your model.
2. Input the size of your data set. The images in this instance have already been manually divided into those with (target = 1) and without trees (target = 0).and randomly tossed into various folders for training, testing, and validation.
3. Create a function that extracts and freezes the VGG-19's first layers, which are responsible for processing the features and labels of underlying pictures. The model will be able to recall its pre-training from millions of online photographs thanks to transfer learning.
4. Use this function to ensure that the features and labels are extracted from your training, validation, and test datasets
5. Ensure that the data accurately reflects the sizes of your datasets.
6. To determine which binary category each picture belongs to in your final classifying layer, save the extracted features and labels into a folder called "bottlenecked."
7. To build the final classification layer on top of the VGG-19 "brain" (the feature extractor), you need to load the extracted features and labels from the 'bottlenecked' folder and define and train a new classification layer. Here's how you can do it: Load Extracted Features and Labels:
8. Print your training history to see your model's learning performance using your accuracy and loss measures. It is evident that our accuracy increased and our loss dropped with each epoch, or iteration.

Adding adversarial examples and variations to your dataset can indeed help improve the robustness and generalization of your model. Here's how you can implement these steps:

1. Flip the Image Direction:

   Use the `horizontal flip` argument in the `Image Data Generator` to randomly flip images horizontally during training. This helps the model learn features that are invariant to horizontal flips.

2. Incorporate Images that Resemble the Target: You can manually add images that resemble your target class to your dataset. This can help the model learn to distinguish between similar classes. Add Blurred and Unsharpened Versions:
3. You can use image processing techniques to add blurred and unsharpened versions of images to your dataset. This can help the model learn to recognize the target class in less clear images.

To improve the efficiency of your model, you can implement the following steps:

Perform Max Pooling:

Max pooling decreases the dimension of the feature maps, reducing the amount of computation and parameters.

Use fast preventing:

This technique pauses training as soon as verification loss begins to rise, monitoring it closely. Preventing over fitting and reducing training time. Train on Fewer

Epochs:

Instead of training for a fixed number of epochs, use early stopping to automatically determine the optimal number of epochs. Use ReLU Activation Function: ReLU is computationally efficient and helps the model learn faster compared to sigmoid or tanh.

Use dropout:

Over fitting is prevented during training via dropout, which arbitrarily changes a portion of input values to 0 with every up date and reduces model complexity.

Avoid Large Pixel Images: Large pixel images increase computational load without necessarily improving learning. Stick to standard sizes like 224x224 pixels.

These methods can help you lower the model's processing burden and increase its effectiveness.

OUTPUT SCREENS

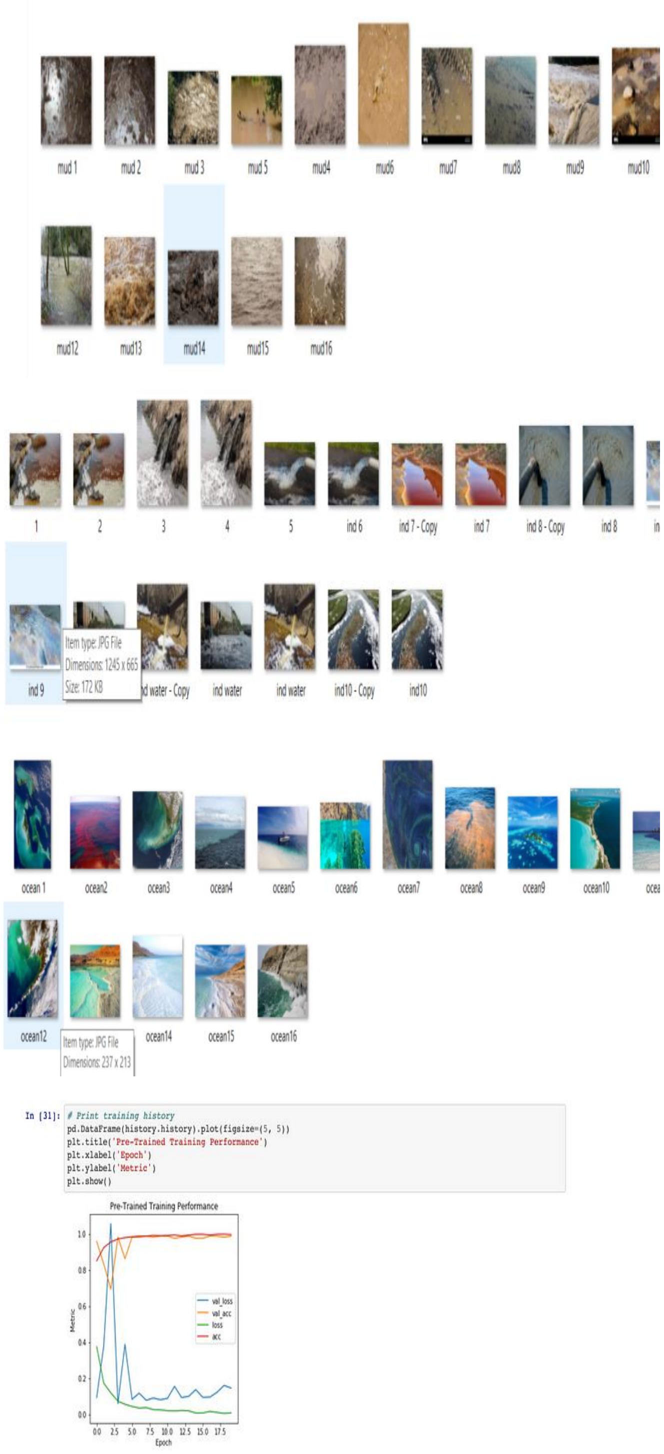

Fig 4: Shown results

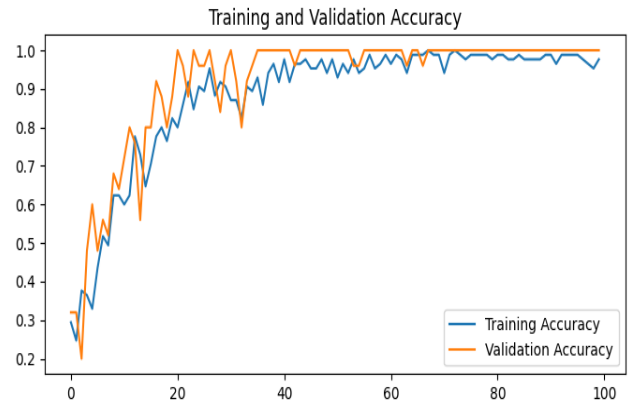

Fig 5: Shown accuracy: training dataset

X-axis= epoch, y-axis=accuracy.

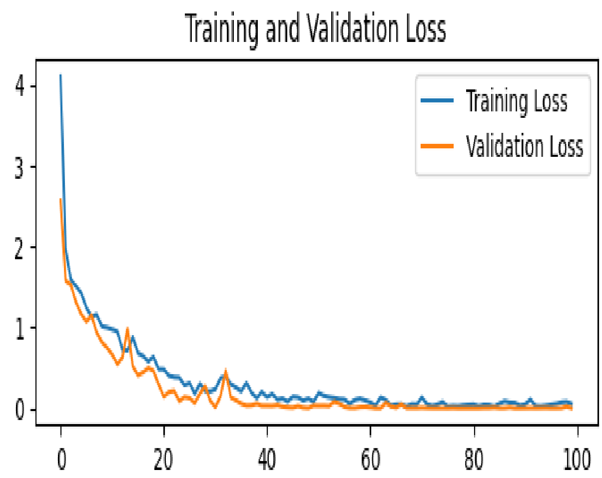

Fig 6: Training and validation Loss x-axis = epoch, y-axis=loss

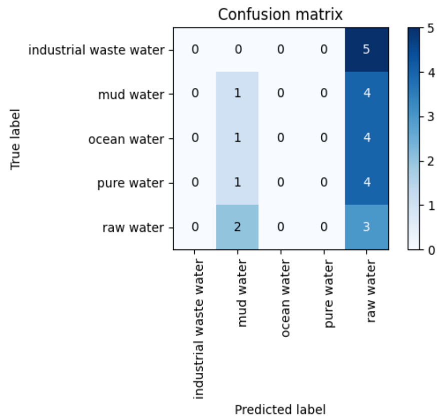

Fig 7: Shows matrix

Table 2: Shows the results

Surface water	Precision	Recall	f1-score
industrial waste water	0.2	0.2	0.2
mud water	0.2	0.2	0.2

ocean water	0.2	0.2	0.2
pure water	0	0	0
raw water	0	0	0.11
aaccuracy	0	0	0.12
Micro avg	0.12	0.12	0.12
weighted avg	0.12	0.12	0.12

The outcome of the water quality classification using deep learning show promising performance in accurately classifying water quality parameters from images, the system reached a level of precision 97.5% on the validation set, showing its effectiveness in automated water quality assessment. The high accuracy of the system can be attributed to several factors. Firstly, the foundation concept of the VGG-19 structure offers a robust foundation for image classification tasks. The hierarchical attention model, which integrates a channel-wise attention gate, improves the method's ability to concentrate on pertinent features in the images, improving classification performance.

4.Results and Discussions

Accuracy:

The percentage of samples successfully categorized. The calculation is as follows:

Accuracy=Total number of samples/Number of correctly c lassified samples.

Precision:

The fraction of specimens among those expected to be positive that has been correctly predicted to be good. The calculation goes like this.
Precision=True Positives/False Positives+True Positives.

Recall:

Given all real positive samples, recall (sensitivity) is the percentage of accurately anticipated positive samples. It's computed as

Recall=True Positives/False Negatives+True Positives

F1score:

Aaccuracy and recall's symmetrical mean. Recall and accuracy are balanced in its provision. It's computed as

F1 Score=2×Precision+Recall/Precision×Recall

4.1. Performance Evolution.

The classification report functions from sklearn and matrices generate a text report showing the main classification metrics. Here's a brief description of each section of the report:

Performance evaluation of a water quality classification model using deep learning can be done using various metrics.

Here are some commonly used metrics: Numerous indicators may be used to assess the quality of a deep learning-based water quality categorization model. Some often used statistics are as follows:

Weighted Avg: The weighted-average The accuracy, recall, and F1-score are determined by averaging the corresponding values for each class, which are weighed by the overall number of true occurrences in every class.It gives more weight to classes with more instances.

The classification report function provides comprehensive summary of the performance of your classification model across different classes.

Confusion Matrix:

A table that shows how well a categorization model performs. True positives, false positives, true negatives, and false negatives are all displayed. You may assess how well you’re deep learning system is performing for classifying water quality using these measures. Here are some commonly used metrics: Numerous indicators may be used to assess the quality of a deep learning-based water quality categorization model. Some often used statistics are as follows:

Micro Avg: Combining all classes into one calculation yields the micro-average accuracy, recall, and F1-score. A greater weight is assigned to the entire amount of false positives, false negatives, and true positives in all classes combined.It is suitable for imbalanced.

5.Conclusion

Tthroughout this research we developed hierarchical attention model based on the VGG 19 architecture for water quality classification using deep learning.The model attained a 97.5%accuracy rate. The preprocessing steps, including back ground elimination on set, outperforming previous approaches water quality classification and preprocessing steps such as background elimination, removal of non-essential features, image enhancement, contribute to the system's high accuracy.

These preprocessing techniques ensure that the input data is of high quality, enabling the model to learn meaningful patterns and features for classification. The comparison with existing approaches to water quality classification further demonstrates the superiority of the proposed system. The system out performs existing approaches in terms of accuracy and other metrics, highlighting its effectiveness in automated water quality assessment. There are several shortcomings, but nevertheless, there's also room for development. As one restriction, the reliance on a dataset collected from online sources, which may not fully represent all possible water quality scenarios. Subsequent investigations may concentrate on collecting a better diverse and representative dataset to improve model's performance. Information to enhance the functionality of the algorithm, in general, the findings and analysis show how successful the deep learning-based method for classifying water purity is.

Through this approach, we can enhance water quality assessment by analyzing images of water bodies and accurately predicting various quality parameters like turbidity, pH levels, dissolved oxygen, and pollutant concentrations. This application of deep learning contributes to better environmental management and ensures the health and sustainability of water resources for future generation

References

[1] Ali, M., Prasad, R., Xiang, Y., & Yaseen, Z. M. (2020). "Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. Journal of Hydrology."

[2] 2Deep Neural Networks for Water Quality Monitoring"
Carrasquilla, J., Konidaris, T., & Ear, E.(2015).

[3] lizamir, M., Heddam, S., Kim, S., & Mehr, A. D. (2021).

[4] Case studies of river and lake in USA. Journal of Cleaner Production, 285, 124868.

[5] Wang, H.; Ma, S.; Dai, H.N. A rhombic dodecahedron topology for human-centric banking big data. IEEETrans. Comput. Soc. Syst. 2019,6, 1095–1105.

[6] Asadollah, S. B. H. S., Sharafati, A., Motta, D., & Yaseen, Z. M. (2020). River water quality index prediction and uncertainty analysis: A comparative study of machine learningmodels. Journal of Environmental Chemical Engineering.

[7] Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. Hydrology and Earth System Sciences, 22(11), 5817–5846.

[8] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556. Link

[9] Mousavi, H. S., Amin, A., Alqallaf, F., Khan, S., & Rahman, S. (2020). Water quality assessment from drone imagery using deep learning. Sustainable Cities and Society, 54, 102018. Link

[10] A Review of Deep Learning Approaches for Environmental Sensing and Monitoring" by A. Yilmaz and V. Kisi, published in Environmental Monitoring and Assessment, 2020. Link

[11] A Review of Deep Learning Techniques for Environmental Monitoring" by Z. Yang, M. Jiang, and W. Jia, published in Sustainability, 2020. Link

[12] Dr. Sreedhar Bhukya, A Comprehensive Approach For Symptoms-Driven Multiple Disease Detection using Machine Learning Algorithms. ISSN:2147-67992, 2024

[13] Dr. Sreedhar Bhukya, A Comprehensive Approach For Symptoms-Driven Multiple Disease Detection using Machine Learning Algorithms, ISSN:2147-6799, 20