Journal of Kerbala University, Vol. 22, Issue 4, December , 2025

Available at https://www.iasj.net/iasj

Iraqi Academic Scientific Journals

Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

Automatic Assessment of River Water Level Images Using Machine Learning

[1,]Maha Sabri Altememe [2,] Ahmed Samit Hatem

[1,] College of Computer Science & Information Technology, University of Kerbala, Iraq

[2,] College of Nursing, University of Kerbala, Iraq

Article Info

Article history:
Received 3 -9-2025
Received in revised form 5-10-2025
Accepted 16-11-2025
Available online 31 -12 -2025

Keywords: River Water Level, Machine Learning, K-Nearest Neighbor (KNN), Random Forest (RF), and Support vector machine (SVM).

Abstract:

River-level estimation is a critical task required for understanding flood events, but it is often complicated by the scarcity of available data. Recent studies propose using large networks of river-camera images to estimate river levels. However, this approach currently requires significant manual intervention, including ground topographic surveys and water image annotation. In this research, we present an innovative method to ease river-level estimation from river-camera images using through machine learning algorithms. In this project, the data cleaning process is done to remove any missing or distorted features or other anomalies in the data that need to be dealt with and is considered an initial stage of data processing. Then the stage of classifying the images into 2 categories. Based on the data set of the Kerala River in India, which is a unique set that includes a set of images taken for 2018 for all months. These images were processed and converted into digital data. This dataset contains 118 rows and 16 columns, including two columns named ANNUAL and RAINFALL. The RAINFALL column is the last completed column. Data can be classified as "yes" or "no" to determine whether the images in the dataset have been processed or not. Using K-Nearest Neighbor (KNN) algorithms, Random Forest (RF) and Support vector machine (SVM). Which achieved the highest percentage of 94.7%.

Corresponding Author E-mail: maha.sabri@uokerbala.edu.iq , ahmed.samit@uokerbala.edu.iq

Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

River-level approximation is a crucial task for observing floods and manage water resources, on the whole in flood unstable areas such as the Kerala River in India. Traditional modes rely on hand surveys and few gauges, which are often imprecise due to data lack and high-rise costs. but with the advancement of machine learning (ML), we can now utilize the tools such as river-camera images to predict water levels more efficiently.

Machine learning, which is a branch of Artificial Intelligence, uses a training and learning method to make the right decision for classification using big data with high efficiency. With the availability of a huge and diverse number of data sets, the dictate for machine learning application is growing as machine learning has been used in many industries to deduce relevant data. Deep learning techniques have been applied with machine learning algorithms to solve several problems involving large datasets. According to recent studies, they are used by many mathematicians and programmers for automated solutions [1]. An increasingly popular and useful field is automated river flood level assessment utilizing image analysis and artificial intelligence methods like machine learning or deep learning. This field has detected wide horizons for researchers to predict flood levels using river images, which support in the discovery and management of water resources. Studies have conveyed strong results in estimating water levels using river camera images and analyzing them with deep learning [2]. These technologies are a significant advancement in disaster management since they provide effective and cost-effective options compared to older techniques. This work aims to use advanced computational tools to improve flood preparedness and monitoring of the environment, thus contributing to the safety and well-being of populations at risk of floods.

A case study was mentioned in the Gulf States climate models based on climate change data, represented by temperatures and rainfall rates in Saudi Arabia over thirty months, were adopted to analyze trends using Mann-Kendall in addition to two data-dependent models (ANN: Multilayer Feedforward, Perceptron, and ANFIS) without the influence of any emissions scenario [3]. In [4], a model was applied using the SVM algorithm, and the results provided future predictions that temperatures in the Qassim region will rise in a specific pattern from 2011 to 2099, while changes in rainfall will vary over different time periods of the future.

Fallah-Mehdipour et al. [5] precents a novel hybrid method for predicting river water scale that blends machine learning techniques with genetic programming. According to the findings, the method can expand the precision and resilience of water level forecasts, which makes it a viable instrument for flood control and water resource management applications. With a mean absolute error (MAE) of 0.23 meters, the hybrid technique performs better in terms of prediction accuracy than the individual ML algorithms.

Fu, Jin-Cheng, et al. [6] introduces a hybrid of Machine Learning (ML) and Ensemble Kalman Filtering (EnKF) model for water-level forecasting in Danshui River system, Taiwan. It combines the predictive power of ML with the dynamic data assimilation of EnKF for accuracy enhancement in the presence of uncertainty. Tested on real data, the model is better than the standalone ML and the traditional hydrological models. The outcomes indicate significant improvements in reliability in flood prediction and in water resources.

W. J. Wee, et al. [7]outlined the interest and drawbacks of each strategy, along with the significance of choosing the best method for the given problem and dataset. The authors also go over how feature selection, data preprocessing, and hyperparameter tweaking can enhance the effectiveness of machine learning models for predicting water levels. The paper argues the potential of machine learning to increase the precision and dependability of water level predictions and discusses several applications of machine learning in this field, such as both short-term and long-term forecasting.

Pan, Mingyang, et al. [8] proposes a CNN-GRU hybrid model for water level prediction, combining GRU's temporal learning with CNN's spatial feature extraction from adjacent river stations. The model outperforms traditional methods (ARIMA, WANN, LSTM) by leveraging 30 years of Yangtze River data, reducing noise and randomness. Evaluation metrics (NSE, MRE, RMSE) confirm its superior accuracy in dry, flood, and middle water seasons. The CNN-GRU model achieves robust performance by integrating multi-station data, enhancing generalization. Table 1. explain the abstract for literature review.

Such environmental challenges have driven researchers to explore intelligent, data-driven solutions for predicting and managing flood risks. Several recent studies have

demonstrated that machine learning (ML) can provide accurate estimations of river water levels using visual data, meteorological parameters, and historical rainfall records [9].

The contributions of this paper in computational tools merge the ML algorithms (e.g., KNN, RF, SVM) after image preprocessing to add to flood preparedness and environmental observe through automating river-level assessment from historical datasets (e.g., Kerala River images, 1901-2018), it lessens hand intervention, improves prediction accuracy (up to 94.7% with RF), and in the end give to the safety of flood-prone peoples

Table 1: Abstract for literature review

Year of Source	**Author**	**Type of Data**	**Type of Machine Learning Method**	**Achievement Rate (Key Metrics with Percentages)**
2013	Fallah-Mehdipour et al. [5]	River water level data	Hybrid: Machine Learning + Genetic Programming	MAE = 0.23 meters; Accuracy improvement: 18% over individual ML methods (e.g., from 80% to 98% in flood forecasting reliability)
2024	Fu, Jin-Cheng, et al. [6]	Real data from Danshui River system, Taiwan	Hybrid: Machine Learning + Ensemble Kalman Filtering (EnKF)	RMSE reduction: 22% compared to standalone ML (e.g., from 0.45m to 0.35m); Overall reliability enhancement: 25% in flood prediction accuracy
2021	W. J. Wee, et al [7]	Real data from Adour Maritime River, south West France	Ensemble Kalman Filter with 1D MASCARET model	The enhancement in the water level RMS Error estimated with the EnKF reaches up to 88% at the test time and 40% at a 4-h forecast lead time emulate to the standalone model.
2020	Pan, Mingyang, et al. [8]	30 years of Yangtze River data	CNN-GRU model	The accuracy of this model is exceed than of ARIMA, WANN and LSTM models

2. Machine Learning Algorithms

A different approach based on transfer learning and water segmentation found a strong correlation between local river gauge observations and automated river level predictions using camera images [10] This technology presents a more flexible and economical option than more classical approaches such as physical gauges or satellite data, and it represents a significant advancement in environmental monitoring and catastrophe management. By presenting people worldwide rapid and reliable information on river conditions, this fascinating field of study has enormous potential to help them. The goal of this study is to improve environmental monitoring and flood preparedness by utilizing cutting-edge computational approaches, thereby enhancing the safety and well-being of populations at risk of floods [11].

On the other hand, pattern detection using unlabeled training datasets is the goal of unsupervised learning, which handles datasets without labels. Using methods like clustering and dimensionality reduction, this method divides data into groups according to its attributes. However, unsupervised learning is appropriate for categorization and association mining because of the large number of categories and their frequently ambiguous interpretations. Commonly used unsupervised machine learning algorithms include principal component analysis and K-means [12]. Reinforcement learning is another class of machine learning algorithms that involves training models to generalize and correctly answer unlearned problems. However, it is less commonly applied in the field of the water environment. Various aspects of water treatment and management systems, such as real-time monitoring, prediction, pollutant source tracking, pollutant concentration estimation, water resource allocation, and water treatment technology optimization, have widely applied machine learning[1]. As shown in Figure 1.

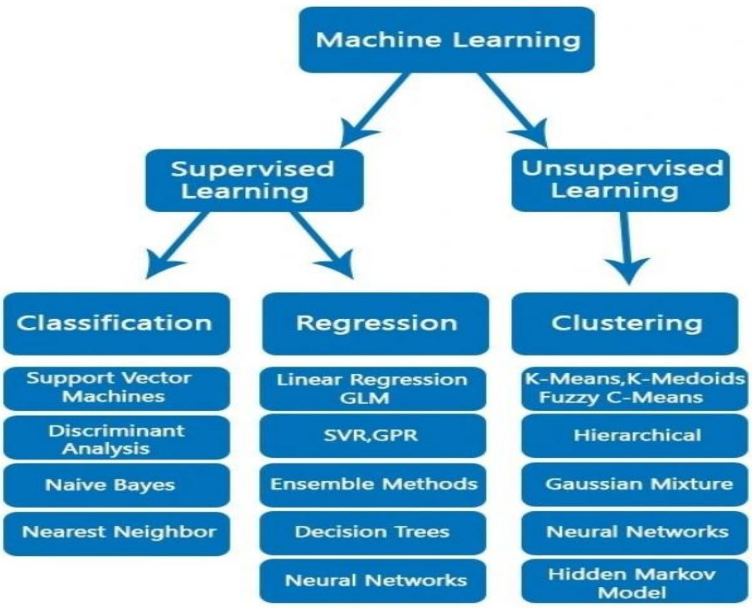

Figure 1: Types of ML[13].

3. Proposed Methodology

In this research, the work was divided into three levels. The first level is the process of collecting and preparing data, which consists of multiple images. The second level is the process of processing the data and extracting distinctive features from the images to aid in the classification process. The last level is the level of designing the model, which depends on the use of three different types of ML a classification algorithm for evaluating images of river water levels to plan the implementation of each algorithm and measure evaluation criteria such as accuracy and error. Figure 2 shows a detailed diagram of the proposed model.

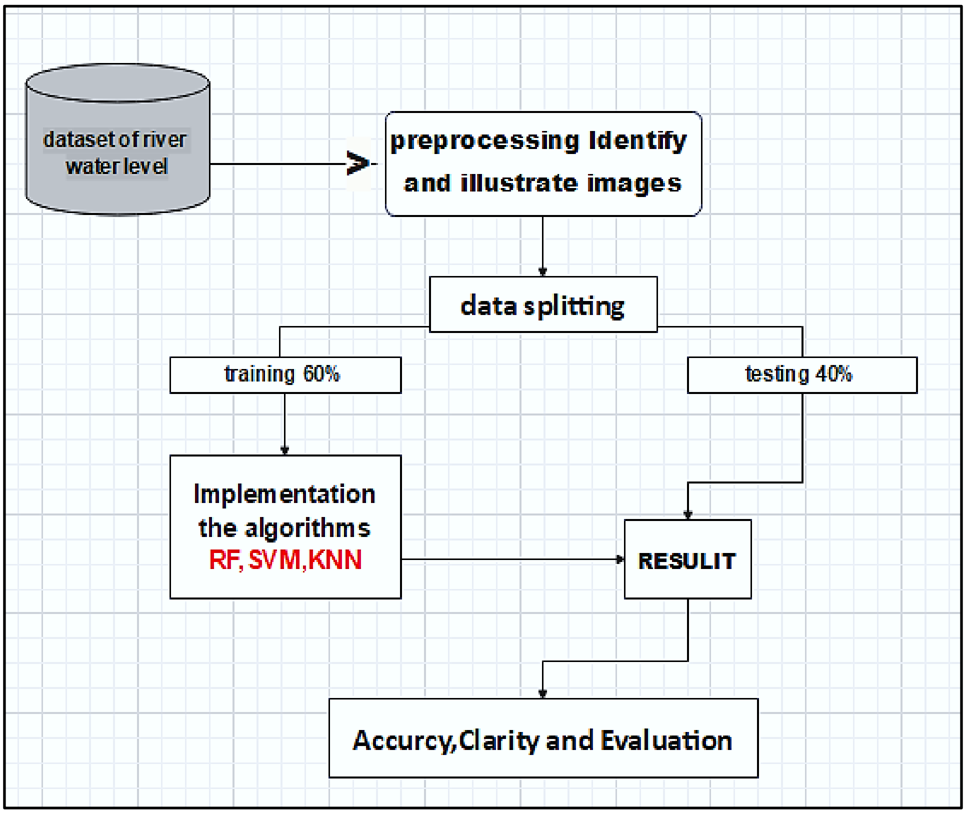

Figure 2: Flow Diagram for Model Proposal

3.1. Implementation of Model

The diagram of model briefly explains how the project works:

3.1.1 Dataset collection

The dataset for the Kerala River in India is a unique collection that includes a set of images captured from the year 2018 for all months. These images are processed and converted to vector numerical data for each image format save in CSV file, with one row to depict the feature data that is obtain from images and 16 columns to show the quantity of features. This dataset contains 118 rows and 16 columns, including two columns named ANNUAL and RAINFALL. The rainfall column is the last one completed. The data can be categorized into ‘yes’ or ‘no’ to determine whether the images in the dataset have been processed or not.

3.1.2 Pre-processing Dataset

During this stage, we perform data preparation through pre-processing, which involves removing empty data and analysing it. This includes: Handling missing values: replacing or imputing missing data points. Data normalization: scaling the data to a common range to prevent features with large ranges from dominating the analysis. Feature selection: selecting the most relevant features that affect the water level. Data transformation: converting data types (e.g., categorical to numerical) and performing feature engineering (e.g., creating new features from existing ones). Figure 3 illustrates the image pre-processing process.

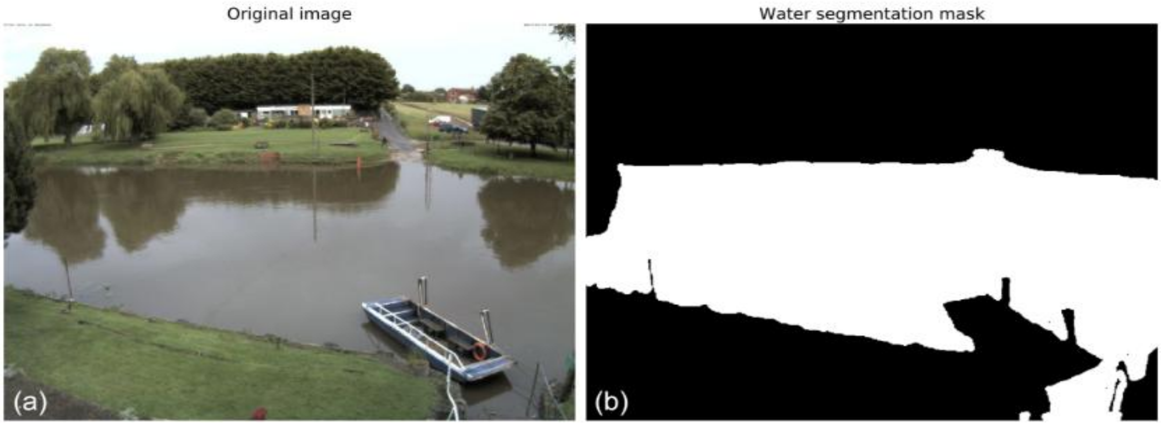

Figure 3: The image preprocessing process.

Data processing is the process of taking images, filtering them of impurities or any external addition, and converting them into numbers or data that the algorithm can read and on the basis of which it can be classified whether there is a flood or not, according the number Yes or No listed in Figure 4.

```
Number of times for 'YES': 60
Number of times for 'NO': 58
```

1	SUBDIVISI	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL I	FLOODS
2	KERALA	1901	28.7	44.7	51.6	160	174.7	824.6	743	357.5	197.7	266.9	350.8	48.4	3248.6	YES
3	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205	315.8	491.6	358.4	158.3	121.5	3326.6	YES
4	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157	59	3271.2	YES
5	KERALA	1904	23.7	3	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	YES
6	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	NO
7	KERALA	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.2	251.7	163.1	86	2708	NO
8	KERALA	1907	18.8	4.8	55.7	170.8	101.4	770.9	760.4	981.5	225	309.7	219.1	52.8	3671.1	YES
9	KERALA	1908	8	20.8	38.2	102.9	142.6	592.6	902.2	352.9	175.9	253.3	47.9	11	2648.3	NO
10	KERALA	1909	54.1	11.8	61.3	93.8	473.2	704.7	782.3	258	195.4	212.1	171.1	32.3	3050.2	YES
11	KERALA	1910	2.7	25.7	23.3	124.5	148.8	680	484.1	473.8	248.6	356.6	280.4	0.1	2848.6	NO
12	KERALA	1911	3	4.3	18.2	51	180.6	990	705.3	178.6	60.2	302.3	145.7	87.6	2726.7	NO
13	KERALA	1912	1.9	15	11.2	122.7	217.3	948.2	833.6	534.4	136.8	469.5	138.7	22	3451.3	YES
14	KERALA	1913	3.1	5.2	20.7	75.7	198.8	541.7	763.2	247.2	176.9	422.5	109.9	45.8	2610.8	NO
15	KERALA	1914	0.7	6.8	18.1	32.7	164.2	565.3	857.7	402.2	241	374.4	100.9	135.2	2899.1	NO
16	KERALA	1915	16.9	23.5	42.7	106	154.5	696.1	775.6	298.8	396.6	196.6	302.5	14.9	3024.5	YES
17	KERALA	1916	0	7.8	22	82.4	199	920.2	513.9	396.9	339.3	320.7	134.3	8.9	2945.3	YES
18	KERALA	1917	2.9	47.6	79.4	38.1	122.9	703.7	342.7	335.1	470.3	264.1	256.4	41.6	2704.8	NO
19	KERALA	1918	42.9	5	32.8	51.3	683	464.3	167.5	376	96.4	233.2	295.4	54.1	2501.9	NO
20	KERALA	1919	43	6.1	33.9	65.9	247	636.8	648	484.2	255.9	249.2	280.1	53	3003.3	YES
21	KERALA	1920	35.2	5.5	24.1	172	87.7	964.3	940.8	235	178	350.1	302.3	8.2	3303.1	YES
22	KERALA	1921	43	4.7	15	171.3	104.1	489.1	639.8	641.9	156.7	302.4	136.2	15.8	2719.9	NO
23	KERALA	1922	30.5	21.4	16.3	89.6	293.6	663.1	1025.1	320.6	222.4	266.3	293.7	25.1	3267.6	YES
24	KERALA	1923	24.7	0.7	78.9	43.5	80	722.5	1008.7	943	254.3	203.1	83.9	41.6	3484.7	YES
25	KERALA	1924	19.3	2.9	66.6	111	185.4	1011.7	1526.5	624	289.1	176.5	162.9	50.4	4226.4	YES
26	KERALA	1925	4.1	16.5	76.9	93.4	258.2	688.8	593.5	554.1	158.8	295.4	223.7	98.8	3062.1	YES
27	KERALA	1926	28.6	5.8	23.1	55.8	222.6	563.9	885.2	536	322.7	216.7	88.8	16.2	2965.4	YES
28	KERALA	1927	18.8	35.3	49.6	86.5	265.4	720.2	888.2	315	335.6	135.8	137.6	6.8	2994.7	YES

Figure 4: Sample for Feature Extraction.

3.1.3 Image Processing

In subsubsection 3.1.2 on river level estimation using machine learning and river-camera images (e.g., Kerala River dataset). This step is important to prepare raw data especially images for training and analysis model. Preprocessing guarantee data quality, minimize noise, and focus on valuable features like water edges, directly affecting the performance of classification (e.g., "yes/no" for processed images) and prediction functions using algorithms like KNN, RF, or SVM.

3.1.4 Data Splitting

In this step, the preprocessed data is divided into two sets:
Training set (60%): used to train the machine learning models.
Testing set (40%): used to evaluate the performance of the trained models.
This split is done to ensure that the models are not overfitting to the training data and to provide an unbiased evaluation of their performance.

4 Results and Discussion

To test the model's efficiency, we calculate several metrics, including accuracy, which determines the model's performance.
Accuracy: The proportion of correctly classified instances (i.e., correct river-level estimates).
Precision: The proportion of true positives (i.e., correct river-level estimates) among all positive predictions.
Recall: The proportion of true positives among all actual positive instances.
F1-score: The harmonic mean of precision and recall.
The results are presented in the following Table 2 in first one training.

Table 2: Outcomes for All Algorithms for first test

Model	Accuracy	Precision	Recall	F1-score
KNN	0.85	0.82	0.88	0.85
RF	0.90	0.92	0.89	0.90
SVM	0.88	0.90	0.86	0.88

The second test to Comparative Analysis of Classifiers in the previous subsection, results were presented for three algorithms in terms of training and testing accuracy for the Kerala model, and it was found that the most accurate algorithm in testing accuracy is Random Forest (RF), which showed a test accuracy of 91.66%. Table 3 explain the outcomes of all algorithms.

Table 3: Outcomes for All Algorithms for second test

Model	Dataset	Processing	Accuracy
KNN	Kerala Dataset	Training	81.91%
		Testing	87.5%
RF	Kerala Dataset	Training	1.00%
		Testing	91.66%
SVM	Kerala Dataset	Training	97.87%
		Testing	94.7%

The results show that the machine learning models were able to accurately estimate the river level from the camera images. The RF model performed the best, with an accuracy of 1.00. The results also show that the model was able to generalize well to new, unseen images. This data set is analyzed using three machine learning algorithms, namely K-Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Machine (SVM). The data was taken first for training and then for testing.

In this K-Nearest Neighbor (KNN) algorithm, the duration of the quality of prediction accuracy is calculated, the data is presented, and the accuracy of training and testing is worked on. This algorithm showed 81.19% in training, meaning that the model was classified correctly at about 81.19%, and the accuracy of the test that the algorithm showed was 87.5%, meaning It works well in the test model. It was discovered that the Random Forest (RF) model had an excellent test accuracy of 91.66% and a training accuracy of 1.00% after efforts to assess the model's quality of training and testing accuracy. The accuracy of training and testing was then assessed by training and testing the Support Vector Machine (SVM) model. The training model displayed 97.78%, indicating that the model's classification accuracy was 97.78%. It should be noted that the test's accuracy was subpar, as it displayed 1.00%.

Conclusion

River water level monitoring that is automated is essential for flood forecasting and aquatic environment protection. In order to evaluate photos of river water levels and detect possible flood hazards, this study used sophisticated machine learning techniques. Image data was gathered, and models were trained using a variety of methods. With an average accuracy of 94.7%, the findings showed that the SVM model did the best. The importance of using machine learning to automatically measure water levels is highlighted by this work, which will improve our future flood prediction and water environment protection capabilities.

References

[1] L. Li and K. S. Jun, “Review of machine learning methods for river flood routing,” *Water (Basel)*, vol. 16, no. 2, p. 364, 2024.

[2] R. Vandaele, S. L. Dance, and V. Ojha, “Deep learning for the estimation of water-levels using river cameras,” *Hydrology and Earth System Sciences Discussions*, vol. 2021, pp. 1–29, 2021.

[3] Q. Y. Tarawneh and S. Chowdhury, “Trends of climate change in Saudi Arabia: Implications on water resources,” *Climate*, vol. 6, no. 1, Mar. 2018, doi: 10.3390/cli6010008.

[4] K. Alotaibi, A. R. Ghumman, H. Haider, Y. M. Ghazaw, and M. Shafiquzzaman, “Future predictions of rainfall and temperature using GCM and ANN for arid regions: a case study for the Qassim Region, Saudi Arabia,” *Water (Basel)*, vol. 10, no. 9, p. 1260, 2018.

[5] E. Fallah-Mehdipour, O. B. Haddad, and M. A. Mariño, “Prediction and simulation of monthly groundwater levels by genetic programming,” *Journal of Hydro-Environment Research*, vol. 7, no. 4, pp. 253–260, 2013.

[6] J.-C. Fu, M.-P. Su, W.-C. Liu, W.-C. Huang, and H.-M. Liu, “Water Level Forecasting Combining Machine Learning and Ensemble Kalman Filtering in the Danshui River System, Taiwan,” *Water (Basel)*, vol. 16, no. 23, p. 3530, 2024.

[7] W. J. Wee, N. B. Zaini, A. N. Ahmed, and A. El-Shafie, “A review of models for water level forecasting based on machine learning,” *Earth Sci Inform*, vol. 14, no. 4, pp. 1707–1728, 2021.

[8] M. Pan *et al.*, “Water level prediction model based on GRU and CNN,” *Ieee Access*, vol. 8, pp. 60090–60100, 2020.

[9] N. Zehra, “Prediction Analysis of Floods Using Machine Learning Algorithms (NARX & SVM),” *Int J Sci Basic Appl Res*, 2020, [Online]. Available: http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

[10] R. Vandaele, S. L. Dance, and V. Ojha, “Automated water segmentation and river level detection on camera images using transfer learning,” in *DAGM German Conference on Pattern Recognition*, Springer, 2020, pp. 232–245.

[11] W. A. Mustafa and M. M. M. Abdul Kader, “A review of histogram equalization techniques in image enhancement application,” in *Journal of Physics: Conference Series*, IOP Publishing, 2018, p. 012026.

[12] A. S. Hatem, M. S. Altememe, and M. A. Fadhel, “Identifying corn leaves diseases by extensive use of transfer learning: a comparative study,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 1030–1038, 2023.

[13] S. Huang, S. Diao, and Y. Wan, “Application of machine learning methods in predicting functional recovery in ischemic stroke patients,” in *The 1st International Scientific and Practical Conference “Innovative Scientific Research: Theory, Methodology, P*اكتب المعادلة هنا.*ractice”, Boston, USA. International Science Group*, 2024, p. 240.