Change Detection in Urmia Lake Using Sentinel-2A Satellite Imagery and Machine Learning Approaches Post-Water Transfer Projects

Hadi Kheiri Gharajeh – 10946669

ABSTRACT

This study investigates the changes in Urmia Lake, Iran, using Sentinel-2 satellite imagery and machine learning algorithms—Random Forest, Decision Tree, and Gradient Boosting—from 2017 to 2024, focusing on periods before and after water transfer projects from the Kani Sib Dam. The analysis revealed a net decrease of approximately 69.60 square kilometres in the lake's water surface area, despite localized improvements following the water transfer efforts. Random Forest proved to be the most accurate classification method for detecting changes in land cover. These findings highlight the limitations of current restoration measures and underscore the importance of advanced remote sensing and machine learning techniques in supporting sustainable water management and conservation strategies for Urmia Lake and similar ecosystems worldwide.

1. INTRODUCTION

Urmia Lake, located in northwest Iran, is the largest saline lake in the Middle East and the third largest globally. Over recent decades, the lake has experienced significant shrinkage due to factors such as drought, increased agricultural water use, and dam construction, which have collectively contributed to severe ecological, economic, and social challenges[1-3]. The drastic decline in the lake's water level has prompted urgent environmental concerns, necessitating various restoration initiatives by Iranian authorities, including the water transfer project from the Kani Sib Dam, inaugurated in June 2024[4].

Remote sensing technology, particularly through Sentinel-2 imagery, plays a crucial role in monitoring and analyzing changes in Urmia Lake over time. Sentinel-2's high-resolution multispectral data facilitates detailed observation of surface water dynamics, land use, and environmental changes, making it an invaluable tool for assessing the impacts of both natural and human activities on the lake[5]. Recent advances in machine learning techniques, such as Random Forest, Decision Tree, and Gradient Boosting, have further enhanced the ability to analyze remote sensing data for environmental monitoring, providing robust methods for detecting land cover changes and water extent variations in complex ecosystems like Urmia Lake[6, 7].

This study aims to analyze the changes in Urmia Lake between 2017 and 2024 using Sentinel-2 images, focusing on the period before and after the second phase of the water transfer project. By applying three different machine learning algorithms—Random Forest, Decision Tree, and Gradient Boosting—the study seeks to evaluate the effectiveness of these methods in detecting changes in the lake's water surface area and surrounding environment. The findings will offer critical insights into the impacts of restoration efforts and help inform future strategies for sustainable water management in the region[8, 9].

Understanding the dynamics of saline lakes like Urmia is vital not only for regional ecological health but also for broader scientific inquiries into the effects of climate change and human intervention on similar ecosystems worldwide. The decline of Urmia Lake mirrors trends observed in other saline lakes globally, such as the Great Salt Lake in the USA and the Aral Sea in Central Asia, where declining water levels have led to increased salinity, biodiversity loss, and significant alterations to local climates[10-12]. These parallels underscore the need for integrated water resource management and environmental conservation strategies.

*Figure 1 : **Urmia Lake***

This study contributes to the understanding of Urmia Lake's ongoing transformation while advancing the application of remote sensing and machine learning in environmental monitoring. By providing a comprehensive analysis of the lake's changes during a critical period, the research highlights the role of technological advancements in addressing environmental challenges and supporting sustainable development goals[13]. The results of this research are expected to guide policymakers and stakeholders in making informed decisions to preserve Urmia Lake and similar ecosystems facing ecological degradation.

In conclusion, the integration of advanced remote sensing techniques with machine learning algorithms presents a powerful approach to monitoring environmental changes. Focusing on Urmia Lake, this study offers valuable insights into the complex interplay of natural and anthropogenic factors influencing the lake's dynamics, emphasizing the importance of continued research and intervention efforts to ensure the sustainable management of this critical water resource[8, 14].

2. MATERIALS AND METHODS

2.1 Study Area

Lake Urmia in northwestern Iran once spanned 5,000 to 6,000 square kilometres, but it has significantly shrunk due to factors such as water mismanagement, climate change, and dam construction[2, 3, 7]. This decline has adversely affected nearly 6 million residents, leading to reduced agricultural productivity, increased soil salinity, and health issues stemming from salt-laden dust storms[8, 9]. The Kani Sib Dam, located southwest of the lake, plays a crucial role in managing water flow to help replenish Lake Urmia, thereby mitigating some of the adverse environmental effects[4]. Figure 2 shows the locations of Urmia Lake and the Kani Sib Dam, highlighting their geographical relationship.

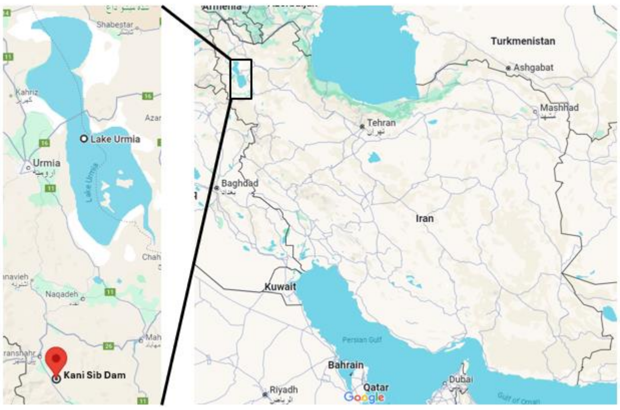

*Figure 2: **Location of Urmia Lake and Kani Sib Dam***

2.2 Data Acquisition

For this study, Sentinel-2 Level-2 satellite images were acquired for two time periods: August 2017 and August 2024. These dates were selected to capture the conditions of Lake Urmia before and after the first and second phases of the water transfer project from the Kani Sib Dam, which began on June 11, 2024. The analysis utilized Bands 2, 3, 4, and 8, providing high-resolution multispectral data essential for distinguishing various land cover types and assessing changes in the lake's water extent[5]. Additionally, a Region of Interest (ROI) shapefile was prepared for each image to facilitate systematic evaluation of land cover transformations and water dynamics over the study period.

Sentinel-2 Bands	Resolution [meters]
Band 2 - Blue	10
Band 3 - Green	10
Band 4 - Red	10
Band 8 - NIR	10

*Table 1: **Different Band set***

2.3 Preprocessing

The preprocessing phase involved cloud masking, mosaicking, and histogram matching to prepare Sentinel-2 images for analysis. Cloud masking was used to remove cloud-covered pixels that can obscure land-cover features, thus enhancing classification accuracy by focusing on unobstructed data[6]. Mosaicking was employed to combine multiple images into a seamless representation of the study area, providing consistent coverage and minimizing edge effects, which improves the reliability of change detection over time. These preprocessing steps were systematically applied to the Sentinel-2 Level-2A data.

To further ensure data consistency, histogram matching was conducted using the Scikit-image library. This process aligned pixel value distributions between images from 2017 and 2024, correcting for differences in lighting, atmospheric conditions, and sensor variations. Such adjustments are essential for ensuring that images are comparable across different time periods, thereby enhancing the performance of machine learning models in detecting changes in land cover and water extent in Urmia Lake.

Figure 3 illustrates the histograms and cumulative distribution functions (CDFs) for the spectral bands (Blue, Green, Red, and Near-Infrared (NIR)) used in the analysis. By aligning these distributions, the preprocessing steps directly improve the consistency and accuracy of temporal analyses, ultimately supporting the study's ability to monitor changes in Urmia Lake with higher precision.

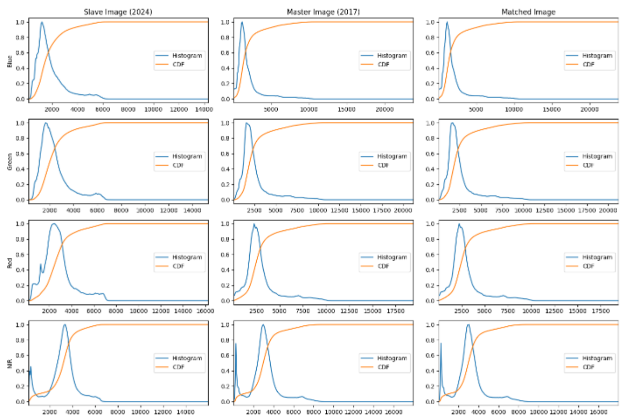

*Figure 3: **Histogram matching for 2017 and 2024***

2.4 Training Data

A GeoPackage (ROI) file with three labelled polygons—Non-Water-Salt, Water, and Salt—was prepared for both image sets to support the supervised classification of Sentinel-2 imagery. These polygons were strategically selected to capture different land cover types, enabling the extraction of spectral signatures from the images[7]. Each pixel within these polygons was labelled according to its land cover class, creating a robust training dataset essential for machine learning classifiers like Random Forest and Decision Tree to accurately recognize unique spectral characteristics[6]. This process ensures precise analysis of the study area and aids in monitoring changes in land cover and water distribution, providing critical insights into Urmia Lake's environmental dynamics.

- **Class 1: Non-Water-Salt** - This class includes all areas that are neither water nor salt, such as soil, urban areas, vegetation, and other landforms.

- **Class 2: Water** - This class covers areas containing water, including the interior regions of Urmia Lake, rivers, and other water bodies.

- **Class 3: Salt** - This class represents the salty regions around the lake that have been exposed due to the lake's drying up, characterized by high salt concentrations.

2.5 Classification Approaches

2.5.1 Random Forest (RF)

Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs for improved accuracy and stability. It is highly effective in remote sensing due to its ability to handle high-dimensional data and minimize overfitting[6]. Its robustness against noise and capacity to capture nonlinear relationships make it ideal for complex tasks like distinguishing land cover types around Urmia Lake.

2.5.2 Decision Tree (DT)

Decision Tree is a simple and intuitive supervised learning method that uses a tree-like model to make decisions based on input features. It is particularly useful for multiclass problems and provides a clear visualization of decision rules. However, Decision Trees can be prone to overfitting and may require pruning or other techniques to improve generalization. Despite these challenges, Decision Trees are computationally efficient and effective for quick classifications.

2.5.3 Gradient Boosting (GB)

Gradient Boosting is an advanced machine learning technique that builds models incrementally by adding weak learners, usually decision trees, to correct errors from previous iterations, thus enhancing overall prediction accuracy[6]. This method excels at capturing complex data patterns but requires careful tuning of parameters to avoid overfitting and handle its computational demands effectively. Its ability to refine predictions iteratively makes it particularly useful for nuanced environmental data analysis, such as detecting subtle changes in land cover and water extent in Urmia Lake.

➢ Reason of Chosing these Tree Algotithms:

1. **Handling of High-Dimensional Data**: Random Forest and Gradient Boosting are well-suited for high-dimensional remote sensing data due to their robustness against noise and ability to capture complex data patterns[6]. This makes them ideal for accurately classifying land cover changes in Urmia Lake.

2. **Accuracy vs. Computational Efficiency**: Decision Trees offer simplicity and speed, making them a good baseline for comparison with advanced ensemble methods like Random Forest and Gradient Boosting, which generally provide higher accuracy at the cost of increased computational demands[5, 6]. This trade-off highlights the balance between accuracy and computational efficiency in choosing models for environmental analysis (Fathian et al., 2016).

3. **Suitability for Multiclass Classification**: Random Forest, Decision Tree, and Gradient Boosting are suitable for multiclass classification, making them effective for distinguishing multiple land cover types around Urmia Lake [6, 7].

3. RESULTS DISCUSSION

3.1 Classification Accuracy

The classification accuracy of three machine learning models—Random Forest (RF), Decision Tree (DT), and Gradient Boosting (GB)—was evaluated using Sentinel-2 imagery to analyze changes in Urmia Lake. The dataset was divided into 70% for training and 30% for testing using Scikit-learn with a random state of 42. Table 2 shows the overall accuracy of each model, with Random Forest achieving the highest accuracy for both 2017 and 2024, as shown in Figure 4. This indicates that Random Forest is the most effective model for classifying land cover types and detecting changes in the lake's water and surrounding areas.

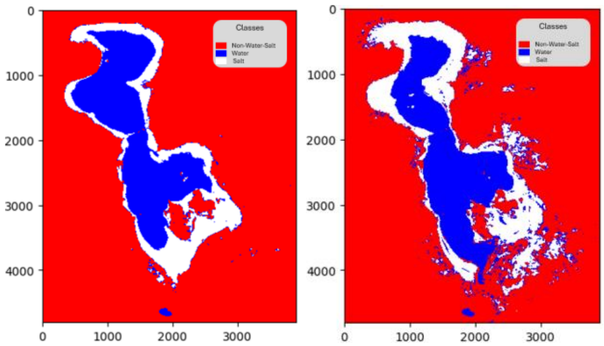

*Figure 4: **Random Forest Classification for 2017 and 2024***

The confusion matrix in Figure 5 provides a performance summary of the classification model for the year 2017. The matrix includes the following classes:

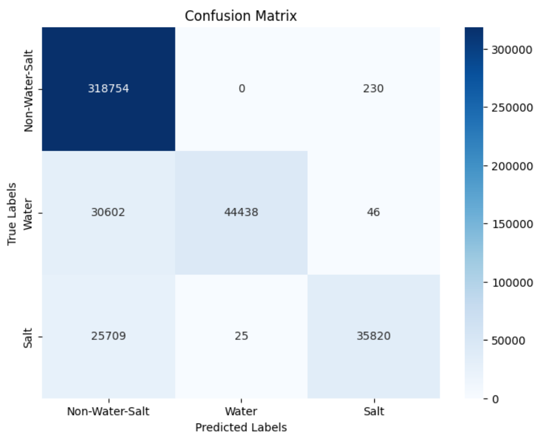

*Figure 5: **Confusion Matrix of RF for 2017***

- **Non-Water-Salt**: The model accurately predicted 318,754 instances as Non-Water-Salt. There were 230 instances incorrectly classified as Salt, and none were misclassified as Water.

- **Water**: The classifier correctly identified 44,438 instances of Water. However, 30,602 Water pixels were misclassified as Non-Water-Salt, and 46 as Salt, indicating some confusion between water and non-water areas.

- **Salt**: The model correctly classified 35,820 instances as Salt. However, 25 Salt instances were misclassified as Water, and 25,709 as Non-Water-Salt, indicating a significant amount of confusion with the Non-Water-Salt class.

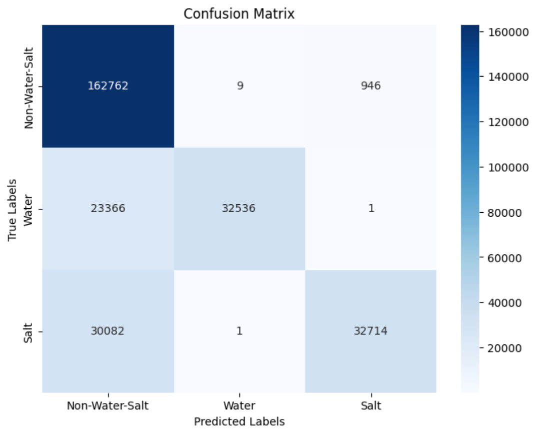

*Figure 6: **Confusion Matrix of RF for 2024***

The confusion matrix in Figure 6 summarizes the performance of the classification model for the year 2024. Here's a brief description of each class:

- **Non-Water-Salt**: The model correctly classified 162,762 instances as Non-Water-Salt. However, 9 instances were misclassified as Water and 946 as Salt, indicating very minimal confusion with other classes.

- **Water**: The classifier accurately identified 32,536 instances as Water. However, there were 23,366 Water instances misclassified as Non-Water-Salt and 1 as Salt, demonstrating a notable degree of misclassification with the Non-Water-Salt class.

- **Salt**: The model correctly classified 32,714 instances as Salt. Nevertheless, 30,082 Salt instances were misclassified as Non-Water-Salt, and 1 as Water, showing significant confusion with the Non-Water-Salt category.

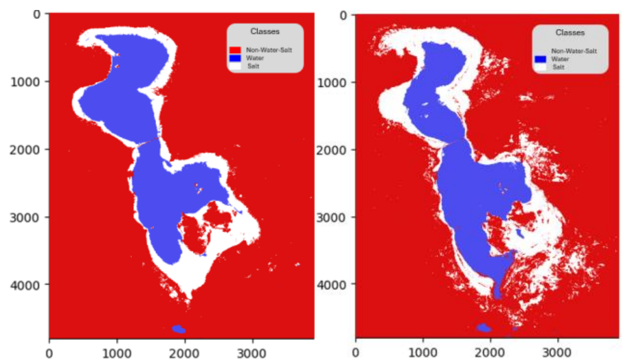

*Figure 7: **Result Sieve Enhancement***

3.2 Sieve:

As shown in Figure 7, the Sieve tool was applied to both images as a post-classification step. This tool helps refine classification results by replacing isolated pixel values with those of the largest neighboring patch, effectively removing small, isolated patches that could be misclassified. This process enhances the overall accuracy of the classification, leading to more reliable change detection results.

3.3 Change Detection

Figure 8 presents a detailed change detection map of Lake Urmia, illustrating land cover transitions between 2017 and 2024, which align with the periods before and after the inauguration of the first and second phases of the water transfer project from the Kani Sib dam, completed on February 24, 2023, and June 11, 2024, respectively. The map highlights key land cover changes, such as transitions from Non-Water-Salt to Water, Water to Salt, and other notable transformations. These changes are categorized into nine distinct classes, each represented by a unique color, as detailed in the legend positioned to the side of the map. This comprehensive legend facilitates the interpretation of the spatial distribution and extent of land cover changes. The change detection map provides valuable insights into the impacts of the water transfer project on the environmental conditions surrounding Lake Urmia. It also enhances our understanding of the spatial dynamics of water and salt coverage, highlighting areas that may require further monitoring or intervention to promote sustainable lake management and restoration efforts.

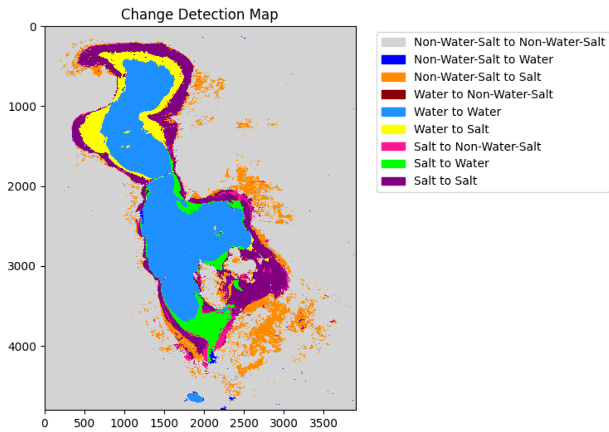

*Figure 8: **Detailed Change Map***

3.4 Result and Discussion

- **Model Performance Comparison:** Random Forest achieved the highest classification accuracy, demonstrating robustness in handling high-dimensional data typical of remote sensing imagery. Its ensemble approach reduces overfitting and captures complex relationships in the data, making it particularly suitable for distinguishing subtle changes in land cover. Gradient Boosting, although slightly less accurate, effectively corrected errors of previous models in sequential stages, which allowed it to capture more complex patterns but also made it more prone to overfitting without careful tuning. Decision Tree, being the simplest model, provided the least accurate results, mainly due to its susceptibility to overfitting and its inability to handle noise as effectively as the ensemble methods.

- **Implications of Misclassifications:** The confusion matrices highlight areas where each model struggled, with notable misclassifications in Random Forest between 'Non-Water-Salt' and 'Water' classes in 2024, likely due to similar spectral characteristics caused by water level changes or salt crusts. Gradient Boosting also had difficulty distinguishing 'Salt' from 'Non-Water-Salt' areas, indicating spectral overlap issues. These errors suggest a need for refined feature selection or additional preprocessing, such as tailored spectral indices, and could be addressed by incorporating extra data layers (e.g., temperature or soil moisture) or adjusting model parameters to better handle spectral ambiguities.

- **Spatial Changes in Urmia Lake:** The analysis revealed notable spatial variability in water coverage changes across Urmia Lake. In the northern part, there was a significant increase in water-covered areas, likely influenced by the effectiveness of the water transfer project and favorable hydrological conditions in this region. Conversely, the southern part experienced a noticeable decrease in water coverage, suggesting that the water transfer project had limited impact or that other factors, such as evaporation rates or lower inflows, might have contributed to the decline.

3.5 Calculations of Changes in Water Coverage Area

I. **Net Change in Water Area (in pixels):** To compute the net change in water area, we consider the number of pixels representing the increase and decrease in water coverage.

Net Change (pixels) = (Increase in water area) − (Decrease in water area)

Given:

<u>Increase in water area</u>: Transition from Non-Water-Salt to Water and Salt to Water:

Increase in water area (pixels)=21,398+324,390=345,788 pixels

<u>Decrease in water area</u>: Transition from Water to Non-Water-Salt and Water to Salt

Decrease in water area (pixels)=6,078+438,103=444,181 pixels

<u>Therefore:</u>

Net Change (pixels)=345,788−444,181=−98,393 pixels

II. **Conversion of Pixel Count to Square Meters:** Each pixel represents an area based on the image resolution. Given that the pixel resolution is $\boldsymbol{30\ meters \times 23.55\ meters}$, the area per pixel is:

Area per pixel=30×23.55=706.5square meters $(\boldsymbol{m^2})$

III. **Net Change in Water Area (in square meters):**

Net Change (m2) =Net Change (pixels)×Area per pixel

<u>Substituting the values:</u>

Net Change (m2) =−98,393×706.5=−69,599,894.5 $\boldsymbol{m^2}$

IV. **Conversion to Square Kilometres:** To convert the area from square meters to square kilometres:

$$\boldsymbol{Net\ Change\ (km^2) = \frac{Net\ Change\ (m^2)}{1,000,000}}$$

$$\boldsymbol{Net\ Change\ (km^2) = \frac{-69{,}599{,}894.5}{-69{,}599{,}894.5} = -69.60km^2}$$

In general, from Figure 8, you can observe that:

Decrease in the Northern Part: There is a decrease in the water-covered area in the northern part of Lake Urmia, possibly due to increased evaporation rates or reduced inflows in this region, which were not fully countered by restoration efforts.

Increase in the Southern Part: The increase in water coverage in the southern part is likely due to the water transfer from the Kani Sib Dam and other sources that provided additional water to this region, highlighting localized impacts of the intervention.

Overall Reduction in Water-Covered Area: Overall, despite localized increases, the lake's total water-covered area decreased by approximately 69.60 square kilometers from 2017 to 2024. This suggests that the lake's overall water loss outweighs gains in specific areas, leading to a net reduction.

4. CONCLUSIONS

This study effectively utilized Sentinel-2 satellite imagery and machine learning algorithms—Random Forest, Decision Tree, and Gradient Boosting—to monitor changes in Urmia Lake from 2017 to 2024, revealing a net reduction in water surface area by approximately 69.60 square kilometres. Random Forest emerged as the most accurate method, underscoring its suitability for complex environmental monitoring. Despite localized improvements from water transfer projects, the overall decline indicates that these measures alone are insufficient to reverse the broader trends of desiccation and ecological degradation. The findings highlight the ongoing challenges of managing saline lakes under the pressures of climate change and human intervention, demonstrating the importance of integrated water resource management and advanced monitoring techniques. These insights can guide sustainable strategies not only for Urmia Lake but also for similar ecosystems worldwide.

REFERENCES

1. Agh, N., et al., *Effects of salinity on survival, growth, reproductive and life span characteristics of Artemia populations from Urmia Lake and neighboring lagoons.* Pakistan journal of biological sciences: PJBS, 2008. **11**(2): p. 164-172.
2. Karbassi, A., et al., *Environmental impacts of desalination on the ecology of Lake Urmia.* Journal of Great Lakes Research, 2010. **36**(3): p. 419-424.
3. Eimanifar, A. and F. Mohebbi, *Urmia Lake (northwest Iran): a brief review*. Saline systems, 2007. **3**(1): p. 5.
4. Dariane, A., et al., *Urmia Lake desiccation and the signs of local climate changes.* Journal of Hydraulic Structures, 2019. **5**(2): p. 1-17.
5. Lillesand, T., R.W. Kiefer, and J. Chipman, *Remote sensing and image interpretation*. 2015: John Wiley & Sons.
6. Hussain, M., et al., *Change detection from remotely sensed images: From pixel-based to object-based approaches.* ISPRS Journal of photogrammetry and remote sensing, 2013. **80**: p. 91-106.
7. Fathian, F., et al., *Trends in hydrological and climatic variables affected by four variations of the Mann-Kendall approach in Urmia Lake basin, Iran.* Hydrological Sciences Journal, 2016. **61**(5): p. 892-904.
8. Khazaei, B., et al., *Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy.* Journal of Hydrology, 2019. **569**: p. 203-217.
9. Sattari, M.T., et al., *Trend and abrupt change analysis in water quality of Urmia Lake in comparison with changes in lake water level.* Environmental monitoring and assessment, 2020. **192**: p. 1-16.
10. Wurtsbaugh, W.A., et al., *Decline of the world's saline lakes.* Nature Geoscience, 2017. **10**(11): p. 816-821.
11. Barnes, B.D. and W.A. Wurtsbaugh, *The effects of salinity on plankton and benthic communities in the Great Salt Lake, Utah, USA: a microcosm experiment.* Canadian Journal of Fisheries and Aquatic Sciences, 2015. **72**(6): p. 807-817.
12. Micklin, P., *The future Aral Sea: hope and despair.* Environmental Earth Sciences, 2016. **75**: p. 1-15.
13. Schmidt, M., R. Gonda, and S. Transiskus, *Environmental degradation at Lake Urmia (Iran): exploring the causes and their impacts on rural livelihoods.* GeoJournal, 2021. **86**(5): p. 2149-2163.
14. Chaudhari, S., et al., *Climate and anthropogenic contributions to the desiccation of the second largest saline lake in the twentieth century.* Journal of Hydrology, 2018. **560**: p. 342-353.