

Prediction of Groundwater Changes Using Artificial Intelligence Models Based on Remote Sensing Images of Surface Water Bodies

Dadajonova Zilola Botirjon kizi^{1,*}, Khamdamov Ulugbek Salim ugli², Nabiye
Raul Timurovich¹

¹ Tashkent University of Information Technologies named after Muhammad
al-Khwarizmi, Tashkent, Uzbekistan

² KDB Bank Uzbekistan, Tashkent, Uzbekistan

* Corresponding author

ziloladadajonova@mail.ru, ulugbek.khamdamov@gmail.com,
ssddts35@gmail.com

Abstract. The study analyzes the annual changes in surface water bodies based on multi-platform remote sensing satellite imagery and links these changes to the depletion of groundwater resources. Among transformer-based architectures, the SegFormer model was selected and trained using water body images from 2020 to 2025. The model identified a linear decreasing trend in water surface area and predicted a decline in groundwater reserves for 2026. According to the results, the water-covered area decreased by 19.6 % during 2020–2025 and is expected to decline by 23.9 % by 2026, with an average annual reduction rate of approximately 3.9 %.

Keywords: Transformers · SegFormer · Remote sensing imagery · Geology · Hydrology · Hydrogeology

1 Introduction

Groundwater remains a primary source of water in many parts of the world, enabling irrigation and supporting rural socio-economic activities [2]. It serves as a vital component of numerous ecosystems and plays a crucial role in the lives of billions of people worldwide [11]. Groundwater is essential in hydrological processes, water resource management, ecological sustainability, and climate change adaptation. In particular, it has a significant impact on the ecosystems of the regions under consideration, which necessitates a deep understanding of its dynamics. Changes in groundwater storage dynamics are of great importance in hydrology, as they directly affect the availability, quality, and sustainability of this resource. Furthermore, in the context of increasing anthropogenic activities and the effects of climate change, accurate and comprehensive assessment of groundwater storage variations is critical for effective water resource management and the development of climate adaptation strategies [5,14]. The object of this study is the pre-Tashkent region. In this context, groundwater monitoring

is carried out based on hydrogeological images obtained through remote sensing methods, and recommendation systems are developed for the rational use of natural resources [6, 9, 15].

2 Materials and Methods

2.1 Data for In Silico Assessment

In recent years, transformer-based architectures have demonstrated high effectiveness in accurately segmenting natural environment images. Previous studies have primarily focused on the segmentation and mapping of surface water sources; however, the relationship with changes in groundwater reserves has not been sufficiently explored [12]. Based on this, we first conduct an analysis of image segmentation models (Fig. 1 and Table 1).

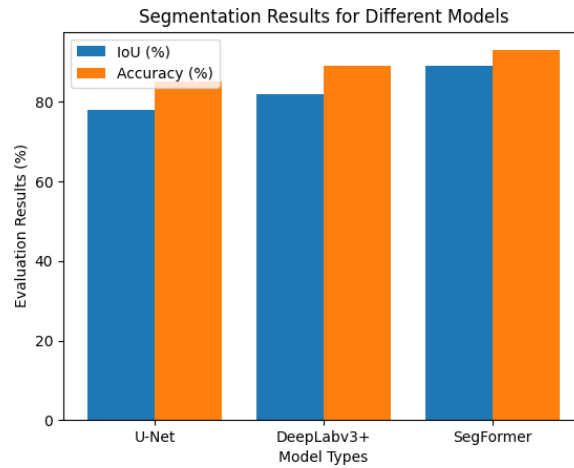


Fig. 1: Comparative analysis of segmentation models

The obtained results show that the SegFormer model achieved higher accuracy in determining the volume of surface water resources from remote sensing images compared to the U-Net and DeepLabv3+ models [3, 13, 16]. Based on the analysis of the models, it can be concluded that, using SegFormer-B0, the sharp decrease in surface water resources can be analyzed, which makes it possible to predict the decline of groundwater in this and similar regions [15].

Table 1: Comparative analysis of segmentation models by accuracy

Model	IoU (%)	F1-score	Accuracy (%)
U-Net	76.5	0.81	83.2
DeepLabv3+	80.7	0.81	89.3
SegFormer-B0	87.3	0.89	93.2

2.2 Determination Rules

During our research, based on analytical results, we use the SegFormer model to achieve high accuracy in processing remote sensing images. The SegFormer (Segmentation Transformer) model is a neural network architecture consisting of two main components, each with its own mathematical formulation [16]. Its primary objective is to perform semantic segmentation, i.e., assigning each pixel of an image to a corresponding class. The mathematical model of SegFormer is mainly expressed through the functions of its two core modules: the Mix Transformer (MiT) encoder (hierarchical Transformer encoder) and the lightweight All-MLP decoder (multi-layer perceptron decoder). The encoder divides the input image into overlapping patches of size 4×4 [16], which can be implemented using convolution operations:

$$P_i = \text{Norm}(\text{Conv}(I_i)), \quad (1)$$

where I_i represents the part of the image at the i -th stage, with kernel size $K = 7$ and stride $S = 4$. As a result, a sequence of tokens is generated. In conventional transformers, the self-attention mechanism has very high computational complexity, $O(N^2C)$, where N is the number of tokens and C is the channel dimension:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_{\text{head}}}}\right) V. \quad (2)$$

To reduce this complexity, SegFormer employs a special mechanism called sequence reduction, which decreases the sequence length of the key (K) and value (V) matrices. In the efficient attention of SegFormer (SRA), K and V are reduced to K' and V' , where R is the reduction ratio:

$$K' = \text{Linear}(\text{Reshape}(K, R)). \quad (3)$$

The $\text{Reshape}(K, R)$ function transforms K into a new sequence of length N/R (or it can be implemented using 2D convolution). Thus, SRA is computed as follows:

$$\text{SRA}(Q, K, V) = \text{Softmax}\left(\frac{Q(K')^\top}{\sqrt{d_{\text{head}}}}\right) V'. \quad (4)$$

This reduces the computational complexity to $O(N^2C/R)$. In the next stage, the Mix-FFN (Mix Feed-Forward Network) is applied. The Mix-FFN in SegFormer does not use positional encoding. Instead, it incorporates spatial information by introducing a 3×3 depthwise convolution into a standard feed-forward network (FFN) layer:

$$F_{\text{out}} = \text{MLP}(\text{DWConv}_{3 \times 3}(\text{GELU}(\text{MLP}(F_{\text{in}})))) + F_{\text{in}}, \quad (5)$$

where F_{in} denotes the input feature, MLP refers to a multi-layer perceptron (linear layers with GELU as the activation function), and $\text{DWConv}_{3 \times 3}$ represents a 3×3 depthwise convolution.

Using the results obtained from the above expressions, the decoding process is carried out. The objective of the decoder is to fuse multi-scale features extracted from the encoder and produce a pixel-level segmentation mask. In the next stage, feature alignment and upsampling are performed. Features at each scale, F_i , are first projected to a common channel dimension C , after which all features are resized to match the highest-resolution feature F_1 (i.e., at $\frac{1}{4}$ scale). The channel projection process maps each feature $F_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i}$ to a unified channel dimension C [16]:

$$\hat{F}_i = \text{Linear}_{C_i \rightarrow C}(F_i), \quad \forall i \in \{1, 2, 3, 4\}. \quad (6)$$

Upsampling: all refined feature maps $\hat{F}_i$ are resized to the same spatial resolution as $\hat{F}_1$, i.e., $\frac{H}{4} \times \frac{W}{4}$:

$$\hat{F}_i^{\text{up}} = \text{Upsample}_{\frac{H}{4} \times \frac{W}{4}}(\hat{F}_i), \quad \forall i \in \{2, 3, 4\}. \quad (7)$$

In the above formulation, $\hat{F}_1^{\text{up}} = \hat{F}_1$. After completing the feature alignment and upsampling stage, the next step is feature fusion. All upsampled feature maps are concatenated and then passed through an additional linear layer for refinement [16]:

$$F = \text{Linear}_{4C \rightarrow C} \left(\text{Concat} \left(\hat{F}_1^{\text{up}}, \hat{F}_2^{\text{up}}, \hat{F}_3^{\text{up}}, \hat{F}_4^{\text{up}} \right) \right). \quad (8)$$

In the next stage, segmentation mask prediction is performed. The final fused feature F is projected through another linear layer to a channel dimension of N_{cls} (number of classes). This function produces class probabilities for each pixel:

$$M = \text{Linear}_{C \rightarrow N_{\text{cls}}}(F), \quad (9)$$

where $M \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times N_{\text{cls}}}$ represents the segmentation mask. The segmentation model, denoted by the function $S(\cdot)$, generates the mask M based on the input image I :

$$M = \text{Decoder}(\text{Encoder}(I)). \quad (10)$$

To obtain the final segmentation mask, after applying the Softmax function, the spatial dimensions of M are upsampled to the original size of the input image, $H \times W$, using interpolation methods such as bilinear or nearest-neighbor interpolation:

$$M_{\text{final}} = \text{Upsample}_{H \times W}(M). \quad (11)$$

In the above formulation, $M_{\text{final}} = S(I)$.

Transformers were originally developed to solve sequence-to-sequence (Seq2Seq) problems such as text generation and machine translation. The Transformer model, which employs self-attention along with encoder and decoder blocks (Fig. 2), represents a next-generation approach designed to transform one sequence into another [4, 17]; transformer encoders are now also widely applied to image analysis tasks [1].

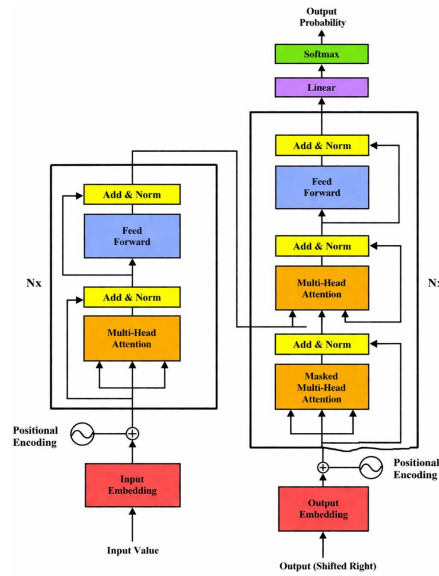


Fig. 2: General architecture of Transformers

Based on this architecture, the SegFormer architecture is constructed (Fig. 3). It uses a hierarchical encoder with efficient self-attention and Mix-FFN blocks together with a lightweight MLP decoder to map an input image to a pixel-level segmentation mask.

3 Results and Discussion

Based on the analysis results, it can be observed that the SegFormer model demonstrates higher accuracy. Therefore, using the SegFormer model, we perform the prediction of groundwater changes based on surface water imagery.

In this study, to obtain reliable results, the parameters of the transformer model are carefully tuned. In this process, 3,200 images collected from the pre-Tashkent region are used, and at the initial stage, suitable images are selected through preprocessing [9]. Examples are shown in Fig. 4.

Based on the images above, representations of surface water reserves and flowing water volumes were obtained across different years. Using this data, the annual decrease in water resources is analyzed, and its impact on groundwater is predicted. For this purpose, the SegFormer model described above is utilized. The model parameters are configured as follows: the AdamW optimizer [10] is used for optimization, cross-entropy loss [7] for error evaluation, a batch size of 8, and 100 epochs for training. The evaluation metrics include F1-score and accuracy [17].

As shown in Fig. 5, the loss values for both training and validation steadily decrease during the training process. This indicates good convergence of the

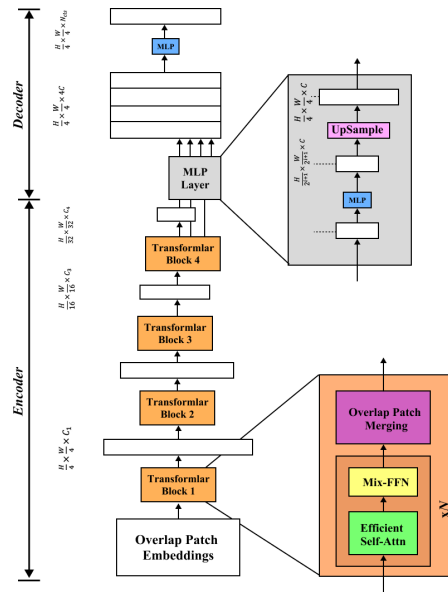


Fig. 3: Architecture of the SegFormer model

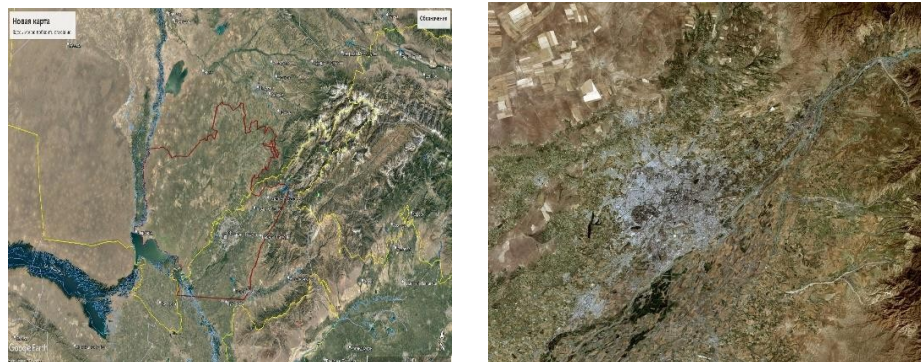


Fig. 4: Remote sensing images of the pre-Tashkent region (dataset)

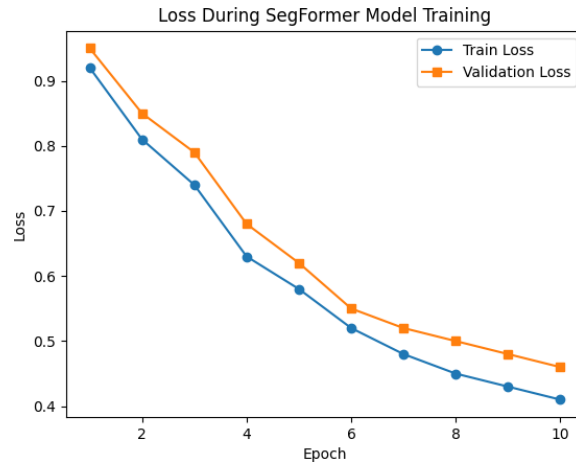


Fig. 5: Training loss during image processing

model. The results demonstrate that regions containing water layers are accurately segmented. Based on the trained model, we predict the impact of changes in surface water body images over the years on groundwater levels and their overall condition [6].

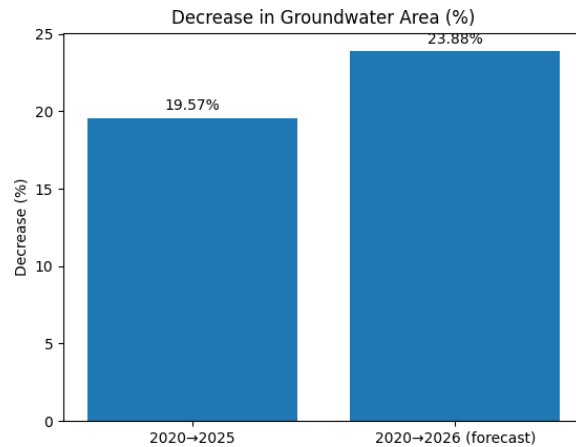


Fig. 6: Prediction of groundwater decrease over the years (%)

The following results were obtained through prediction (Fig. 6). The water area was 512.4 km^2 in 2020, while in 2025 it decreased to 412.1 km^2 . Based on these results, a SegFormer-based prediction for 2026 estimates the water area to be 390.05 km^2 . Accordingly, the total decrease during 2020–2025 is calculated as 19.57 %, while the overall decrease for 2020–2026 reaches 23.88 %. The average annual reduction rate is approximately 3.915 %. Trend-based forecasting of this kind has also been applied to regional agricultural indicators in Uzbekistan [8].

4 Conclusion

In this study, the SegFormer model was used to analyze remote sensing images in order to assess the condition of groundwater resources and predict their future changes. During the research, images from the period 2020–2025 were analyzed, and a linear trend in surface water area change was identified, which enabled the prediction of groundwater depletion for 2026. In 2020, the water-covered area was 512.4 km², while in 2025 it decreased to 412.1 km². According to the linear prediction based on the SegFormer model, the area is expected to decline to 390.05 km² by 2026. These results indicate a decrease of 19.6 % over the period 2020–2025 and 23.9 % over 2020–2026. The average annual reduction rate is estimated to be approximately 3.9 %. It was also observed that water-covered zones significantly decreased from 2025 to 2026, accompanied by an increase in dry land areas. This phenomenon is likely associated with reduced precipitation, rising temperatures, and disruptions in the natural balance of groundwater systems. The obtained results demonstrate the strong potential of the SegFormer model for application in geological, hydrological, and hydrogeological monitoring systems.

References

1. Agarwal, A.K., Hamid, A.B.B.A., Ather, D., Tiwari, R.G., De, I., Krishna, K.R.: Lesion-aware ordinal transformer for diabetic retinopathy classification from fundus images. *Biomedical and Pharmacology Journal* **19**(1), 233–245 (2026). <https://doi.org/10.13005/bpj/3349>
2. Agarwal, A.K., Ather, D., Astya, R., Parygin, D., Garg, A., Raj, D.: Analysis of environmental factors for smart farming: An Internet of Things based approach. In: 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART). pp. 210–214. IEEE (2021). <https://doi.org/10.1109/SMART52563.2021.9676305>
3. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proc. European Conference on Computer Vision (ECCV) (2018)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al.: An image is worth 16×16 words: Transformers for image recognition at scale. arXiv:2010.11929 (2020)
5. Duggal, S., Ather, D., Yusupov, I., Sagar, P., Kumar, R., Jain, V.: Leveraging digital technologies and system dynamics for achieving sustainable development goals across diverse sectors. In: 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS). pp. 1935–1939. IEEE (2024). <https://doi.org/10.1109/ICTACS62700.2024.10841149>
6. Famiglietti, J.S., Rodell, M.: Water in the balance. *Science* **340**(6138), 1300–1301 (2013)
7. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016)
8. Kler, R., Ather, D., Singh, G., Chaudhary, N., Arora, M.: Predicting agricultural growth in Jizzax region using advanced machine learning techniques: An ARIMA-based approach. In: 2023 4th International Conference on Computa-

- tion, Automation and Knowledge Management (ICCAKM). pp. 1–5. IEEE (2023).
<https://doi.org/10.1109/ICCAKM58659.2023.10449567>
9. Kumar, R., Ather, D., Sagar, P., Yusupov, I., Duggal, S., Jain, V.: Integrating predictive models and remote sensing for sustainable agriculture, energy, and environmental management. In: 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS). pp. 1930–1934. IEEE (2024).
<https://doi.org/10.1109/ICTACS62700.2024.10841299>
10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: Proc. International Conference on Learning Representations (ICLR) (2019)
11. McDonald, R.I., Weber, K.F., Padowski, J., Flörke, M., Schneider, C., Green, P.A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G., Montgomery, M.: Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Global Environmental Change* **27**, 96–105 (2014)
12. Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S.: High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2015)
14. Scanlon, B.R., Ruddell, B.L., Reed, P.M., Hook, R.I., Zheng, C., Tidwell, V.C., Siebert, S.: The food-energy-water nexus: Transforming science for society. *Water Resources Research* **53**(5), 3550–3556 (2017)
15. Wada, Y., van Beek, L.P.H., van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P.: Global depletion of groundwater resources. *Geophysical Research Letters* **37**(20) (2010)
16. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and efficient design for semantic segmentation with transformers. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)
17. Xu, H., Lin, Z., Liu, X., Wang, Y., et al.: A survey on semantic segmentation: From CNN to transformer and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022)