ANN-Based Predictive Modeling of Aquifer Dynamics in Quetta Valley

Areeb Ul Haq [1*], Abdul Latif Qureshi1 [1], Arjumand Zehra Zaidi [1]

[1] U.S.-Pakistan Center for advanced Studies in Water, Mehran UET, Jamshoro;
* Correspondence: areebshaikh125@gmail.com

Abstract

Monitoring and understanding groundwater level (GWL) dynamics in the Quetta Valley have gained significant attention due to the increasing reliance on groundwater resources. Accurate forecasting of GWL is critical for effective resource management, mitigating risks such as overexploitation, water quality degradation, and land subsidence. This study employs an Artificial Neural Network (ANN) model to predict GWL fluctuations using data collected from 14 monitoring stations over a 40-year period (1980–2020), encompassing a wide range of GWL changes. The dataset was divided into two subsets: 77.23% for training and 22.76% for testing the model. Model performance was assessed quantitatively using root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). Additionally, qualitative evaluations were conducted through time-series line plots and scatter plots, offering insights into the model's predictive accuracy. The findings provide a deeper understanding of GWL dynamics in the Quetta Valley and demonstrate the potential of ANN models as effective tools for sustainable groundwater management.

Keywords: Aquifer fluctuations; Artificial Neural Networks (ANNs); Quetta Valley

1. Introduction

Aquifers serve as vital water sources for ecosystems and human societies, especially in arid and semi-arid regions where surface water is limited or irregular. In such contexts, groundwater becomes the primary supply for domestic, agricultural, and industrial use. The Quetta Valley, characterized by a semi-arid climate with erratic and scarce rainfall, heavily depends on its groundwater resources to support a growing population and expanding agricultural activities. However, the dynamics of aquifer systems shaped by both natural hydrological processes and increasing anthropogenic pressures pose substantial challenges to sustainable water resource management in the region.

Recent studies highlight the complexities inherent in groundwater flow, recharge mechanisms, and the impacts of climate variability and human interventions on aquifer sustainability (Ahmad et al., 2024; Ahmad et al., 2023; Waseem Muhammad et al., 2023). For instance, multitemporal satellite imagery analysis reveals climate change-induced spatiotemporal variations in land use and land cover, directly affecting recharge and extraction patterns (Ahmad et al., 2024). Similarly, hydrological risk assessments combining hydrodynamic models such as HEC-RAS have illustrated how compound effects of flow and precipitation peaks under changing climate conditions threaten reservoir and groundwater system stability (Ahmad et al., 2023).

Artificial intelligence processes, especially Artificial Neural Networks (ANNs), have become more popular in hydrological modeling in response to these challenges because they are able to identify complex, non-linear relationships in data without a detailed prior understanding of what is happening in a system (Herbert et al., 2021). ANNs have been effectively used to predict groundwater levels, quality and system responses in equally complicated settings. They allow the combination of various datasets such as precipitation, land use transitions, extraction volumes, and climatic indicators that would allow aquifer behavior to be more precisely predicted (Ahmad et al., 2024).

Moreover, the newest developments in machine learning have been applied to the supply forecasting of water and water risk assessment to enhance the resilience of water resource man-management during uncertainty (Herbert et al., 2021; Muqeet et al., 2023). Water demand and supply also take the form of integrated hydrological modeling with socioeconomic scenarios that are used to inform sustainable planning in places such as the Upper Indus Basin, which is also similar in problems to Quetta Valley (Ahmad et al., 2025).

This study aims to develop an ANN-based predictive model to analyze and forecast aquifer dynamics in the Quetta Valley. By leveraging machine learning techniques and incorporating comprehensive environmental and anthropogenic data, the research seeks to improve groundwater management efficiency. The model will serve as a practical decision-support tool for planners and policymakers to devise strategies ensuring sustainable groundwater usage, thereby contributing to the long-term water security of the Quetta Valley in the face of climatic and developmental pressures.

1.1. Research Area

The chosen study area is Quetta, the capital of Pakistan's Balochistan province. The Quetta valley measures a total of 603 km^2 and is located between the latitude of 30 00-30 30 N and 66 40-67 15 E (Khair et al., 2012). (Figure 1).

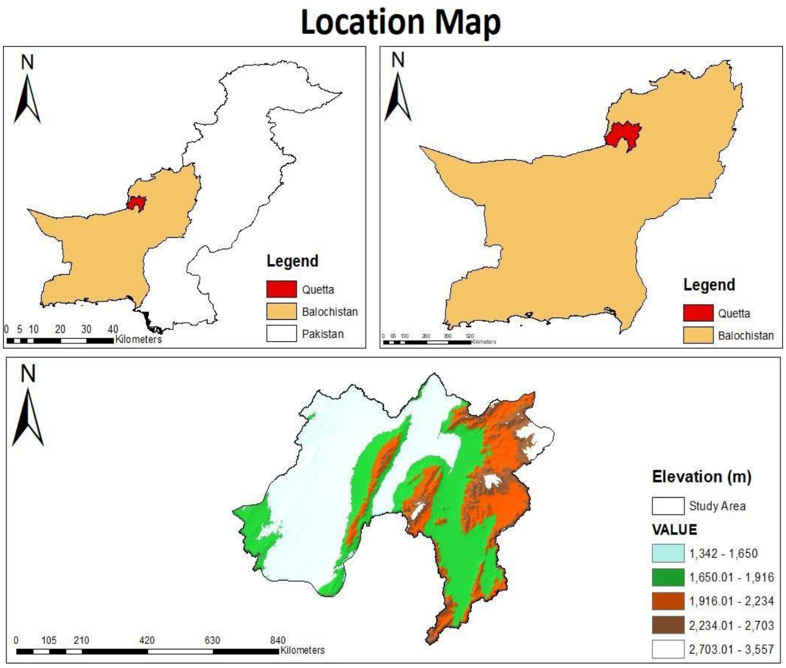

Figure 1. Location Map of the Study Area

1.2. Hydrogeology

There are two aquifers beneath the Quetta Valley. The uppermost layer, the unconfined alluvial aquifer, which forms the valley floor, is made up of a mixture of clay, sand, and gravel. Between 150 and 300 feet below the surface, it is located. From the southern to the northern portion of the valley and from the valley shoulders to the center, the alluvial aquifer's thickness progressively increases (Aftab et al., 2022). The second aquifer, a restricted aquifer made of limestone formation, ranges in depth from 350 ft to 1,200 ft (Imran et al., 2021). The hard-rock formations present on the rock lime-stone and conglomerates are highly jointed and fractured having well-developed sec-ondary porosity and permeability, owing to the active tectonic of the Chaman Fault System. The aquifers directly replenish along the mountain slopes in the form of exposed limestone formations as a result of precipitation and surface runoff (Table 1).

Table 1: Hydrostratigraphic Zonation and Depth Characteristics

Aquifer Types	Lithology	Depth (m)
Shallow Aquifer	Gravel, sand, and silt layer	45 - 91
Deep Aquifer	Limestone strata	106 -- 365

2. Methodology

2.1. Data Collection

The study used groundwater level, precipitation, and temperature as input variables. The Water and Sanitation Agency (WASA) provided piezometer data from the Quetta Valley from 1990 to 2022, which were used to assess groundwater level variations. Monitoring frequency varied from monthly to annual, with nine observation wells yielding a total of 492 data points. The observation well location map is shown in Figure 2.

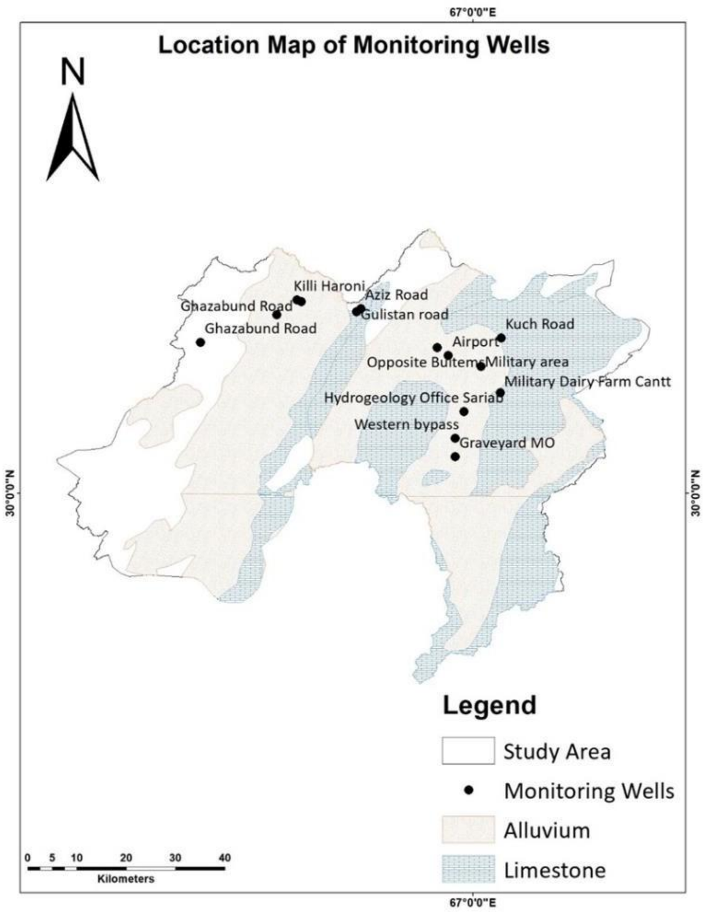

Figure 2. Location map of monitoring wells

2.2. Model Deployment (ANN)

Predictions of groundwater level are made using an empirical, and data-driven Artificial Neural Network (ANN) model. ANN architecture consists of three layers; input layer, hidden layer and the output layer. The feed-forward multilayered perceptron (MLP), which is a popular method of hydrological parameters, was used to train and test ANN model in this study. The use of MLP in the ANN models shows that the model has the ability to address nonlinear, complex relationships involving variables in hydrological forecasting and analysis. This practice shows that machine learning methods can enhance the accuracy and usefulness of frameworks in managing and predicting water sources. MLP is most advantageous when the model to develop is non linear and this is the case in this research to predict. It trains cor-relations between inputs and outputs with a high volume of data. One node (neuron) in a layer of an MLP is linked to all the other nodes in the next layer by some weight (Figure 3 illustrates the ANN structure of the prediction).

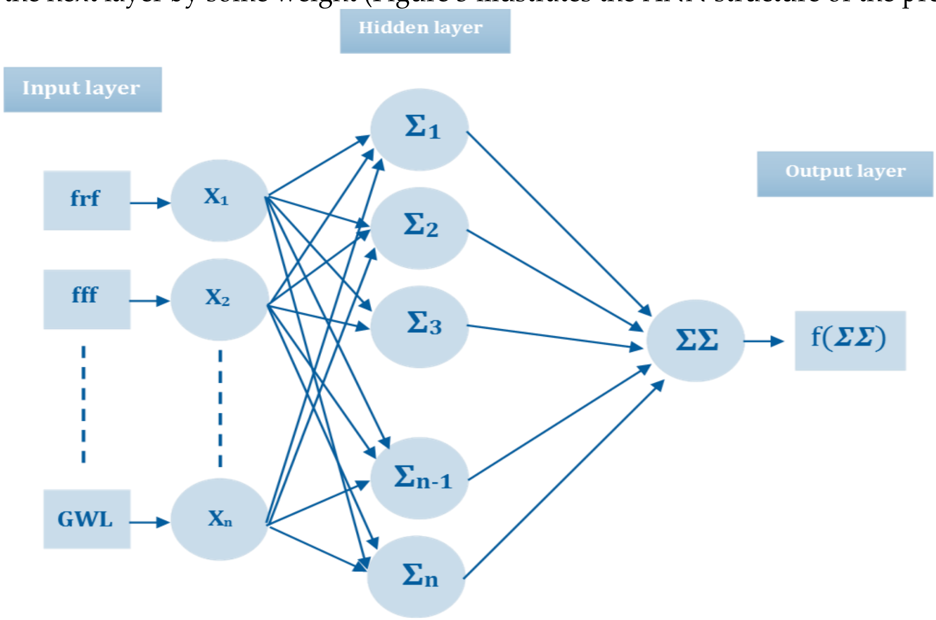

Figure 3. Artificial Neural Network architecture for prediction.

2.3. Training and Testing of data

The nature of data and its preparation are also significant in order to be appropriately trained and test data. The dataset was separated into two portions 77.23/22.76 per cent training and testing respectively as well as validating the model developed.

2.4. Model Calibration

During the training phase at the beginning, the activation functional employed in the hidden layers was the Rectified Linear Unit (ReLU). ReLU activation function is implemented element-wise on the output of each neuron. The formula for ReLU is:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{1}$$

When we have an ANN to predict the level of ground water, where characterizes the output (or weighted sum) of a neuron in the network, the ReLU function will simply yield: Mathematically, this can be written as:

a = 1,

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{Otherwise} \end{cases} \tag{2}$$

This activation job introduces non-linearity to the model and supports the network learn complex patterns in the data. It's worth noting that ReLU has been widely used in deep learning models, including those applied to time series prediction tasks such as groundwater level forecasting.

2.5. Model Validation

We validated the model after deploying and calibrating it with several statistical measures such as Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) to determine the reliability and predictive ability of the deployed model (Table 2). The R2 is between 0 and 1 with 1 being the ideal model and above 0.8 being an acceptable model with good prediction. A negative value of less than 0.5 represents an unacceptable model with a weak predictive power. In the case of RMSE and MAE, however, there is no predetermined range; lower values are said to be better to the performance of a model, and values close to 0 or even 0 is described as a perfect model.

The mathematical precision ions of various statistics applied in the measurement of prediction precision are as follows.

Table. 2: Evaluation Metrics for ANN Model Performance

Performance Indicators for Artificial Neural Network Models		
(RMSE)	Root Mean Squared Error	$\{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{hi}+\mathrm{hi}')\}^{1/2}$
(MAE)	Mean Absolute Error	$\frac{1}{n}\sum_{i=1}^{n}\lvert(\mathrm{hi}+\mathrm{hi}')\rvert$
(R^2)	Coefficient of Determination	$1-\frac{\sum_{i=1}^{n}(\mathrm{hi}+\mathrm{hi}')^2}{\sum_{i=1}^{n}(\mathrm{hi}+\mathrm{hi}')^2}$

Where n = the number of observations or data points, hi = the actual groundwater level for the i^{th} observation, hi′ = the predicted groundwater level for the i^{th} observation and h = the mean of the actual groundwater levels.

3. Results and Discussions

3.1. Assessment of Rainfall Influence on Static Water Levels

The paper provides an in-depth study of the rainfall and static water level changes in alluvium and limestone aquifers between 1980 & 2020 (Fig 4). There are significant changes in both parameters in this period, which are essential in the process of un-derstanding the groundwater dynamics in these aquifers.

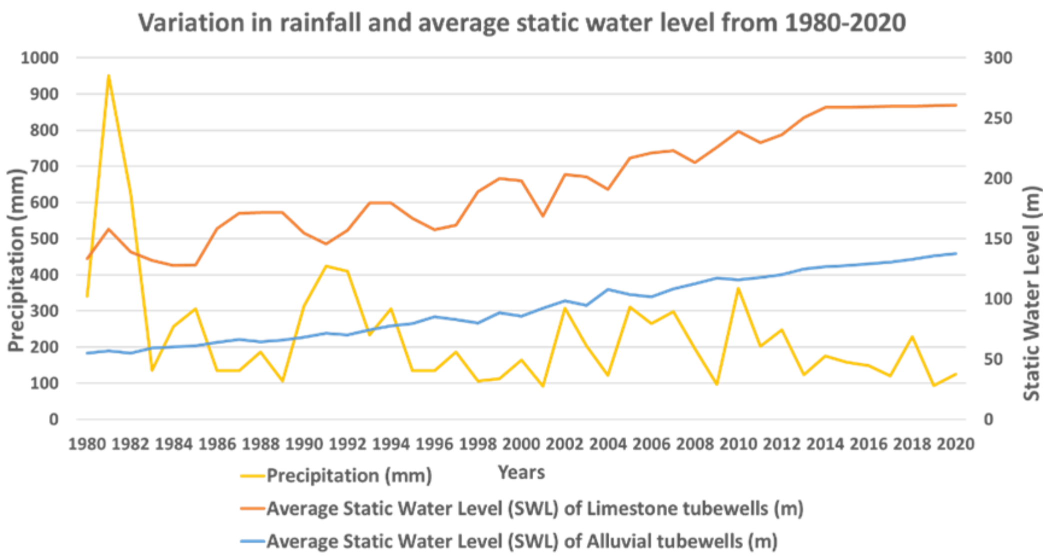

Figure 4. Trends in Rainfall and Mean Static Water Levels (1980–2020) within Alluvial and Limestone Aquifers.

3.2. Evaluation of ANN Model Predictions

The prediction of the static water levels of the two types of aquifers was undertaken using Artificial Neural Network (ANN) models. The ANN model forecasts of the alluvium and limestone aquifers, respectively are shown in Figure 5 and Figure 6. These predictions are important to the future water resource management based on their accuracy.

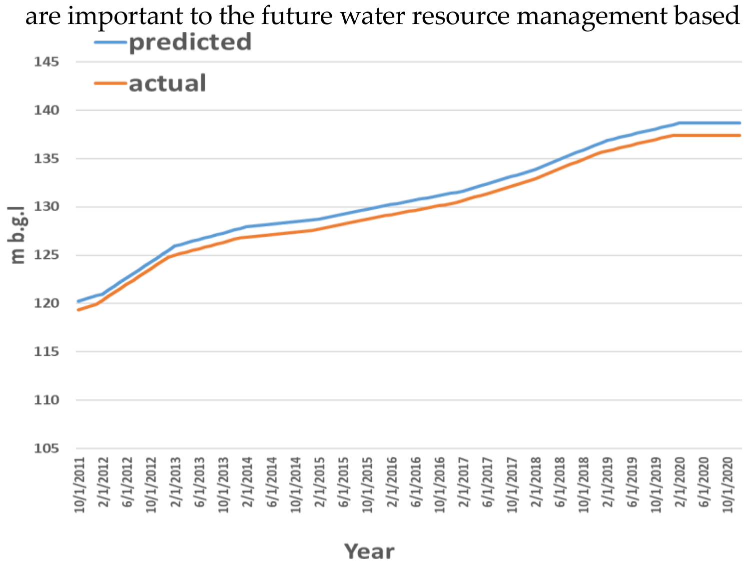

Figure 5. Artificial Neural Network–Based Prediction of Groundwater Levels in the Alluvial Aquifer.

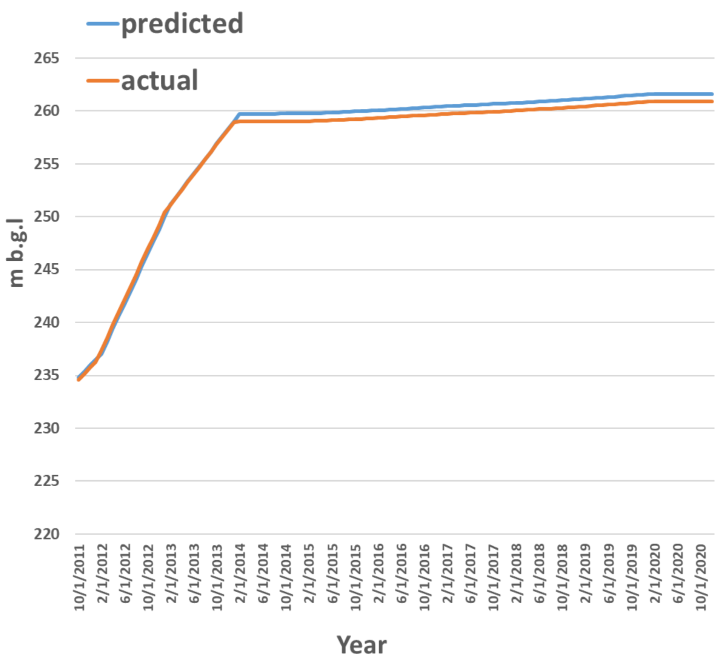

Figure 6. ANN–Based Prediction of Groundwater Levels in the Limestone Aquifer

3.3. Real and Predicted Water Level Compared

. The paper also compares the actual and predicted water table of the two aquifers. The alluvium and limestone aquifers have scatterplots as per figures 7 and 8 which reveals the efficiency of the model in forecasting water levels within the study period.

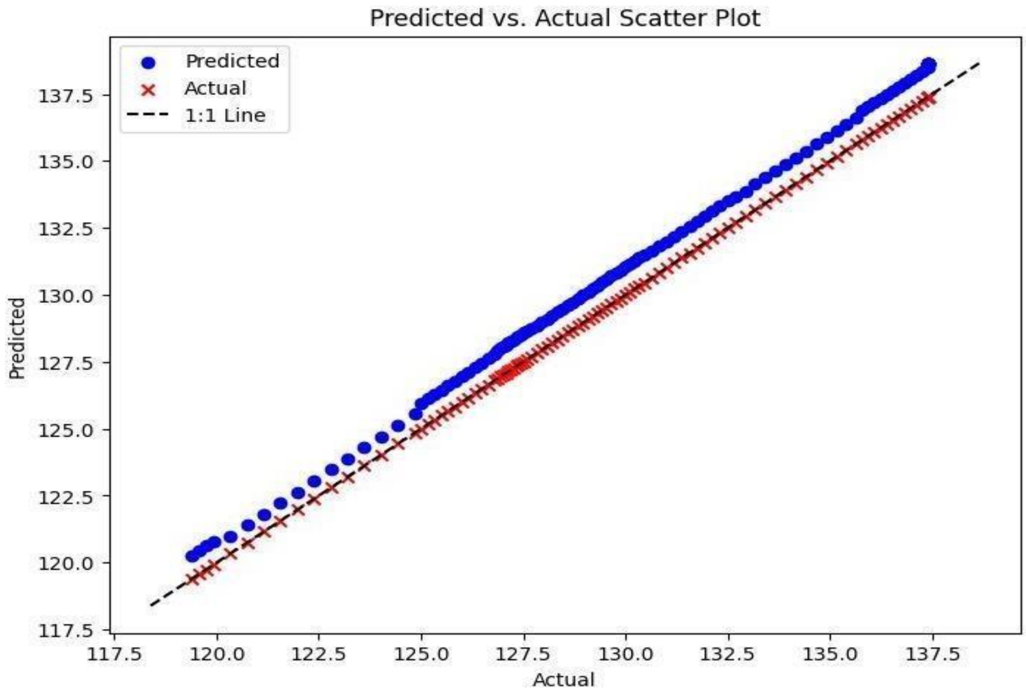

Figure 7. Comparison between Measured and ANN-Predicted Groundwater Levels in Alluvium

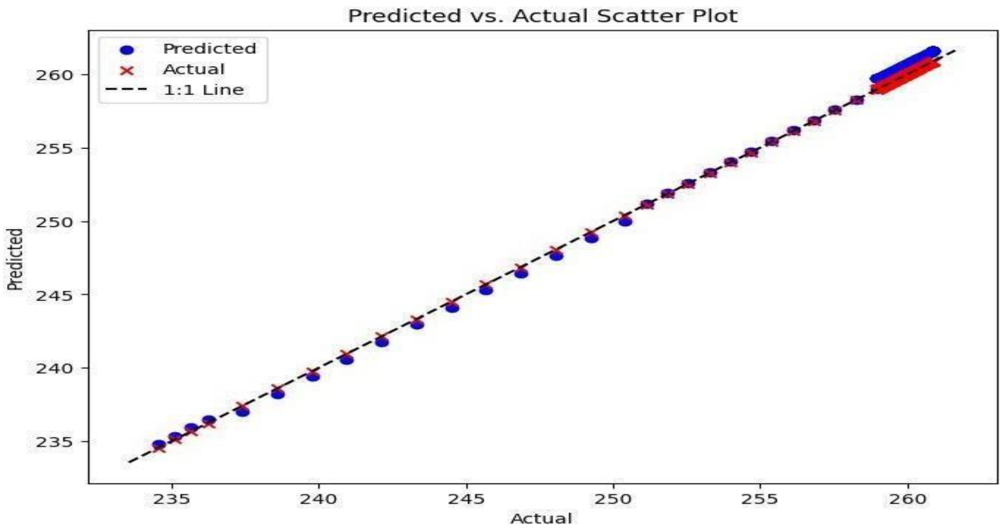

Figure 8. Evaluation of Predicted versus Actual Groundwater Levels in the Limestone Aquifer

3.4. Evaluation Metrics for ANN-Based Predictions

The performance of the ANN model was evaluated using three statistical indicators: the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) (Table 3). Results indicate that the limestone aquifer ($R^2 = 0.991$) achieved a higher predictive accuracy compared to the alluvial aquifer ($R^2 = 0.961$). This suggests that the ANN model demonstrates a stronger reliability in forecasting groundwater levels within the limestone aquifer.

Table 3: Evaluation Metrics for ANN-Based Predictions

Statistical Evaluation			
Aquifer Type	R^2	RMSE	MAE
Limestone	00.991	00.646	00.602
Alluvium	00.961	01.013	01.003

3.5. Groundwater Decline Trends and Socioeconomic Context

Figure 9 highlights spatial variations in groundwater decline rates, with central Quetta experiencing declines up to 2.815 m/year (limestone) and 1.476 m/year (alluvium). These trends correspond to intensive groundwater extraction driven by population growth, agricultural expansion, and unregulated well drilling over the past two decades.

The decline is exacerbated by erratic rainfall patterns linked to climate variability, underscoring the urgency for integrated water management approaches that consider both environmental and socio-economic factors.

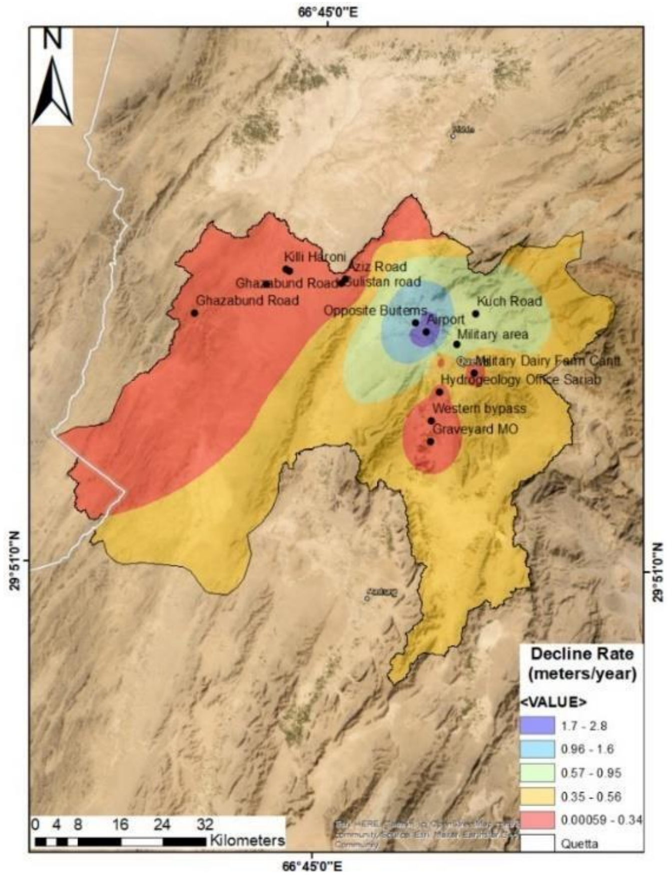

Figure 9. Spatial Distribution of Groundwater Depletion in Quetta.

3.6. Temporal Analysis of Groundwater Levels

The region's susceptibility to diminishing groundwater levels is highlighted by a 40-year study of groundwater levels (Figure 10), which offers important insights into the temporal trends and patterns of groundwater changes.

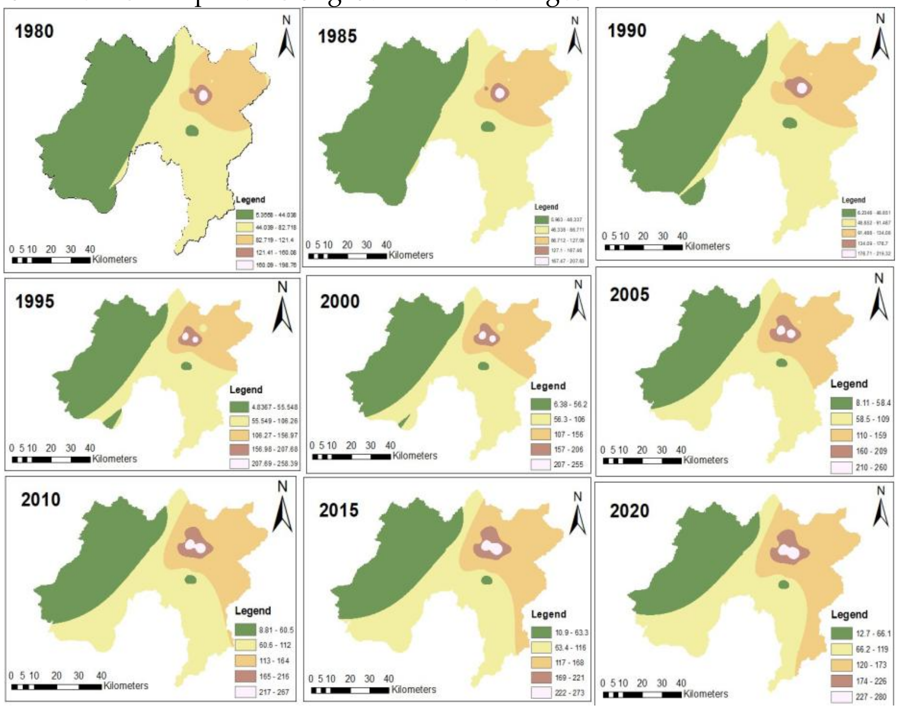

Figure 10. Temporal trends of groundwater levels over 40 years, indicating sustained decline especially in central areas

3.7. Practical Implications for Water Resource Management

The ANN model's accurate groundwater level forecasts enable proactive management strategies. Water authorities like WAPDA can use these predictions to:

- Implement extraction controls based on predicted critical thresholds.
- Plan artificial recharge projects during favorable climatic conditions.
- Issue permits strategically to prevent aquifer overexploitation.
- Integrate predictions into decision-support systems to enhance resilience under uncertainty.

3.8. Limitations and Future Work

The model's spatial accuracy is constrained by the uneven distribution of monitoring wells. Expanding the network and applying geostatistical interpolation methods will enhance predictive coverage. Additionally, integrating socio-economic data and land use change information can improve model comprehensiveness.

4. Conclusion

This study has successfully demonstrated the application of an artificial neural network (ANN) in predicting groundwater levels in the Quetta Valley of Balochistan, Pakistan, with a particular focus on two primary aquifers: the limestone aquifer and the alluvium aquifer. The high R^2 values of 00.992 for the limestone aquifer and 00.962 for the alluvial aquifer, as well as the low mean absolute error (MAE) and root mean square error values, demonstrated the ANN model's remarkable prediction accuracy. These outcomes demonstrate how well the model can predict groundwater processes in intricate hydrogeological environments.

Our findings highlight a concerning trend of declining water levels in both aquifers, with the limestone aquifer experiencing a decline of 2.815 m/year and the alluvial aquifer at 1.476 m/year. This decline is attributed to extensive extraction through unauthorized tube wells and agricultural wells over the past 20 years, leading to significant aquifer exhaustion. The implications of these declining levels are profound, especially considering the importance of the limestone aquifer as a vital drinking water resource.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ahmad, Izhar, Muhammad Waseem, Sadaquat Hussain, and Megersa Kebede Leta. 2024. "Climate Change Induced Spatiotemporal Variations of Land Use Land Cover by Using Multitemporal Satellite Imagery Analysis." Journal of Water and Climate Change 00(0):1–23. https://doi.org/10.2166/wcc.2024.675.
2. Ahmad, Izhar, Muhammad Waseem, Ammar Ashraf, Megersa Kebede Leta, Sareer Ahmad, and Hira Wahab. 2023. "Hydrological Risk Assessment for Mangla Dam: Compound Effects of Instant Flow and Precipitation Peaks under Climate Change, Using HEC-RAS and HEC-GeoRAS." SN Applied Sciences 5(12). https://doi.org/10.1007/s42452-023-05579-2.
3. Waseem Muhammad, Sareer Ahmad, Izhar Ahmad, Hira Wahab, and Megersa Kebede Leta. 2023. "Urban Flood Risk Assessment Using AHP and Geospatial Techniques in Swat Pakistan." SN Applied Sciences 5(8):215. https://doi.org/10.1007/s42452-023-05445-1.
4. Herbert, Z. C., Z. Asghar, and C. A. Oroza. 2021. "Long-term Reservoir Inflow Forecasts: Enhanced Water Supply and Inflow Volume Accuracy Using Deep Learning." Journal of Hydrology 601:126676. https://doi.org/10.1016/j.jhydrol.2021.126676.
5. Muqeet, M., et al. 2023. "Enhanced cellulose nanofiber mechanical stability through ionic crosslinking and interpretation of adsorption data using machine learning." International Journal of Biological Macromolecules 237:124180. https://doi.org/10.1016/j.ijbiomac.2023.124180.

6. Ahmad, S., et al. 2025. "Assessing water demand and supply in the Upper Indus Basin using integrated hydrological modeling under varied socioeconomic scenarios." Applied Water Science. https://doi.org/10.1007/s13201-024-02310-3.