Abstract—Groundwater is the largest storage of freshwater resources, which serves as the major inventory for most of the human consumption through agriculture, industrial, and domestic water supply. In the fields of hydrological, some researchers applied a neural network to forecast rainfall intensity in space-time and introduced the advantages of neural networks compared to numerical models. Then, many researches have been conducted applying data-driven models. Some of them extended an Artificial Neural Networks (ANNs) model to forecast groundwater level in semi-confined glacial sand and gravel aquifer under variable state, pumping extraction and climate conditions with significant accuracy. In this paper, a multi-layer perceptron is applied to simulate groundwater level. The adaptive moment estimation optimization algorithm is also used to this matter. The root mean squared error, mean absolute error, mean squared error and the coefficient of determination (R²) are used to evaluate the accuracy of the simulated groundwater level. Total value of R² and RMSE are 0.9458 and 0.7313 respectively which are obtained from the model output. Results indicate that deep learning algorithms can demonstrate a high accuracy prediction. Although the optimization of parameters is insignificant in numbers, but due to the value of time in modelling setup, it is highly recommended to apply an optimization algorithm in modelling.

Keywords—Hybrid deep learning model; Groundwater; MLP; ADAM

I. INTRODUCTION

Groundwater is the largest storage of freshwater resources, which serves as the major inventory for most of the human consumption through agriculture, industrial, and domestic water supply, [1-2]. With the growth of computer science and hardware, the data-driven methods (e.g., regression analysis, statistics, gray theory, machine learning) have been extensively studied and broadly used in lots of areas with good-enough results, [3-5]. In the fields of hydrological, reference [6] applied a neural network to forecast rainfall intensity in space-time and introduced the advantages of neural networks compared to numerical models. Then, many researches have been conducted applying data-driven models. Also, an Artificial Neural Networks (ANNs) model is extended to forecast groundwater level in semi-confined glacial sand and gravel aquifer under variable state, pumping extraction and climate conditions with significant accuracy, [7]. Reference [8] found Feedforward Neural Network (FNN) models to predict monthly and quarterly time-series water levels at a monitoring well with higher results than the models developed by statistical regression methods. Moreover, the capabilities of Auto-Regressive Moving Average, Adaptive Neuro-Fuzzy Inference System, ANN, Support Vector Machine and Gene Expression Programming are investigated for groundwater level forecast, [9]. Reference [10] used hybrid models by synthesizing Ensemble Empirical Mode Decomposition (EEMD) with data-driven models to forecast groundwater level fluctuations. Thus, Deep Learning (DL) algorithms have shown high potential in areas such as speech recognition, computer vision, natural language processing and intelligent recommendation. Reference [10] applied deep learning algorithms to enhance simulations of large-scale groundwater flow in IoTs. Also, groundwater storage is estimated from seismic data using deep learning, [11]. Reference [12] used Multi-Layer Perceptron and Long Short-Term Memory beside with one statistical regression model, Seasonal Autoregressive Integrated Moving Average to predict the model. Three approaches are compared for optimizing the models’ hyper parameters, including two surrogate model-based algorithms and also a random sampling method, [13]. Results indicated that the optimization of the hyper parameters leads to the reasonably more accurate performance of all models and forecast the groundwater level. Reference [14] used deep learning to predict groundwater depth fluctuations. As founded on this research, the deep learning model is an intelligent tool to forecast groundwater depths. In addition, advanced AI techniques can preserve resources and labor conventionally employed to estimate different features of complex groundwater systems. Multilayer...
Perceptron is used with Particle Swarm Optimization to Predict Landslide Susceptibility, [15]. The results demonstrated that the MLP-PSO model had the most accuracy in prediction. Also, the combination of MLP-PSO algorithm optimization increased the accuracy compared to the MLP model. Reference [16] applied Long Short-Term Memory (LSTM) deep neural network to reconstruct missing groundwater level data and also increase the time series length. An appropriate methodology is exerted to predict the water flow and water level parameters of the Shannon River in the long term based on deep learning. They employed a machine learning model based on deep convolutional neural networks (CNN). The proposed model predicted water flow and also water level from 2013–2080, [17]. Reference [18] studied the feasibility of executing deep learning emulators for contaminant transport models and a generalized flow based on the hydrogeology of an aquifer in South Australia. A data-driven model is employed to forecast groundwater levels in an arid basin using Artificial Neuron Networks (ANNs), Support Vector Machines (SVMs), and M5 Model Tree, [19]. Reference [20] compared nonlinear autoregressive networks with exogenous inputs (NARX) with popular state-of-the-art Deep Learning techniques such as long short-term memory (LSTM) and convolutional neural networks (CNN) to predict groundwater level. A multilayer perceptron (MLP) model, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model, and long short-term memory (LSTM) model are compared to predict slope movements. Results indicated that the moving-average models (SARIMA) performed better than the Machine Learning models (MLP and LSTM) during both training and test, [21].

In this study, a Multi-Layer Perceptron (MLP) is applied to simulate groundwater level. Also, the Adaptive moment estimation (Adam) optimization algorithm is used to this matter. Afterward, this hybrid optimal approach is applied to forecast the groundwater level in the Najafabad aquifer. Finally, to be precise about the accuracy, a comprehensive study of the presented model is presented and the obtained results are discussed.

II. METHODOLOGY

A. Case study

The case study aquifer is located in Najafabad plain, Gavkhoni catchment consists of 21 basins. Najafabad study area is located in the middle of this basin and in Isfahan province. The total area of this area is 1754.9 km², of which 679.2 km² are highlands and 1075.7 km² are plains. The alluvial aquifer is spread over an area of 941 km² (95% of the plain area). The highest point with an altitude of 2953 m is in the northeast and the minimum elevation of 2069 m above sea level is in the western north near the Zayandehrood River. Zayandehrood River enters from the southern part of this area and exits from the eastern part. The water table depth maps of the study region show a dominant northeast-southwest inclination. In the central parts of the plain, the high density of wells created bulls-eye contours and a cone of depression in water table maps (Figure 1). At the southwestern limits of the aquifer, contours align more or less in a northeast-southwest direction. The Najafabad sub-basin has a mainly semiarid climate. The average rainfall is only 155 mm/year, and most of the rainfall takes place in the winter from December to April. While during the summer, there is no effective rainfall. Furthermore, annual potential evapotranspiration is about 1.950 mm.

B. Modelling setup

First, by considering the hydrological conditions of the region, the effective parameters of the groundwater table are identified. Since the study area is one of the semi-arid regions of the country, according to studies, there is no correlation between data such as humidity, solar coefficients and other meteorological data with the groundwater table. Also, due to the seasonality of rivers in the region, river discharge data and infiltration are not considered. Finally, rainfall, temperature and water inputs are used to measure, model and predict the water table in the study area of Najafabad aquifer and aquifers with the same conditions. Modelling with this data is performed in 9 steps (Figure 2). After collecting the necessary inputs for modelling from regional water in Isfahan, the data have been sorted.

![Study area](image1)

Fig. 1. Study area

C. Multi-Layer Perceptron (MLP) model

Multi-Layer Perceptron (MLP) model is a feed-forward neural network that consists of an input layer, at least one hidden layer, and an output layer, [22]. It can model nonlinear patterns by introducing multiple layers and nonlinear activation functions that transform input signals. It has been primarily used for classification and function estimation problems, [23]. MLPs have been used for, e.g., forecasting drought, [24], estimating evapotranspiration, [25], and predicting water flow, [26]. Unlike other statistical models, the MLP does not make any assumptions regarding the prior probability density distribution of the training data, [27].
In an MLP network, all the neurons in the hidden and output layers use nonlinear activation functions to simulate the action potential of biological neurons. So, a Rectified Linear Unit (ReLU) function is applied for the activation function in all of the neurons using as:

\[
\text{ReLU}(x) = \begin{cases} 
  x & \text{if } x > 0 \\
  0 & \text{if } x \leq 0
\end{cases}
\]  

(1)

Compared with logistic sigmoid function and hyperbolic tan function, the ReLU has finer performance. It can solve the challenges of gradient explosion and gradient disappearance, and preserve the convergence rate in a stable state. Since the network is entirely connected, each node in one layer connects to every node in the next layer with a certain weight. Therefore, the output \( c \) of each neuron is:

\[
c = \varphi \left( \sum_i w_i a_i + b \right)
\]

(2)

where \( a_i \) and \( w_i \) are the inputs and weights of the neuron respectively, \( b \) is the bias of the current neuron and \( \varphi \) is the activation function ReLU.

**D. Adaptive Moment Estimation (Adam)**

Adaptive Moment Estimation (Adam) optimization is applied to calculate the adaptive training rate of the parameters, [28]. This method, in addition to storing the descending mean of the square of the past of the gradient or \( v_i \) and the mean of the descending of the past of the gradient or \( m_i \), is kept as a momentum. This is while the momentum can be seen as a ball sliding on a sloping surface that has no friction and therefore can be placed at the minimum error level, [29]. Both parameters of the mean descending average of the square of the gradient and the average of the descending of the past of the gradient can be calculated as follows:

\[
m_i = \beta_1 m_{i-1} + (1 - \beta_1) g_i
\]

\[
v_i = \beta_2 v_{i-1} + (1 - \beta_2) g_i^2
\]

(3)

\( m_i \) and \( v_i \) are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively, hence the name of the method. As \( m_i \) and \( v_i \) are initialized as vectors of zeros, the authors of Adam observe that they are biased towards zero, especially during the initial time steps, and especially when the decay rates are small (i.e. \( \beta_1 \) and \( \beta_2 \) are close to 1). They counteract these biases by computing bias-corrected first and second moment estimates:

\[
\hat{m}_i = \frac{m_i}{1 - \beta_1^i}
\]

\[
\hat{v}_i = \frac{v_i}{1 - \beta_2^i}
\]

(4)

They then use these to update the parameters just as we have seen in Adadelta and RMSprop, which yields the Adam update rule:

\[
\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_i} + \varepsilon} \hat{m}_i
\]

(5)

Default values of 0.9 for \( \beta_1 \), 0.999 for \( \beta_2 \), and 10\(^{-8}\) for \( \varepsilon \) has been modified. They show empirically that Adam works well in practice and compares favorably to other adaptive learning-method algorithms.

**III. RESULTS AND DISCUSSION**

The settings of the multi-layer perceptron model are based on the active function of Relu. In this research, network training is based on 80% of available data and testing is based on the remaining 20%. The last column of the model, which represents the groundwater level or output of the model, is shown in Figure (3). The information in this column is obtained from the impact level of each well in the aquifer by a mathematical formula within GIS software. On the other hand, we assume that the inputs of each model include temperature, precipitation and weighted average of groundwater level (in order to reduce the number of features and since the features of reservoirs are one-dimensional, weighted average is used and their output is groundwater level.

In this research, the Python software is used to study the simulations. So, there are many benefits to use a Python environment such as the availability of tables and graphs and many other steps applied in modeling, called libraries. Different libraries are used for various purposes, which will be described. The Pandas library is used to work with *xlsx and *.xls input formats in modeling. The Numpy, Sklearn, and Tensorflow.keras libraries are employed for the modeling section, and the Matplotlib libraries are exerted to display the
output in the form of a graph (Table 1). Adjustment of hyperparameter coefficients is mentioned below. Due to the vast application in various modeling fields, 500 neurons in the hidden layer are applied in the configuration of the multilayer perceptron model (Table 2). Also, the active function of this model is a Rectified linear unit. Moreover, the input layer has 100 nodes, which corresponds to the 100 potential distinguished predictors. There are 500 neurons in the hidden layer. All of the weights and biases of the linkages among the potential variables and groundwater recharge values are improved when the parameters of the network are learned based on the training data. Then the predicted values are computed based on the learned weights and biases when the input data is applied. The output layer only had one neuron, which represented the values of groundwater recharge.

![Original Data](image)

**Fig. 3.** Water level measured in 175 months

| Libraries Used | Specifications |
|----------------|-----------------|
| Pandas         | Dataframes/importing CSV files |
| Numpy          | Scientific Computing |
| Sklearn        | Model implementation |
| Matplotlib     | Plotting of numerical mathematics |
| Tensorflow.keras | High-level Neural Network APIs |

**TABLE II.**多层感知器模型配置

| Model summary | Neurons | Activation | Optimizer | inp_size | verbose |
|---------------|---------|------------|-----------|----------|---------|
| MLP           | 500     | ReLu       | ‘adam’    | 1        | False   |

The results can be studied by comparing the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and coefficient of determination (R²) as shown in Table (3). As stated before, network training is based on 80% of available data, and testing is based on the remaining 20%. The last column of the data, which represents the groundwater level or output of the model, is shown in Figure (3). The information in this column is obtained from the impact level of each well in the aquifer by a mathematical formula within GIS software. On the other hand, the inputs of each model are considered as temperature, precipitation, and weighted average of groundwater level (in order to reduce the number of features, the weighted average is used to obtain the one-dimensional features of wells) and their output is groundwater level. The modeling is performed using a Multi-Layer Perceptron (MLP) model learning simulator in Python. As you can see in Figure (4), the green line represents the MLP model, which is also comparable to the original data.

**TABLE III.** 多层感知器模型输出

| RMSE     | MAE    | MSE    | R²     | Amount | Model   |
|----------|--------|--------|--------|--------|---------|
| 0.7308   | 0.5564 | 0.5341 | 0.9436 | 80%    | Train   |
| 0.7337   | 0.5632 | 0.5383 | 0.9517 | 20%    | Test    |
| 0.7313   | 0.5578 | 0.5349 | 0.9458 | 100%   | Total   |

**Fig. 4.** Model simulation with multi-layer perceptron

**IV. CONCLUSION**

Modern time-series modeling and forecasts are mostly consisting of artificial intelligence models. In this study, a deep learning structure was conducted to estimate groundwater level. To this end, a Multi-Layer Perceptron (MLP) combining an Adam optimization is exerted. The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R²) were used to evaluate the accuracy of the simulated groundwater level. The total value of R² and RMSE are 0.9458 and 0.7313 respectively which were obtained from the model output. Results indicated that deep learning algorithms can demonstrate a high accuracy prediction. Although the optimization of parameters was insignificant in numbers, due to the value of time in modelling setup it is highly recommended to apply an optimization algorithm in modeling.
REFERENCES

[1] Famiglietti, J. S., 2014. The global groundwater crisis. Nature Climate Change, 4(11), pp.945-948.
[2] Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Lebanc, M., Famiglietti, J.S., Edmunds, M. and Konikow, L., 2013. Ground water and climate change. Nature climate change, 3(4), pp.322-329.
[3] DeVries, Pheobe M. R., Fernanda Viégas, Martin Wattenberg, and Brendan J. Meade. 2018. Deep learning of aftershock patterns following large earthquakes. Nature, 7720, pp. 632-634.
[4] Chen, C., He, W., Zhou, H., Xue, Y. and Zhu, M., 2020. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Helhe River Basin, northwestern China. Scientific reports, 10(1), pp.1-13.
[5] Zhou, H., Chen, C., Liu, H., Qin, F. and Liang, H., 2019. October. Proactive Knowledge-Driven Dialogue System on Pointer Network. In CCF International Conference on Natural Language Processing and Chinese Computing, pp. 724-735.
[6] French, M. N., Krajewski, W. F. and Cuykendall, R. R., 1992. Rainfall forecasting in space and time using a neural network. Journal of hydrology, 137(1-4), pp.1-31.
[7] Coppola Jr, E. A., Rana, A. J., Poulton, M. M., Szidarovszky, F. and Uhl, V. W., 2005. A neural network model for predicting aquifer water level elevations. Groundwater, 43(2), pp. 231-241.
[8] Uddamerr, V., 2007. Using statistical and artificial neural network models to forecast potentiometric levels at a deep well in South Texas. Environmental Geology, 51(6), pp. 885-895.
[9] Shiri, J., Kisi, O., Yoon, H., Lee, K. K. and Nazemi, A. H., 2013. Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. Computers & Geosciences, 56, pp.42-44.
[10] Gong, Y., Wang, Z., Xu, G. and Zhang, Z., 2018. A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. Water, 10(6), pp. 730.
[11] Lăhişavara, T., Malehmır, A., Pasanen, A., Kärkkäinen, L., Huttunen, J. M. and Hesthaven, J. S., 2019. Estimation of groundwater storage from seismic data using deep learning. Geophysical Prospecting, 67(8), pp. 2115-2126.
[12] Kochhar, A., Singh, H., Sahoo, S., Litoria, P. K. and Pateriya, B., 2021. Prediction and forecast of pre-monsoon and post-monsoon groundwater level: using deep learning and statistical modelling. Modelling Earth Systems and Environment, pp. 1-13.
[13] Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Fabyishenko, B. and Agarwal, D., 2020. Surrogate optimization of deep neural networks for groundwater predictions. Journal of Global Optimization, pp. 1-29.
[14] Kumar, D., Roshni, T., Singh, A., Jha, M. K. and Samui, P., 2020. Predicting groundwater depth fluctuations using deep learning, extreme learning machine and Gaussian process: a comparative study. Earth Science Informatics, 13(4), pp. 1237-1250.
[15] Li, D., Huang, F., Yan, L., Cao, Z., Chen, J. and Ye, Z., 2019. Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, bp neural network, and information value models. Applied Sciences, 9(18), pp. 3664.
[16] Yu, M. T., Jardani, A., Massei, N. and Fournier, M., 2021. Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. Journal of Hydrology, 597, pp. 1257-1276.
[17] Ali, Z., Hussain, I., Faisal, M., Nazir, H. M., Hussain, T., Shad, M. Y., Mohamd Shoukry, A. and Hussain Gani, S., 2017. Forecasting drought using multilayer perceptron artificial neural network model. Advances in Meteorology.
[18] Yu, X., Cui, T., Sreekanth, J., Mangeon, S., Doble, R., Xin, P., Rassam, D. and Gilfedder, M., 2020. Deep learning emulators for groundwater contaminant transport modelling. Journal of Hydrology, 590, pp. 125-141.
[19] Huang, M. and Tian, Y., 2015. October. Prediction of groundwater level for sustainable water management in an arid basin using datadriven models. In International Conference on Sustainable Energy and Environmental Engineering (SEE), Bangkok.
[20] Wunsch, A., Liesch, T. and Broda, S., 2020. Groundwater Level Forecasting with Artificial Neural Networks: A Comparison of LSTM, CNN and NARX. Hydrology and Earth System Sciences Discussions, 1-23.
[21] Kumar, P., Pathania, A., Agarwal, S., Mali, N., Singh, R., Chaturvedi, P., Uday, K. V. and Dutt, V., 2020. Predictions of weekly slope movements using moving-average and neural network methods: a case study in Chamoli, India. In Soft Computing for Problem Solving, pp. 67-81.
[22] Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford university press.
[23] Trenn, S., 2008. Multilayer perceptrons: Approximation order and necessary number of hidden units. IEEE transactions on neural networks, 19(5), pp. 836-844.
[24] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30.
[25] Tabari, H. and Talaee, P. H., 2013. Multilayer perceptron for reference evapotranspiration estimation in a semiarid region. Neural Computing and Applications, 23(2), pp. 341-348.
[26] Araujo, P., Astry, G., Ferrerio-Lage, J. A., Mejuto, J.C., Rodriguez-Suarez, J. A. and Soto, B., 2011. Multilayer perceptron neural network for flow prediction. Journal of Environmental Monitoring, 13(1), pp. 35-41.
[27] Gardner, M. W. and Dorling, S. R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric environment, 32(14-15), pp. 2627-2636.
[28] Kingma, D. P., and Ba, J. L. 2015. Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, pp. 1–13.
[29] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In Advances in Neural Information Processing Systems, 30.