Predicting aquatic ecosystem health using machine learning algorithms

Dr. Priya Vij[1*]; Patil Manisha Prashant[2]

Received: 16 August 2024; Revised: 16 September 2024; Accepted: 11 November 2024; Published: 10 December 2024

Abstract

The most instinctive natural resource on the planet is water. Hydrological researchers must take the vital step of predicting water level and taking the necessary action to prevent the impending water crisis in order to avoid the shortage of water. Several hydrological research have demonstrated the potential of Artificial Neural Networks (ANN) as a new technology for anticipating groundwater levels. The extended feature and flawed data set for prediction result in an unpredictable and inconsistent aspect. To prevent a water crisis, strict groundwater management measures must be implemented. Artificial Neural Networks (ANN) are effective in three areas, according to recent research: they can handle very huge data sets, complex computing challenges, and training at the discrete level of representation or depiction. Strong and fast hardware is desired for deep learning stimulation. A single-core CPU will not be able to create reliable predictions with a flawed and inaccurate data set of input variables; on the other hand, a multi-core CPU will make predictions more efficiently. A multi-processor system with thousands of computer units called cores is called a Graphical Processing Unit (GPU).

Index terms: Aquatic ecosystem, Health, Machine learning

1*- Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India. Email: ku.priyavij@kalingauniversity.ac.in, ORCID: https://orcid.org/0009-0005-4629-3413
2- Research Scholar, Department of CS & IT, Kalinga University, Raipur, India. Email: patil.manisha@kalingauniversity.ac.in, ORCID: https://orcid.org/0009-0003-1140-0261
*Corresponding author
DOI: 10.70102/IJARES/V4S1/7

Background

Water is one of the most vital natural resources for the continuation of life on Earth. A precise water forecast will assist the groundwater management authority in making conservation decisions, and a long-term forecast will confirm seasonal and cyclical variations in groundwater levels (Uddin *et al.,* 2024). Aside from the fact that moving water from a lake to a water basin has a significant impact, variations in water level are a complex anomaly for hydrological situations with several constrained characteristics, such as climate variables. Nevertheless, the accurate presentation of the aquatic environment and the model's ability to detect variations in measurement are necessary for the method-based approach to be effective (Deng, Chau and Duan, 2021). These methods are quite slow and typically need a lot of time. Numerous research employing various statistical methods have been proposed. In the current water crisis, accurate and consistent groundwater level forecasting is essential. Future forecasting helps the water management authority plan strategies and make the right judgements in this regard by providing valuable information. In addition to helping scientists create essential tools for improving science, deep learning models have been utilised to tackle significant transdisciplinary problems. This study demonstrates how deep learning models are increasingly being used in water studies (Zhu *et al.,* 2022). Data from previous studies served as the foundation for the research methodology employed in this study. The future can be predicted using the hydrological research and records that are already available. Time-series analysis is one of the most widely utilised forecasting techniques. The future can be predicted using this approach, which uses statistical analysis of multiple parametric parameters (Jaiswal and Pradhan, 2023). The variables under consideration must have a time interval between their values. This inquiry is carried out under the assumption that past interactions will recur in the future. Time-series analysis just requires a few time-related components. There are several different ways to forecast time-series data (Gambín *et al.*, 2021).

Analysis And Design

The process of extracting relevant information from the raw data collected from the source and supplying it to deep learning models is known as preprocessing (Arrighi and Castelli, 2023). It refers to updating, managing, or removing data in order to enhance overall performance. It's not always the case that we receive the data in the format we want; in order to perform any kind of computation, the data must be formatted or cleaned in accordance with the model (Kwon, Kang and Pyo, 2024). Data must be cleaned and formatted since real-world data frequently contains errors, noise, and missing information.

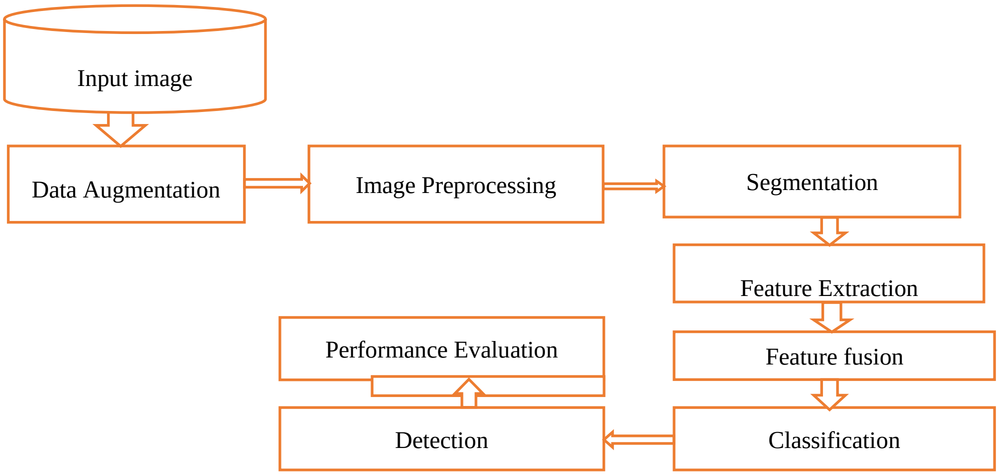


Figure 1: Block diagram of proposed model.

For this reason, data is pre-processed and cleaned before training, which also helps to increase efficiency and produce more accurate results (Figure 1). For any real-world problem, evaluating a set of data that has not been properly cleansed or separated will yield erroneous findings. Therefore, the first and most important step before any analysis is to feature the nature of the data (Lee *et al.*, 2021). The most important phase in deep learning techniques is pre-processing. If there is a lot of noise in the dataset during preparation, the prediction results could be skewed by duplicate or incorrect data. Parallelism does not always solve issues, but in several domains, such as artificial intelligence (AI), image detection, and ANN. The GPU calculates the distance between each layer by multiplying the input and weight matrices (Uddin *et al.*, 2022). The distance calculation is done by a single thread running in a single kernel. Assuming that the network had two neurones for every 100 inputs, there would be 200 threads in total, each of which would output the neuron's output. In a single block, this algorithm runs the same amount of threads. The number of neurones in the hidden layer is equal to the number of blocks that are displayed. Copying the findings into an array is the second task. Ultimately, the GPU's output is sent back to the main memory. A weighted sum is distributed among the available threads using the synaptic parallel approach. On the other hand, the pattern combines the parallel synapse and neuronal methods. Finding the factors that influence a groundwater prediction's performance is the first stage.

A number of ANN's characteristics make this method a desirable substitute for a variety of challenging issues. The two main ways to employ neural nets are to use an expert system or to apply an algorithmic approach to problem-solving. The main advantage is that it allows computers to make wise judgements for people with little to no human intervention. When there is sufficient information available regarding the data and model, the analytical or algorithmic technique is derived.

Requirement Resources and Results

The new model will be simulated using a single NVIDIA GeForce GTX 1080 Ti GPU 2760 4MB and an Intel Core i7-6850K 3.60 GHz 12-core processor. The deep learning workstation platform is equipped with the Python simulation

DOI: 10.70102/IJARES/V4S1/7

framework and Windows 10 operating system, which have been used to evaluate the proposed system's performance in a variety of scenarios in figure 2.

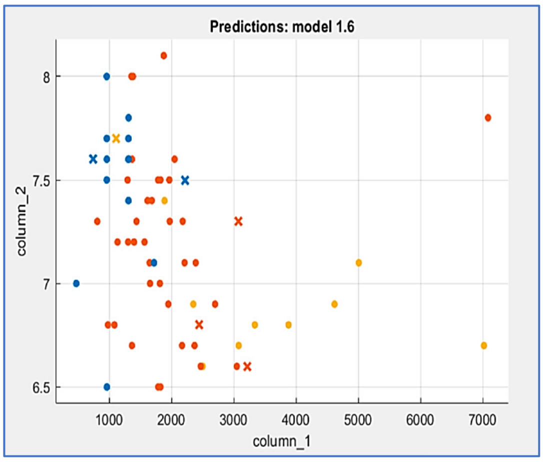
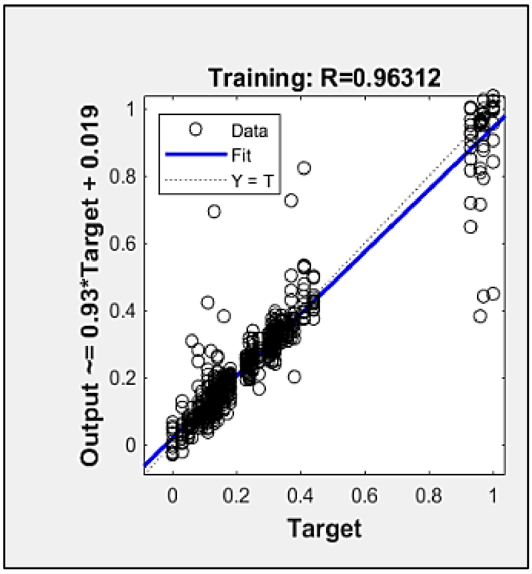
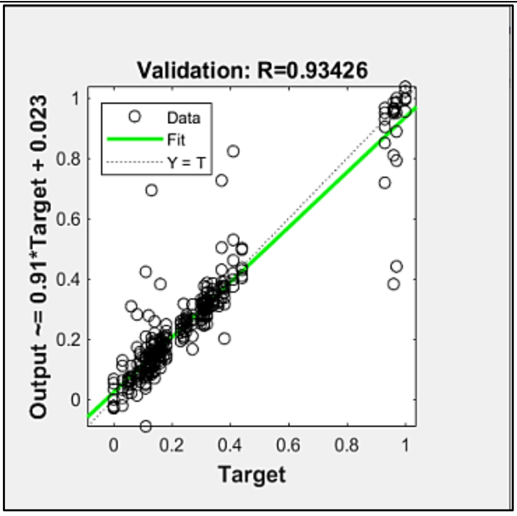
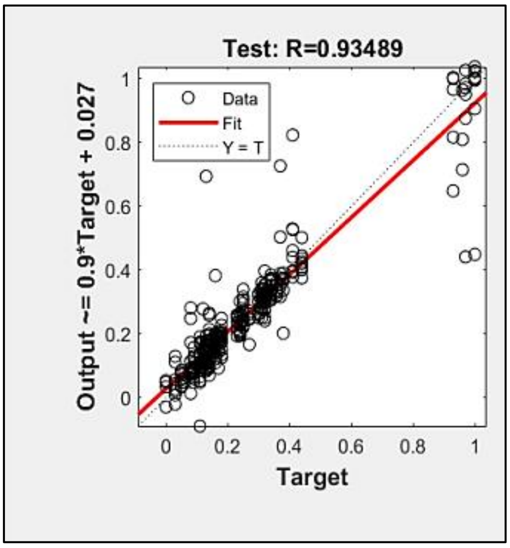


Figure 2: Water quality index.

Expert systems, on the other hand, can be used to solve situations where there is not enough data, analytical background, or relationship to produce a definitive model. When there is a lack of foundational theory and lacking data, ANNs are suitable. Without a particular training technique, neural networks generalised the disorganised data and assimilated clear outputs.







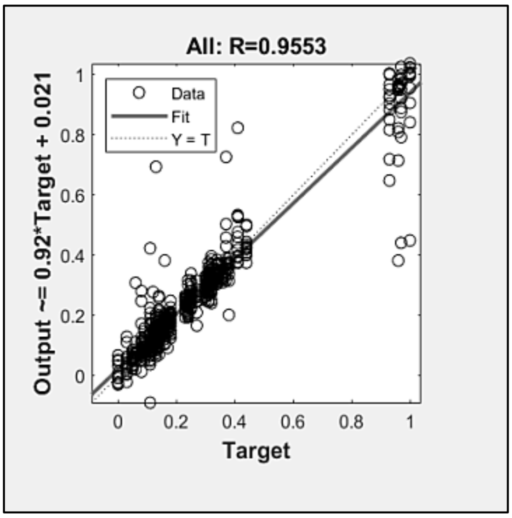


Figure 3: Regression plot.

When data emerges from thorough investigation, it may be nonlinear, non-stationary, or disorganised, making design challenging. In computer vision, where the machine can extract information from photos and movies, neural networks are helpful. Computer vision includes the identification and recognition of images, faces, and visual stimuli. Neural networks are also crucial in the other two primary fields of speech and natural language processing in figure 3.

Conclusions

The Parallelised BPN algorithm has been used for all learning, training, and validation on CUDA. Prediction findings show improved performance with a low execution time, error rate, and high correlation coefficient when using a parallelised algorithm on a GPU. Although the results of the week and ten-week lead error models are satisfactory, errors begin to increase after that. A correlation coefficient of 0.96337 indicates a promising match between the observed and expected output. In Faridabad, a relationship between ground water level and several influencing factors including temperature, precipitation, and water level was examined. Based on the results, it can be concluded that the parallelised algorithm demonstrated its ability to estimate groundwater levels at six observation wells using models with one- and ten-week lead times. According to this study, groundwater levels are trending downward by 0–2 meters year.

References

Arrighi, C. and Castelli, F., 2023. Prediction of ecological status of surface water bodies with supervised machine learning classifiers. *Science of the Total Environment*, *857*, p.159655. https://doi.org/10.1016/j.scitotenv.2022.159655

Deng, T., Chau, K.W. and Duan, H.F., 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. *Journal of Environmental Management*, *284*, p.112051. https://doi.org/10.1016/j.jenvman.2021.112051

Gambín, Á.F., Angelats, E., González, J.S., Miozzo, M. and Dini, P., 2021. Sustainable marine ecosystems: Deep learning for water quality assessment and forecasting. *IEEE access*, *9*, pp.121344-121365. https://doi.org/10.1109/ACCESS.2021.3109216

Jaiswal, H. and Pradhan, S. 2023. The Economic Significance of Ecosystem Services in Urban Areas for Developing Nations. *Aquatic Ecosystems and Environmental Frontiers*, *1*(1), 1-5.

Kwon, Y.S., Kang, H. and Pyo, J., 2024. Estimation of aquatic ecosystem health using deep neural network with nonlinear data mapping. *Ecological Informatics*, *81*, p.102588. https://doi.org/10.1016/j.ecoinf.2024.102588

Lee, S., Kim, J., Lee, G., Hong, J., Bae, J.H. and Lim, K.J., 2021. Prediction of aquatic ecosystem health indices through machine learning models using the WGAN-based data augmentation method. *Sustainability*, *13*(18), p.10435. https://doi.org/10.3390/su131810435

Uddin, M.G., Nash, S., Diganta, M.T.M., Rahman, A. and Olbert,

A.I., 2022. Robust machine learning algorithms for predicting coastal water quality index. *Journal of Environmental Management*, *321*, p.115923. https://doi.org/10.1016/j.jenvman.2022.115923

Uddin, M.G., Nash, S., Rahman, A., Dabrowski, T. and Olbert, A.I., 2024. Data-driven modelling for assessing trophic status in marine ecosystems using machine learning approaches. *Environmental Research*, *242*, p.117755. https://doi.org/10.1016/j.envres.2023.117755

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B. and Ye, L., 2022. A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*, *1*(2), pp.107-116. https://doi.org/10.1016/j.eehl.2022.06.001