CLASSIFICATION OF NITRATE CHEMICAL CONCENTRATION IN GROUND WATER USING CLOUD BASED ARTIFICIAL INTELLIGENCE

D. Leela Dharani[a], Jagadeesan.S[b], N.Bindu Madhavi[c], Senthilkumar N C[d], M. Sunil Kumar[e], K.Prabhu[f]

Article History: Received: 10.09.2022 **Revised:** 15.10.2022 **Accepted:** 05.11.2022

Abstract: In order to ensure effective management and the prevention of groundwater contamination within the watershed, it is required to conduct an accurate susceptibility study. Because it is a resource that is necessary for human existence, the creation of goods from agriculture, and the operation of machinery in industrial settings, groundwater needs to be adequately monitored and preserved for future use. Therefore, it is of the utmost importance that an accurate evaluation of the groundwater contamination vulnerability index be performed using ResNet-18. The groundwater vulnerability index is a helpful tool for analyzing the state of the environment in various parts of the world, and the ResNet-18 model is able to lend a hand in the effort to accomplish this goal. The research indicated that ResNet-18 was a powerful tool for improving the evaluation of groundwater contamination vulnerability, and that it may help in ensuring the safety of the environment by reducing the likelihood of contamination. These findings were presented in the form of conclusions that were drawn from the research. This was demonstrated by the fact that it assisted in improving the assessment of the potential for contaminating groundwater supplies.

Keywords: Chemical, Nitrate, ResNet, Deep Learning

[a]. Assistant Professor, Department of Information Technology, P.V.P Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India.
[b]. Assistant Professor (Senior Grade), School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[c]. Associate Professor, Centre for Distance & Online Education, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India.
[d]. Associate Professor, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[e]. Professor & Programme Head, Department of Computer Science and Engineering, School of Computing, Mohan Babu University (erstwhile Sree Vidyanikethan Engineering College), Tirupathi, Andhra Pradesh, India
[f]. Associate Professor, Department of Civil Engineering, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India.

***Corresponding Author**

E-mail: dharanidonepudi@gmail.com

DOI: 10.31838/ecb/2022.11.11.010

INTRODUCTION

It is very necessary, in order to preserve its cultural and ecological richness, to have a grasp of the mechanisms that regulate and maintain its natural surrounds. This is because this understanding is what allows for the preservation of that diversity. In spite of the immense renown it has received, there has not yet been a complete hydrogeological collection or assessment that has been produced across a wide variety of geographical areas. Because of this, it is becoming increasingly essential that we make use of scientific study in order to gain a deeper understanding of the processes that are at play in the natural world [1].

Groundwater is an attractive solution for the distribution of residential water in big cities, towns, and villages as, in comparison to surface water sources, it has a cheaper starting cost, greater water quality, and availability all year round. Groundwater that is contaminated with nitrate chemicals is becoming more prevalent not only in local communities but also on a worldwide scale [1]. A study conducted on the state of the water supply in the countries of Sub-Saharan Africa came to the conclusion that shallow wells, particularly those located in densely populated metropolitan areas, were more likely to be contaminated by waste from homes. A key cause for concern regarding the quality of Ethiopia groundwater is the presence of high concentrations of microorganisms and nitrate chemicals in shallow, unconfined aquifers located in close proximity to large communities [3].

According to the conclusions of a number of studies, groundwater will continue to function as the primary source of fresh water in spite of changes and variations in the climate. The aquifer and the surface water sources that it feeds are tapped for groundwater extraction in order to provide water for drinking and other household uses [4]. Despite this, the contamination of groundwater has greatly worsened as a result of population pressure and the irresponsible use of water, neither of which has taken into mind the continued existence of ecosystem services in the long run. This is a challenge that can't be disregarded in any way. A number of different variables, including shifts in hydrogeology and actions carried out by humans, are responsible for the poisoning of groundwater [5].

This high level of contamination has been caused by a variety of factors, including those listed above. A first assessment on the quality of the drinking water in Ethiopia and discovered that 32 percent of the wells had measurable amounts of nitrate chemical contamination. In addition, multiple studies have demonstrated that safeguarding groundwater resources is not only a prudent strategy, but also one that is one that is substantially cheaper than cleaning up and rebuilding a polluted aquifer. This is something that has been shown to be the case [6].

The majority of the water that is extracted from the ground and from springs is what is utilized for residential purposes. The concentration of nitrate chemical in the shallow groundwater wells of Bahir Dar City was high, but it showed a decreasing tendency as one proceeded from the city centre to the border. Either the high population density in the city, which is caused by humans, or the differences in the local geology could be the source of this phenomenon [7]. Both of these factors are viable explanations. Two procedures that are vital to groundwater and environmental management are the establishment of a vulnerability map and an assessment of the risk of groundwater pollution. Both of these steps should be completed as soon as possible. Because of these processes, administrators will be able to make decisions that are based on reliable information, which is an important reason why these steps are necessary [8].

There is a wide range of ways that can be utilized in the process of estimating the threat that is posed to groundwater. Some of these methods include process-based simulation models, statistical approaches, and overlay-index methods. In order to successfully complete groundwater risk assessments, it is common practice to make use of overlay index methodologies like DRASTIC [9].

These aspects are listed in the order from most likely to least likely. In order to measure the quantities of bacteria and other physicochemical pollutants in consumer water samples. An examination into the effect that human waste has on the natural environment was carried out. The quality of the ground water and surface water resources and the surrounding peri-urban areas was evaluated, who carried out the assessment. On the other hand, the regional disparities in groundwater pollution induced by hydrogeological and anthropogenic variables are not well documented [10]-[15].

The primary objective of this research was to make use of a modified version of the model in order to determine the degree to which the groundwater is susceptible to being contaminated by nitrate chemicals. This was done in order to determine the degree to which the groundwater is susceptible to being contaminated by nitrate chemicals. To be more specific, our goals were to make maps of intrinsic and particular vulnerability, learn more about where nitrate chemical contamination was happening, and change the Vulnerability Index to take into account the number of nitrate chemicals that we found. All three of these goals were to be accomplished.

PROPOSED METHOD

To create the model for recognizing photographs of nitrate chemical that we are presenting to you here, we made use of a deep residual network that was given the name ResNet-50.

2.1. ResNet-50

Over the course of the previous few years, numerous proposals for backbone networks of exceptionally high quality have been made. ResNet residual structure makes it possible for it to prevent gradient growth and disappearance as more network layers are added. These are two problems that could arise as a result of adding more network levels. It is one of the feature-extraction foundations that is utilized the most and is suited for application in convolutional neural networks (CNNs).

Table 1 outlines the many internal components that make up ResNet-50, along with their respective compositions. In order to make room for a convolution layer (conv1) and a max pooling approach with a stride size of 2, the input image is first scaled to 224 × 224 pixels. After this step, the convolution layer and the max pooling technique are applied. After that, there are a total of four residual layers that are utilized for the feature extraction process (Layer 1, Layer 2, Layer 3, and Layer 4). Following that, we will use a method called global average pooling (GAP) to get a feature that is 1 x 1 x 2048, then we will flatten it and input it into a complete convolutional (FC) layer. The probability of the various types of nitrate chemicals are what are produced as a result of this procedure.

Table 1. ResNet-50 Network structure

Layer	Output Size
conv1	112×112×64
-	56×56×64
Layer1	56×56×256
Layer2	28×28×512
Layer3	14×14×1024
Layer4	7×7×2048
-	1×1×2048
FC	Num_classes

Feature Fusion

The hierarchical structure of the features retrieved by a convolution neural network can be attributed, in part, to the nature of the convolution operation as well as the structure of the network itself. Therefore, the information used by the lower layers, such as colour and border, is quite simple. The features that are collected will be more advanced and more class-specific if there are more layers than there are currently in the model. When organizing items into categories, such as those that may be found in the ImageNet collection, it is frequently important to make use of abstract, high-level attributes.

Nitrate chemicals, on the other hand, come in a wide range of colours, crystal forms, hardness and chemical properties of nitrate chemicals; all of these aspects are intuitively represented in the photographs of nitrate chemicals by their varied hues, outlines, transparency levels, and surface textures. Nitrate chemicals come in a wide range of colours, crystal forms, hardness, and lustre due to the diverse chemical Nitrate chemicals can be discovered in a number of different habitats, such as the soil, the water, and the air. Since this is the case, the approaches that are currently being used to identify nitrate chemicals still have room for development because they do not take into account low-level characteristics such as colours, forms, and textures.

In this study, we present a strategy for the recognition of nitrate chemicals that combines characteristics in order to improve the performance of features, which, in turn, leads to an improvement in the accuracy of recognition. This strategy combines low-level and high-level characteristics in order to increase the performance of features, which in turn, leads to an improvement in the accuracy of recognition.

In particular, ResNet-50 is used to extract four layers of features from the ImageNet dataset from nitrate chemical images. These features come from the ImageNet dataset. Layers 3 and 4 are the ones that provide the high-level features, with the features produced by Layer 4 being the ones that are commonly used for classification purposes due to the fact that they have the largest receptive area and the semantic information.

The high-level features are provided by the layers that provide Layers 3 and 4. It is standard procedure to assign a quality rating that is lower than average to the features that are present at the Layer1 and Layer2 levels. GAP needs to be performed beforehand in order to scale the features to a constant height and breadth in order for feature fusion to be possible. After that, the high-level (FH) and the low-level features (FL) are concatenated together using the method that is used to create the fused features.

Ffused=[FH,FL].

Therefore, the merged features consist of a wide variety of high-level and low-level details, as well as semantic information at a higher level. In the very last, but certainly not least, the model output that was produced by the fusion features is retrieved with the assistance of the Full Connection (FC) layer. This is the appearance of the model that is formed when the features of Layer2 and Layer4 are combined, as seen in the image in Figure 1.

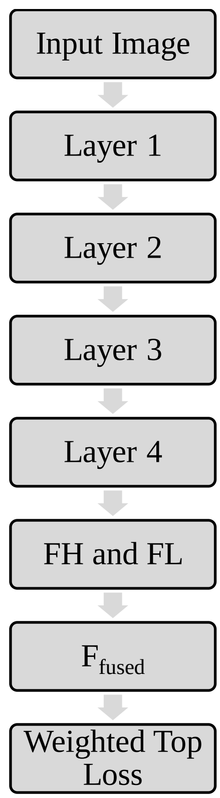

Figure 1. Feature Fusion Recognition

Loss Function

The loss function elucidates the objectives that the process of network learning is attempting to realize in order to be successful. It is common for there to be a significant difference between the proportion of simple samples, which exhibit the unmistakable properties of the nitrate chemicals, and the proportion of difficult samples, in which the features are displayed in a manner that is unclear either as a result of insufficient imaging or as a result of the nitrate chemical own inherently concealed properties. This is because it is typical for there to be a significant disparity between the proportion of simple samples, which exhibit the unmistakable.

In most cases, a significantly higher percentage of the samples are simple. When the network is being trained, it picks up on the obvious characteristics rapidly from the uncomplicated samples. However, the mining of features from the challenging data is neglected since their weights in the total loss are so low. This is because the weights are based on the total loss. Once a network reaches a certain level of learning, the increase in its performance is capped because the existing loss functions are unable to require the network to learn the implicit information that is included in the hard samples. This means that once a network reaches this level of learning, the increase in its performance is capped. As a consequence of this, an increase in a network performance can no longer be improved upon once it has reached the level of learning.

Some nitrate chemicals display significant intra-class variances but only little inter-class differences, which can lead to confusion as a result of the subtle differences that exist between the two groups of nitrate chemicals. As a consequence of this, giving more weight to these challenging data sets has the potential to bring about improvements in the accuracy of the model. Zhang proposes using top-k loss as a solution to the issue of OHSM that manifests itself throughout the process of online face identification. The selection of challenging samples with high loss values as training samples is a crucial part of the OHEM technique. This is done in order to identify the parameters of the network.

In this work, we make an effort to incorporate top-k loss, which shall henceforth be referred to as Losstop-k, into our investigation of nitrate chemical image recognition. This will be done in order to better understand the results of our investigation. The softmax loss for the ith sample in a batch is denoted by Li and may be written as,

$$L_i = -\log \frac{\left(e^{a_i}\right)}{\sum_{j=1}^{class} e^{a_j}}$$

$$Loss = \frac{1}{N}\sum_{i=1}^{N} L_i$$

$$Sort\left(L_i\right) = \left\{L_1', \ldots, L_N'\right\}$$

where ai is the number of calves in the first vector indicating the model output for the ith sample. Li indicates the softmax loss for the ith sample in a batch. Li is the notation used to indicate the soft maximum loss for the thirty-first sample in a batch.

According to equation (2), the softmax loss is the total number of losses that take place inside the batch, where N - total number of items in the batch. correspondingly, when the integers are put in ascending order, we write Li′. In Equation 4, where k is a percentage, the value of Losstop-k is the kN loss average values that are largest.

$$Loss_{topk} = \frac{1}{k \times N}\sum_{i=1}^{k\times N} L_i$$

The only loss values that are employed in the top-k loss strategy to determine the values of the training batch learning parameters are the ones that come from the top-kN examples, which are also referred to as the hard samples. These loss values are the only ones that are used. Utilizing backward propagation in order to derive gradients from this challenging data is one way to make the network more efficient, which paves the way for improving its overall performance. Because of this, the network will be able to assign more weight to hard data and unearth additional implicit information in a manner that is both more effective and efficient.

The performance of the network has been significantly enhanced, and it is now able to make more effective use of all of the data that is already at its disposal. The first kxN loss values in this function indicate tough samples, whereas the latter NkxN loss values indicate straightforward data. Within the context of the equation, the interval [0,1] is used to represent the weight coefficient. When =1, the top-k loss is equal to the softmax loss, and when =0, it is equal to the top-k loss. However, when =2, it is equal to the weighted top-k loss.

$$Loss_{topk_w} = \frac{1}{k \times N}\sum_{i=1}^{k\times N} L_i' + \alpha \frac{1}{\left(1-k\right)\times N}\sum_{i=k\times N+1}^{N} L_i'$$

Performance Evaluation

It is possible for the findings of the evaluation to be influenced due to the subjective nature of the modified model. This holds true for whatever kind of grading or weighting system you can think of. In order to validate the influence that the parameter has assessment, the purpose of the sensitivity analysis is to analyze the data. It is common practice in the field of groundwater vulnerability assessment to do a sensitivity analysis on a single parameter.

This is done so that a better knowledge of the effect of subjectivity and weights assigned to each parameter in the assessment may be gained. In order to investigate the effect that each variable has on the vulnerability index, we opted for and made use of the parameter analysis for the purpose of this inquiry. On the basis of Equation 7, the following is an explanation of how to carry out a single parameter sensitivity analysis (SPSA):

W=(PrPw)/V×100

where

W – overall parameter weight,

Pr - rating value

Pw – individual parameter weight

V - overall vulnerability index.

Map removal sensitivity analysis

A sensitivity analysis of map removal is what is used to exhibit the level of sensitivity that is associated with leaving out one or more maps, and it is what is used to illustrate this level of sensitivity. It was necessary to conduct a map-removal-sensitive analysis in order to determine whether or not the model included all of the essential components. The equation for the map removal sensitivity measure is shown below in (Eq. 6), and it is given in terms of the variation index S.

S=(VN-1−V‘n-1)/ V×100

where

S - sensitivity index,

V` - perturbed index, and

V - unperturbed index,

N - data layers to compute V

n - data layers to compute V`.

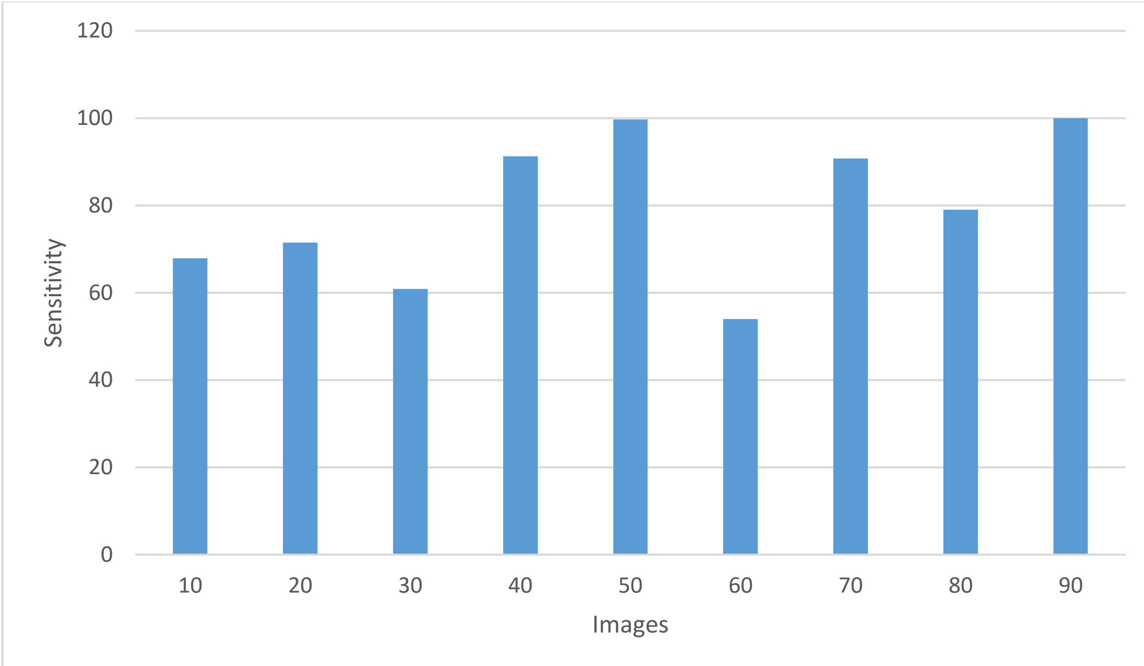

Figure 2: Single Parameter Sensitivity

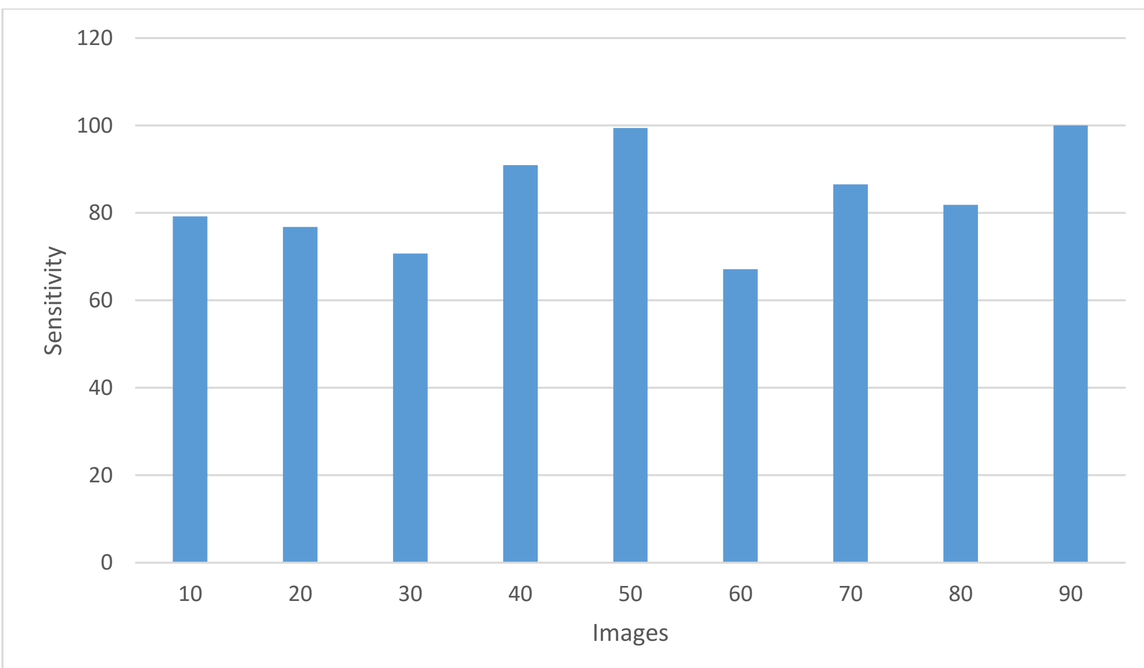

Figure 3: Map Removal Sensitivity

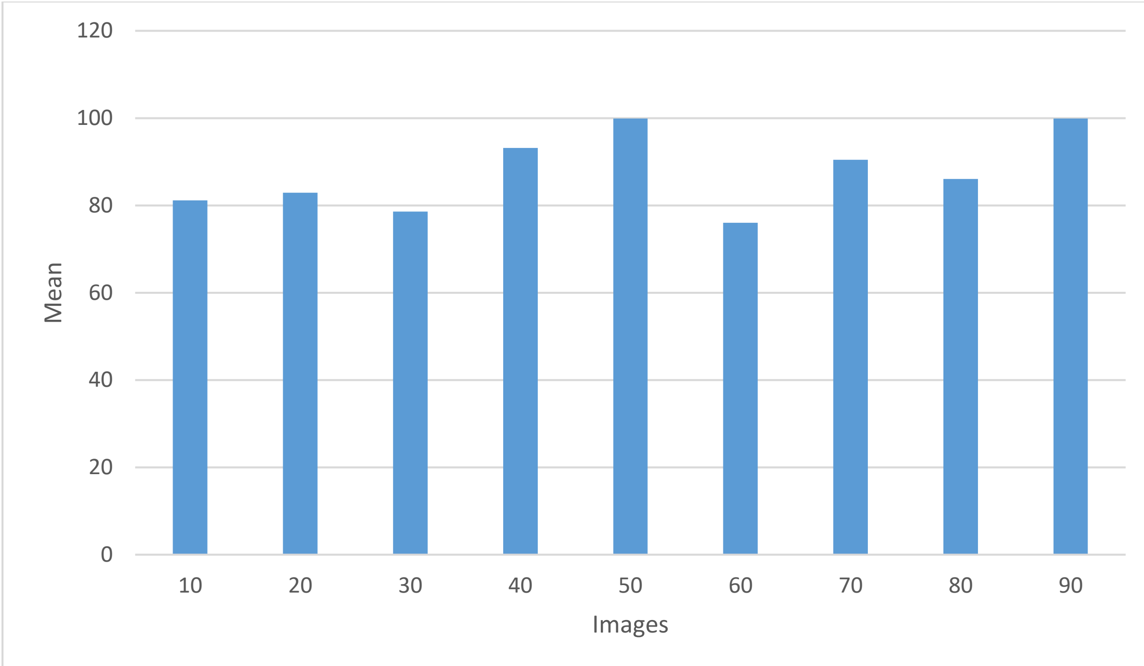

Figure 4: Mean of ResNet-18

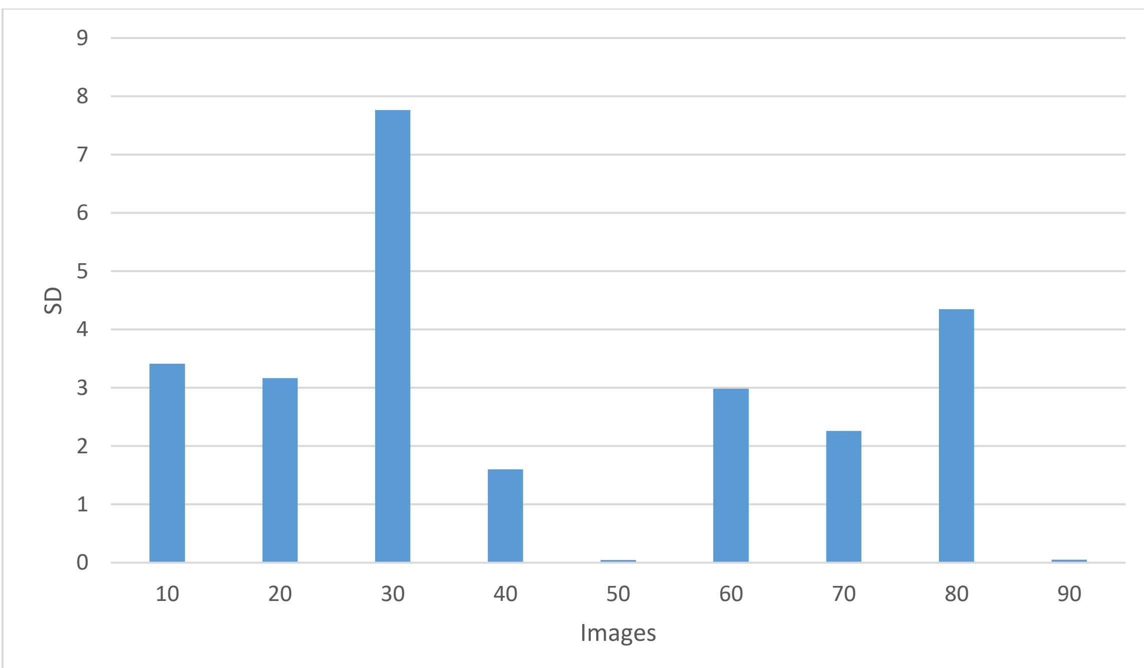

Figure 5: Standard Deviation

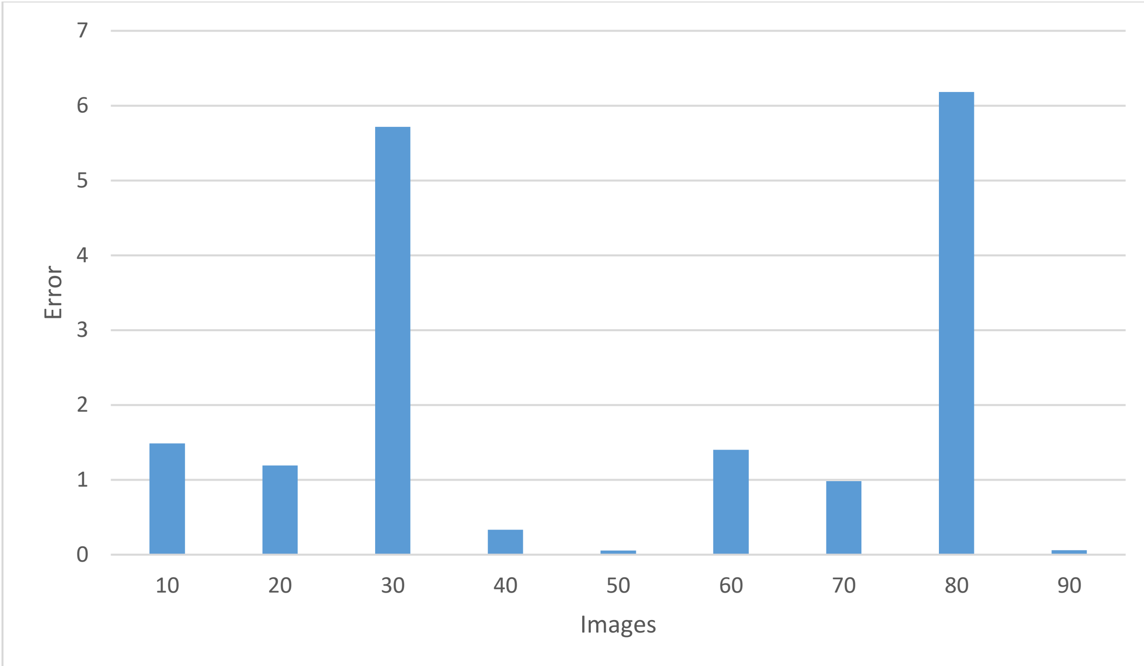

Figure 6: Prediction Error

Every one of the parameters had a consistency ratio that was less than 0.1 within the framework of the weighted pairwise comparison. This served as evidence that the accumulated expert opinions were credible and appropriate for use in the ResNet-18 approach that was utilized for the process of producing risk maps.

The Analytical Hierarchy Process Index and the nitrate chemical levels in the environment were shown to have a correlation, and the value of the Pearson coefficient for this association was equal to 0.53. In comparison to the index as well as the other approaches, this specific strategy has demonstrated a far greater link.

It is possible to incorporate human reasoning into the control algorithm as a means of coping with the unpredictability of the parameters. This can be performed without jeopardizing the algorithm security in any way. According to the updated model, 31% of those who were evaluated were classified as having a high vulnerability, 39% were classified as having a moderate vulnerability level, and 30% were classified as having either a low susceptibility range. When compared to the traditional method, this approach significantly broadens the scope of vulnerable regions within the high and moderate-to-high classes.

CONCLUSIONS

It is a resource that is necessary for human existence, the creation of goods from agriculture, and the operation of machinery in industrial settings, groundwater needs to be adequately monitored and preserved for future use. Therefore, it is of the utmost importance to conduct an accurate evaluation of the groundwater contamination vulnerability index using ResNet-18. This evaluation must be completed as soon as possible. The groundwater vulnerability index is a helpful tool for analyzing the state of the environment in various parts of the world, and the ResNet-18 model is able to lend a hand in the effort to accomplish this goal.

The goal of this study is to characterize the hydrogeochemical mechanisms that influence nitrate chemical and fluoride enrichment, as well as the geographical distribution, temporal distribution, and incidence of those mechanisms. This information will be gathered through the course of this study. The total hazard index provided a numerical portrayal of the potential harmful effects on one health that could arise from consuming groundwater that was severely contaminated. These potential adverse effects could be the result of drinking contaminated groundwater. The findings of this study help determine the severity of unique sensitive zones at a given area. This is essential for planning and executing effective approaches to ameliorate groundwater quality and human health, which are both affected by poor groundwater quality.

REFERENCES

i. Ali, M., Prasad, R., Xiang, Y., Khan, M., Farooque, A. A., Zong, T., & Yaseen, Z. M. (2021). Variational mode decomposition based rando
ii. m forest model for solar radiation forecasting: New emerging machine learning technology. Energy Reports, 7, 6700-6717.
iii. Dehghani, R., Torabi Poudeh, H., & Izadi, Z. (2022). Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. Modeling Earth Systems and Environment, 8(2), 2599-2613.

iv. K. Lakshmi Narayanan, R. Santhana Krishnan, Y. Harold Robinson, E. Golden Julie, S. Vimal, V. Saravanan, M. Kaliappan, "Banana Plant Disease Classification Using Hybrid Convolutional Neural Network", Computational Intelligence and Neuroscience, Volume 2022, Article ID 9153699, 13 pages https://doi.org/10.1155/2022/9153699, 2022.

v. Jamei, M., Ahmadianfar, I., Karbasi, M., Jawad, A. H., Farooque, A. A., & Yaseen, Z. M. (2021). The assessment of emerging data-intelligence technologies for modeling Mg+ 2 and SO4− 2 surface water quality. Journal of environmental management, 300, 113774.

vi. Tiyasha, T., Bhagat, S. K., Fituma, F., Tung, T. M., Shahid, S., & Yaseen, Z. M. (2021). Dual water choices: The assessment of the influential factors on water sources choices using unsupervised machine learning market basket analysis. IEEE Access, 9, 150532-150544.

vii. Alavi, J., Ewees, A. A., Ansari, S., Shahid, S., & Yaseen, Z. M. (2022). A new insight for real-time wastewater quality prediction using hybridized kernel-based extreme learning machines with advanced optimization algorithms. Environmental Science and Pollution Research, 29(14), 20496-20516.

viii. M. Rajalakshmi, V. Saravanan, V. Arunprasad, C.A.T. Romero, O.I. Khalaf, C. Karthik, "Machine Learning for Modeling and Control of Industrial Clarifier Process", Intelligent Automation & Soft Computing, Vol. 32, No. 1, 2022.

ix. Wang, S., Peng, H., & Liang, S. (2022). Prediction of estuarine water quality using interpretable machine learning approach. Journal of Hydrology, 605, 127320.

x. Takeda, T., Kato, J., Matsumura, T., Murakami, T., & Abeynayaka, A. (2021). Governance of artificial intelligence in water and wastewater management: The case study of Japan. Hydrology, 8(3), 120.

xi. Wu, Y., Sun, L., Sun, X., & Wang, B. (2022). A hybrid XGBoost-ISSA-LSTM model for accurate short-term and long-term dissolved oxygen prediction in ponds. Environmental Science and Pollution Research, 29(12), 18142-18159.

xii. Chia, S. L., Chia, M. Y., Koo, C. H., & Huang, Y. F. (2022). Integration of advanced optimization algorithms into least-square support vector machine (LSSVM) for water quality index prediction. Water Supply, 22(2), 1951-1963.

xiii. Bhagat, S. K., Tiyasha, T., Kumar, A., Malik, T., Jawad, A. H., Khedher, K. M., ... & Yaseen, Z. M. (2022). Integrative artificial intelligence models for Australian coastal sediment lead prediction: An investigation of in-situ measurements and meteorological parameters effects. Journal of Environmental Management, 309, 114711.

xiv. Chen, J., Chen, S., Fu, R., Li, D., Jiang, H., Wang, C., ... & Hicks, B. J. (2022). Remote sensing big data for water environment monitoring: current status, challenges, and future prospects. Earth's Future, 10(2), e2021EF002289.

xv. Zhao, D., Wang, J., Jiang, X., Zhen, J., Miao, J., Wang, J., & Wu, G. (2022). Reflectance spectroscopy for assessing heavy metal pollution indices in mangrove sediments using XGBoost method and physicochemical properties. CATENA, 211, 105967.