JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Deep Learning Based Optimal Deep belief network for Water Quality Prediction

[1]J. Charles, [2]G. Vinodhini, [3]R. Nagarajan
[1]Research scholar, Department of Computer and Information sciences, Annamalai University, Annamalai Nagar.
[2]Assistant Professor, Department of Information Technology, Annamalai University, Annamalai Nagar.
[3]Assistant Professor, Department of Computer and Information sciences, Annamalai University, Annamalai Nagar.
jcharles.1404@gmail.com, vinodhini.g.t@gmail.com, rathinanagarajan@gmail.com

Abstract

Recently, water quality monitoring becomes vital to increase protection and management of water resources. Under the related laws and regulations, environmental protection department agency monitors rivers, lakes, streams, and other kinds of water body for assessing water quality condition. The valid and high-quality data produced from the monitoring activity helps water resource manager to understand the pollution control needs, present pollution situation and energy consumption problems. In this context, this study designs a new class imbalance handling with deep belief network (CI-DBN) technique to estimate the water quality (WQ). The proposed CI-DBN model involves different processes such as pre-processing, class imbalance (CI) handling, DBN based prediction, and artificial rain drop algorithm (ARA) based hyperparameter handling. The DBN model reviews the class balanced input data and performs the prediction process. Finally, the ARA is used to optimally choose the hyperparameter values of the DBN model. A wide-ranging simulation analysis is carried out and the experimental results showcased the significate of the CI-DBN model compared to the recent WQI predictive approaches interms of various measures.

Keywords: Water quality index, Deep learning, DBN model, Parameter tuning, Artificial intelligence, Soft Computing.

1. Introduction

Presently, the water quality predictive model is classified into two classes based on their inherent characteristics: mechanism and non-mechanism (data-driven) water quality predictive methods. The mechanism model is constraint by biological, chemical, physical, and other features of the water environment system, and is based on the system structure data [1-4]. Mostly, data-driven model has considerable impacts on the prediction of water quality parameter [5]. The main method are the time sequence model, gray theory predictive model [6], regression predictive methods (like support vector machines (SVMs)) [7], and the

artificial neural networks (ANNs) predictive model [8]. But the first three models have certain shortcomings namely low prediction accuracy, poor generalization ability, and low calculation accuracy. Recently, the deep learning (DL) algorithm has gained considerable interest in modeling water quality. Artificial neural network (ANN) is a type of machine learning (ML) technique that is realized using the extensive parallel connection of self-adaptive simple unit for stimulating the biological nervous system. It is the base of DL algorithm, and it has the ability to fully fit complex nonlinear relationships and the benefits of good robustness.

Subsequently, variety of advanced techniques such as Neuro-Fuzzy Inference Scheme (ANFIS), ML, ANN, and statistical analysis tools, are investigated for the development of WQ parameter predictive methods [8]. A couple of researchers has adjudicated ANN method as the reliable and most preferred method for the improvement of WQ parameter prediction methods due to their remarkable appropriateness to nonlinear and irregular circumstances. In the study on a backpropagation (BP) neural network (NN), i.e., a general representation of ANN, and established that its improved methodologies have powerful pertinence to WQ parameter prediction with clear advantages in prediction non-linear issues [9]. Other variations, the RBFNN that is widely employed in aquaculture, offer the advantages of an unsophisticated architecture, the capacity to universally estimate random function with accuracy, and faster training speed [10].

Charles et al. [11] introduces an ML based WQ predictive method isolation forest (IF) using OANFIS approach. The suggested strategy includes pre-processing at an early phase to convert the information into a well-suited format. Furthermore, IF related outlier detective system is applied for removing the outlier that exists in the information. For predictive method, OANFIS classification is employed in which the parameter of the ANFIS approach is tuned through CPSO method. In Mariammal [12], introduced an IOT based solution for checking and predicting the WQ and aware the client beforehand the water gets contaminated. The presented methodology employs IoT and enhanced NN system for prediction. It comprises different embedded sensor nodes such as color, conductivity, pH, and turbidity. The estimated sensor value is saved in the dataset and focused on predictive investigation. The Cat swarm optimization (CSO) based NN system is applied to forecast the quality outcome.

Haghiabi et al. [13] explores the efficiency of AI techniques includes SVM, ANN, and group technique of data handling (GMDH) for predicting WQ components. To propose the SVM and ANN, various kinds of kernel and transfer functions have been investigated, correspondingly. Review the outcomes of SVM and ANN designated that the two methods have appropriate efficiency to predict WQ component. Chen et al. [14] presented an approach named TrAdaBoost-DBN that incorporates advanced DL method by DBN and instance-related TL via through TrAdaBoost. This technique gets the complete benefits of the DBN technique and TL techniques, such as effective capacity of capturing the long-term dependences amongst time series and the flexibility of leveraging the interrelated knowledge out of wide-ranging data sets for filling in largescale successive information.

This study designs a new class imbalance handling with deep belief network (CI-DBN) method to estimate the water quality (WQ). The proposed CI-DBN model involves different processes such as pre-processing,

class imbalance (CI) handling, DBN based prediction, and artificial rain drop algorithm (ARA) based hyperparameter handling. The DBN model receives the class balanced input data and performs the prediction process. Finally, the ARA is used to optimally choose the hyperparameter values of the DBN model. An extensive range of simulation analysis was executed and the experimental results showcased the significate of the CI-DBN model compared to the recent WQI predictive approaches interms of various measures.

2. The Proposed Model

In this study, a novel CI-DBN model was devised for the prediction of WQ. The proposed CI-DBN model incorporates a set of processes namely preprocessing, ADASYN based CI handling, DBN based prediction, and ARA based hyperparameter tuning. The DBN method receives the class balanced input data and performs the prediction process. Finally, the ARA is used to optimally choose the hyperparameter values of the DBN model.

2.1. ADASYN Model

It can be present to utilize ADASYN [15] that is a growth of SMOTE. ADASYN is previously initiated suitable from clinical imaging application to analysis of retinal health, analysis of focal liver lesion, and recognition of premature delivery. The synthetic instances were generated dependent upon majority of nearest neighbors using k-NN technique. This technique utilizes weighted dissemination to varying minority class instances dependent upon its level of difficulty from trained [16]. It takes further synthetic sample to instance in minority class that is strenuous for training than individual's case which is easy for training.

2.2. Design of DBN Model

A DBN is classified into two phase supervised and unsupervised pre-training. Initially, it only learns feature with input value without labels. The procedure is conducted in the following as follows. The input value learn the primary hidden layer x that sequentially learn the next hidden layer. Then, tuning with an error backpropagation using a label with supervised fine-tuning has been takes place [24].

The RBM can be denoted as follows:

$$E(v,h) = -\sum_{i,j} v_i\, h_j w_{ij} - \sum_i b_i\, v_i - \sum_j b_j\, h_j$$

Whereas v_i indicates the binary state of $i\text{-}th$ visible node, h_j denotes the binary state of $j\text{-}th$ hidden nodes , w_{ij} represents the weight between $i\text{-}th$ and $j\text{-}th$ nodes , b_i characterizes the bias term of $i\text{-}th$ visible nodes, and b_j represent the bias term of $j\text{-}th$ hidden nodes.

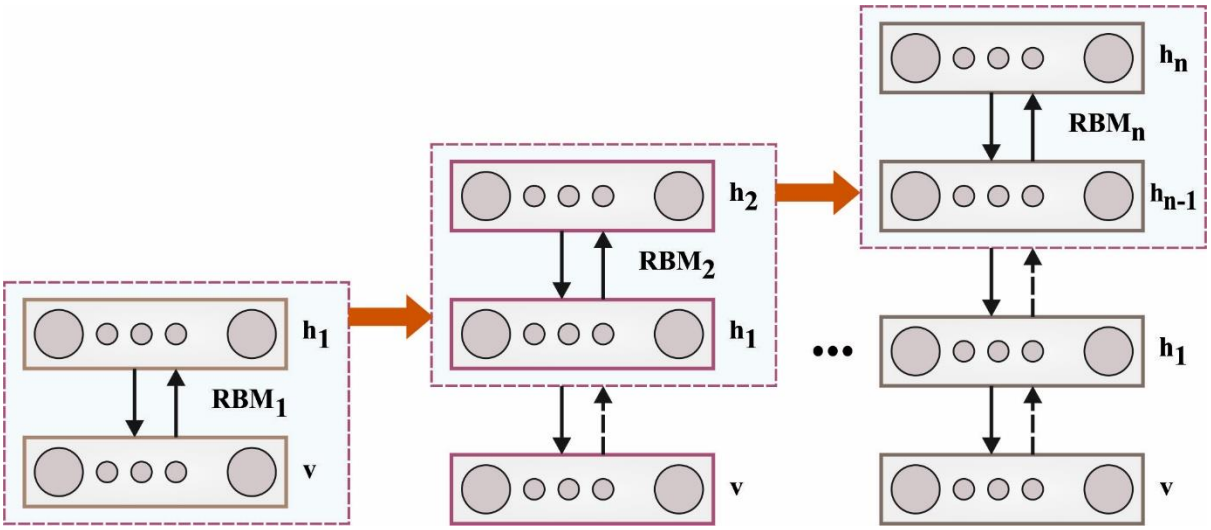

Fig. 1. Overview of DBN

2.3. Design of ARA based Hyperparameter Tuning

For effectively tuning the hyperparameters of the DBN model, the ARA is applied. The ARA stimulates the changing procedure of raindrop acquired by the observance of natural rainfall procedure. The main concept is to follows the raindrop to occupy the minimum energy state with the maximum number-the raindrop pool (RP) [17]. As many metaheuristic methods, ARA initiates by a population initialization by arbitrarily locating N vapor in a searching region, as well as all the vapors have a respective location determined in the following:

$$Vapor_i = \left(x_i^{(1)}, \cdots, x_i^{(d)}, \cdots, x_i^{(D)}\right), i = 1,2, \cdots, N. \quad (1)$$

Whereas N represent population sizes, D denotes the dimensional of problems, and $x_i^{(d)}$ indicates the location of the ith vapor in the d dimensions.

It can be considered for simplicity, that the raindrop location was the geometric center of ambient water vapor. Hence, its location is determined by Eq. (2):

$$Raindrop = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^{(1)}, \cdots, \frac{1}{N}\sum_{i=1}^{N} x_i^{(d)}, \cdots, \frac{1}{N}\sum_{i=1}^{N} x_i^{(D)}\right) \quad (2)$$

Assume $Raindrop^{(d_i)}$ represent the location of Raindrop in the $d_i th$ dimension, where $d_i (i = 1,2,3,4)$ can be arbitrarily selected from the set $\{1, 2, \cdots, D\}$. Now, $New\_Raindrop^{(d_1)}$ is attained by a linear combination of $Raindrop^{(d_2)}$, $Raindrop^{(d_3)}$ and $Raindrop^{(d_4)}$, and the other elements in $New_$Raindrops are similar to Raindrops. Consequently, the $New_$Raindrop is described by Eq. (3):

$$\begin{cases} New\_Raindrop^{(d)} = Raindrop^{(d_2)} + \varphi \cdot (Raindrop^{(d_3)} - Raindrop^{(d_4)}), if\ d = d_1; \\ New\_Raindrop^{(d)} = Raindrop^{(d)}, otherwise. \end{cases} \quad (3)$$

Let φ be arbitrary value within $(-1,1)$, $d = 1,2, \cdots, D$. Once the $New_$Raindrops contact the ground, it would be split into many smaller raindrops due to the quality and speed. Next, this small raindrop ($Small -$

$Raindrop_i, i = 1,2, \cdots, N$) would be flying in each direction. Therefore, $Small\_Raindrop_i$ is given by Eq. (4) [18]:

$$Small_{Raindrop_i} =$$

$$New\_Raindrop + sign(\alpha - 0.5) \cdot \log(\beta) \cdot (New\_Raindrop - Vapor_k) \quad (4)$$

Under the action of gravity, this $Small\_Raindrop_i (i = 1,2, \cdots, N)$ would flow from higher altitude to lower altitude direction, and many would ultimately halt at the location having low altitude (that is, best solution). The RP is intended for tracking this low position obtained until now, as well as the update of RP.

Additionally, the flowing direction of raindrop d_i for $Small\_Raindrop_i$ $(i = 1,2, \cdots, N)$ was created on the basis of linear integration of 2 vectors $d1_i$ and $d2_i$, where d_i, $d1_i$ and $d2_i$ are defined by:

$$d1_i = sign(F(RP_{k_1}) - F(Small\_Raindrop_i)) \cdot (RP_{k_1} - Small\_Raindrop) \quad (5)$$

$$d2_i = sign(F(RP_{k_2}) - F(Small\_Raindrop_i)) \cdot \left(RP_{k_2} - Small_{Raindrop}\right) \quad (6)$$

$$d_i = \tau_1 \cdot rand1_i \cdot d1_i + \tau_2 \cdot rand2_i \cdot d2_i \quad (7)$$

In which RP_{k_1} and RP_{k_2} denotes candidate solution in RP $(k_1, k_2 \in \{1,2, \cdots, |RP|\})$, τ_1 and τ_2 represent step parameter of $Small\_Raindrop_i$ flowing, $rand1_i$ and $rand2_i$ indicates uniformly distributed arbitrary number from the within (0,1), F denotes FF. Consequently, $New\_Small\_Raindrop_i (i = 1,2, \cdots, N)$ is given by:

$$New\_Small_{Raindrop_i} = Small_{Raindrop_i} + d_i \quad (8)$$

But, $Small\_Raindrop_i$ couldn't be in the flowing in a realtime ecosystem. It can be essential to present a variable Max_$Flow$_Number for controlling the maximal amount of flow. Next, they would stay in the location with a comparatively low elevation or evaporate afterward some flowing.

3. Results and Discussion

The proposed model is simulated using Python tool and the results are validated through four datasets namely monsoon, PSTM, PREM, and summer datasets. The dataset includes 15 attributes and 35 instances under each dataset.

Table 1 Predictive results of CI-DBN Method

Dataset	***Sens.***	***Spec.***	***Accu.***	***Prec.***	F_{score}	***Kappa***
Monsoon Dataset	99.68	91.69	96.93	98.33	98.96	89.10
PSTM Dataset	95.78	97.12	97.48	99.24	97.94	92.59
PREM Dataset	94.29	93.98	97.33	97.93	99.41	93.07
Summer Dataset	97.27	98.10	97.72	99.66	98.38	89.88
Average	**96.76**	**95.22**	**97.37**	**98.79**	**98.67**	**91.16**

Table 1 provides an overall WQI predictive result analysis of the CI-DBN model on four datasets.

Fig. 3 offers a comparative $sens_y$, $spec_y$, and $accu_y$ analysis of the CI-DBN model on four datasets. On the test monsoon dataset, the CI-DBN model has offered $sens_y$, $spec_y$, and $accu_y$ of 99.68%, 91.69%, and 96.93% respectively. Besides, on the test PSTM dataset, the CI-DBN model has obtained $sens_y$, $spec_y$, and $accu_y$ of 95.78%, 97.12%, and 97.48% respectively. Moreover, on the test PREM dataset, the CI-DBN model has resulted to $sens_y$, $spec_y$, and $accu_y$ of 94.29%, 93.98%, and 97.33% correspondingly.

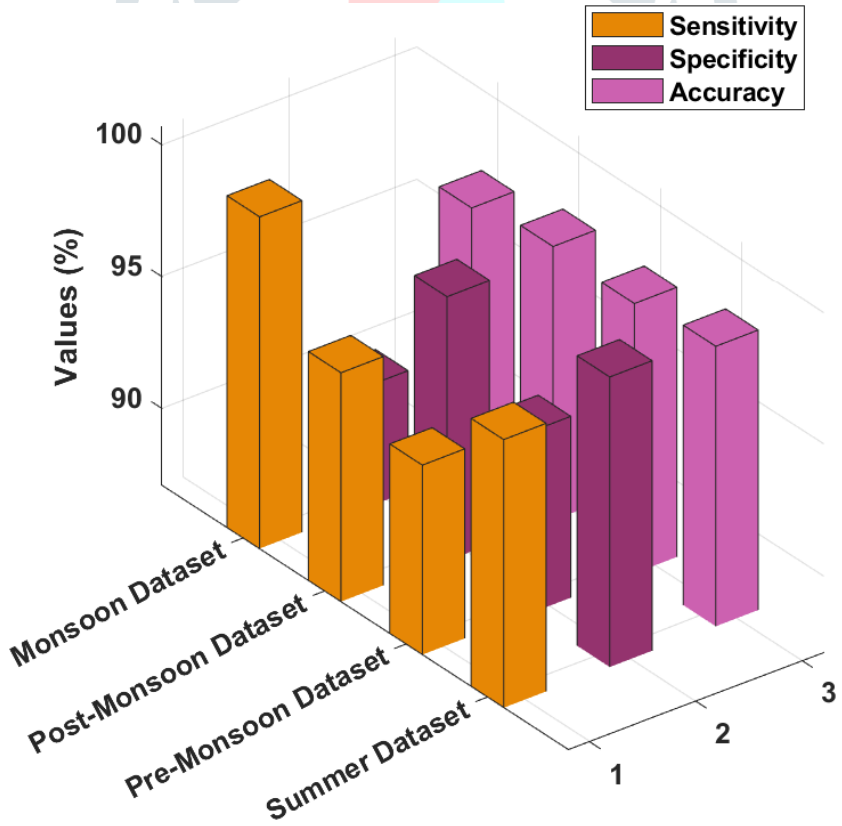

Fig. 3. Result analysis of CI-DBN method with different measures

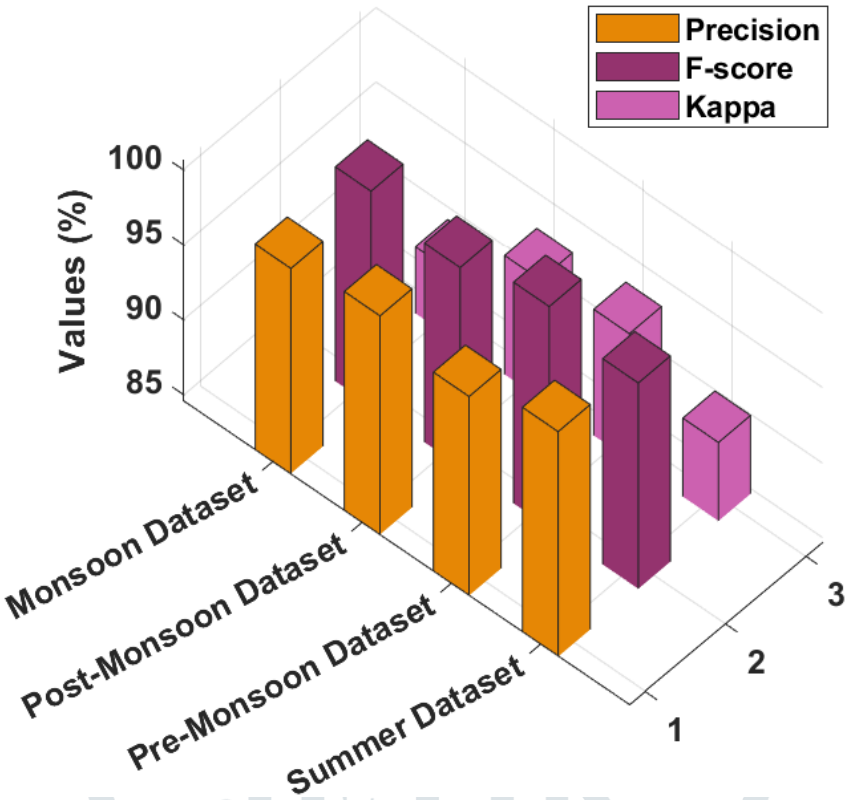

Fig. 4. Result analysis of CI-DBN technique with 4 datasets

Fig. 4 renders a comparative F_{score} *and kappa* study of the CI-DBN model on four datasets. On the test monsoon dataset, the CI-DBN model has offered sF_{score} *and* of 98.96% and 89.10% respectively. Moreover, on the test PSTM dataset, the CI-DBN model has accomplished F_{score} *and* of 97.94% and 92.59% respectively. Eventually, on the test PREM dataset, the CI-DBN model has accomplished $sens_y$, $spec_y$, and $accu_y$ of 99.41% and 93.07% respectively.

Average result analysis of the CI-DBN model takes place under four datasets in Fig. 5. The results exhibit that the CI-DBN method has resulted in average $sens_y$, $spec_y$, $accu_y$, F_{score} *and kappa* of 96.76%, 95.22%, 97.37%, 98.79%, 98.67%, and 91.16% respectively.

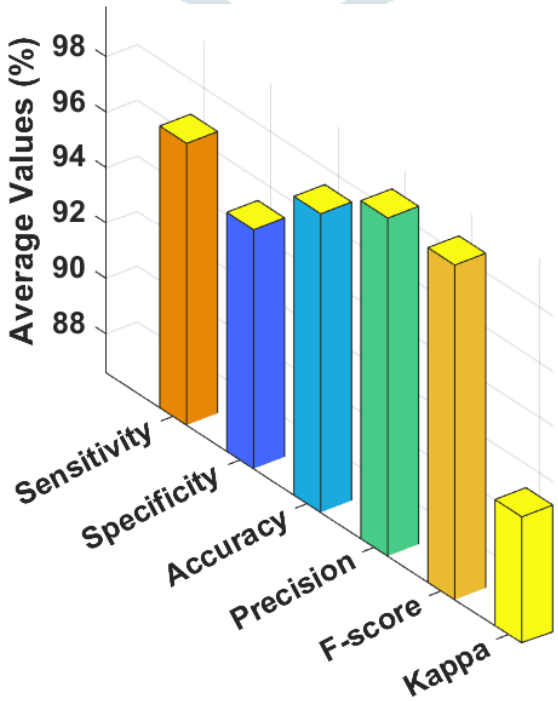

Fig. 5. Average analysis of CI-DBN technique with distinct measures

An accuracy graph of the CI-DBN model on the test four dataset has appeared in Fig. 6. The figure showcased the improvements of the validation accuracy than the training accuracy on the test data. A clear loss graph of the CI-DBN model on the test four dataset is reported in Fig. 7. The figure stated that the validation loss seems to be lower compared to training loss on the test data.

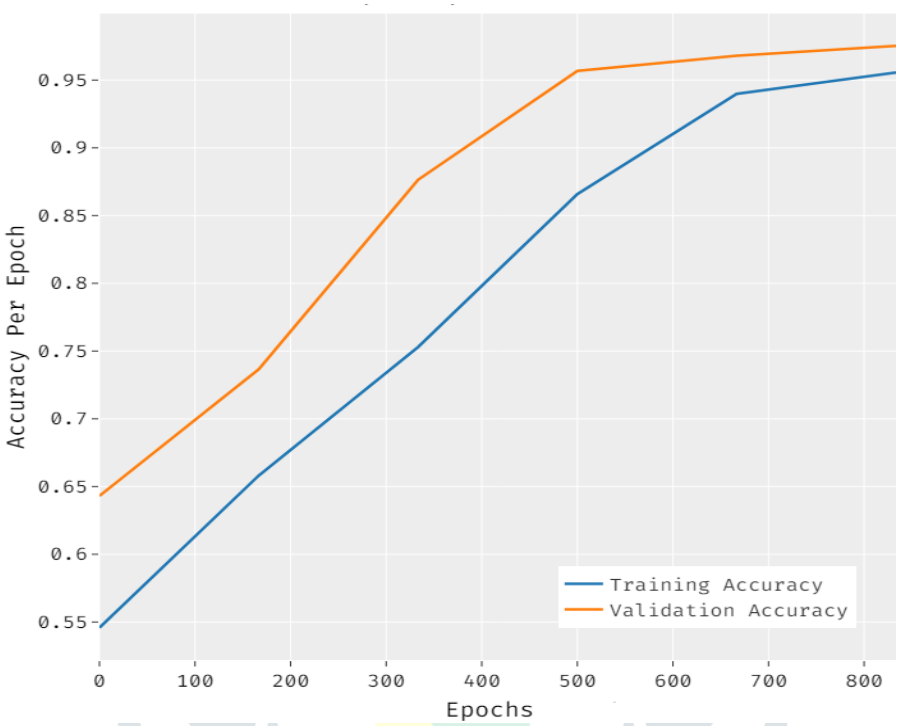

Fig. 6. Accuracy graph analysis of CI-DBN technique

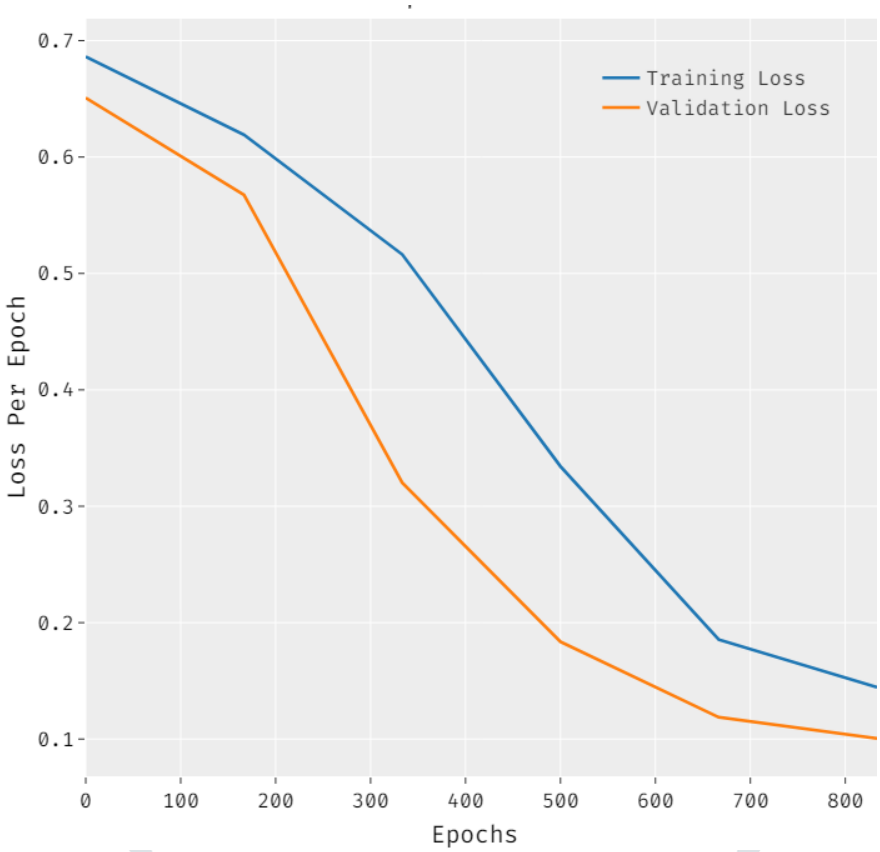

Fig. 7. Loss graph analysis of CI-DBN technique

Figs. 8-9 offers the detailed $sens_y$ and $spec_y$ analysis of the CI-DBN model with recent methods [19-23]. The outcomes reported that the RF method has obtained lower values of $sens_y$ and $spec_y$. In line with, the NB, GBT, and C4.5 models have reached slightly increased values of $sens_y$ and $spec_y$. Followed by, the DT and ANN models have tried to accomplish reasonable values of $sens_y$ and $spec_y$. However, the presented CI-DBN model has outperformed the other approaches with the superior $sens_y$ and $spec_y$ values of 96.76% and 95.22% correspondingly.

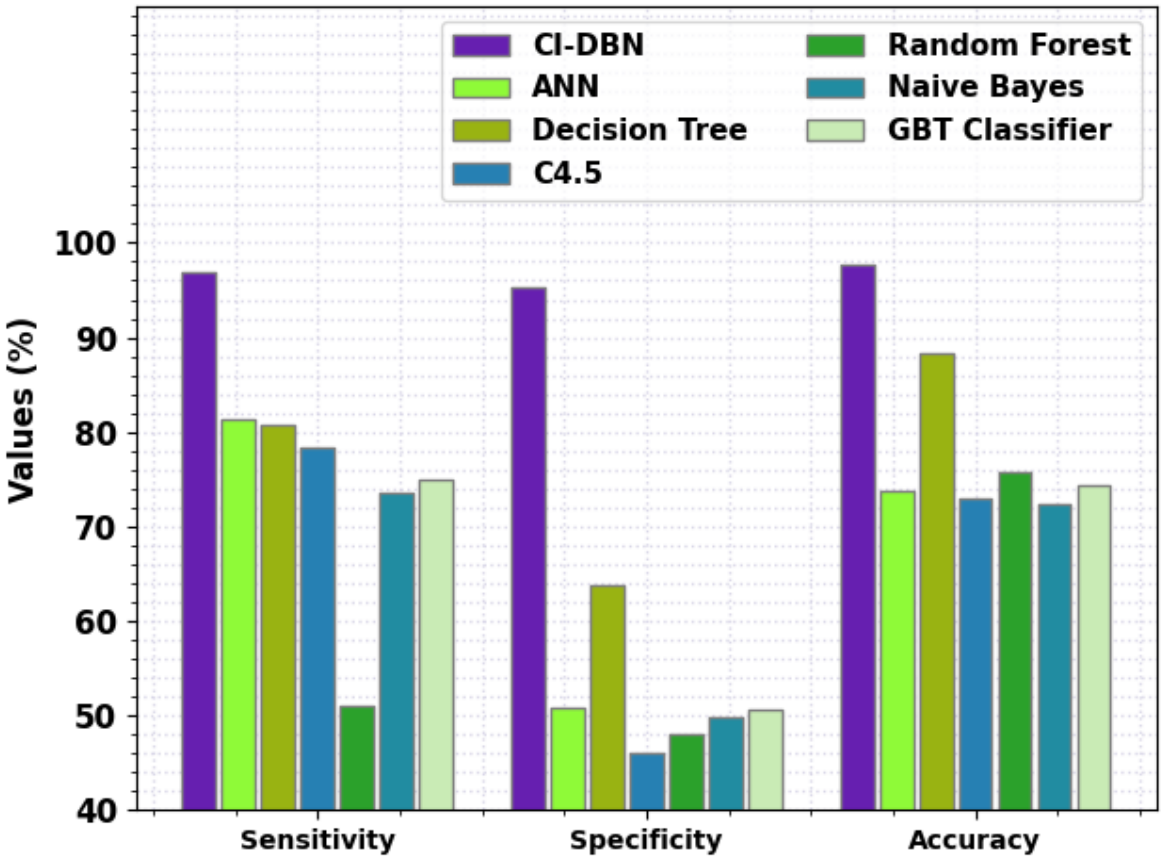

Fig. 8. Comparative analysis of CI-DBN technique with existing approaches-I

The figure also suggested the detailed acc_y and $prec_n$ analysis of the CI-DBN method with recent techniques. The results demonstrated that the RF approach has reached lesser values of acc_y and $prec_n$. Along with that, the NB, GBT, and C4.5 techniques have reached somewhat enhanced values of acc_y and $prec_n$. Afterward, the DT and ANN methods have tried to accomplish reasonable values of acc_y and $prec_n$. But, the presented CI-DBN algorithm has exhibited the other techniques with the superior acc_y and $prec_n$ values of 97.37% and 98.79% correspondingly.

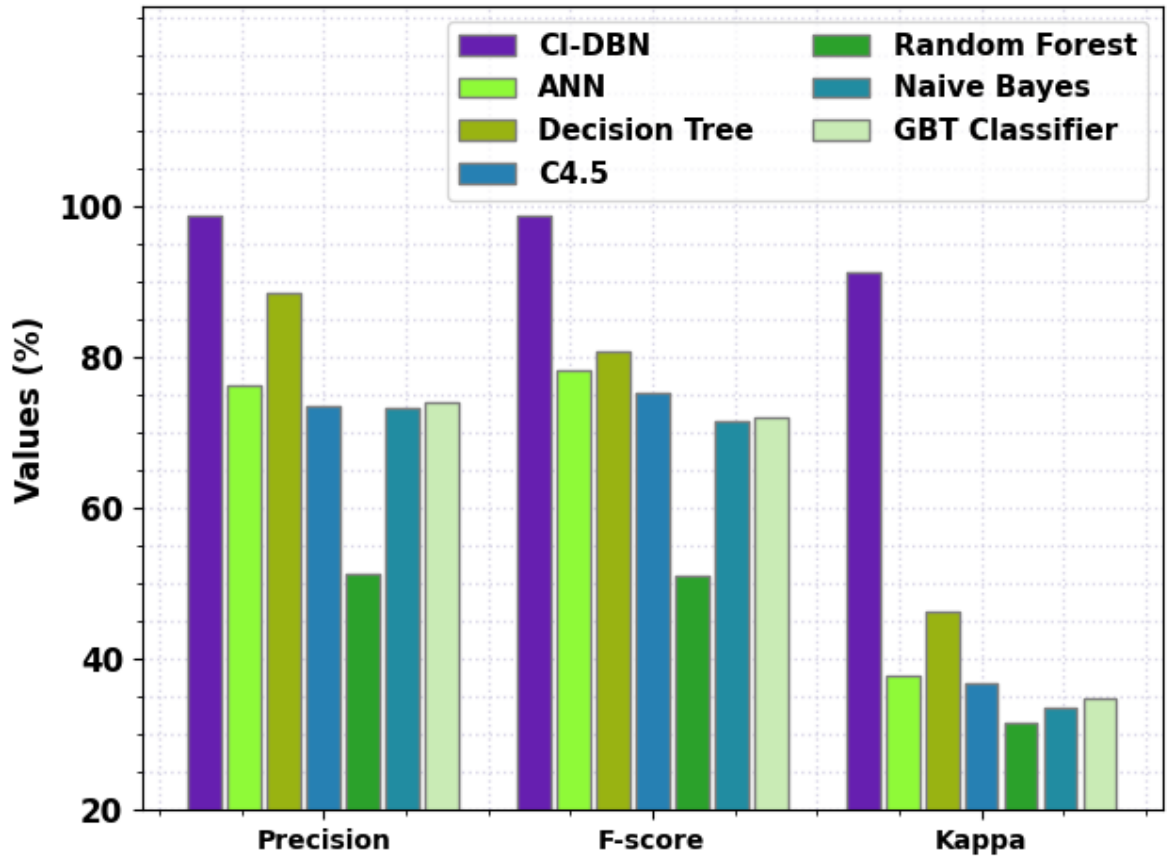

Fig. 9. Comparative analysis of CI-DBN technique with existing approaches

The outcomes depicted that the RF technique has reached lesser values of F_{score} and $kappa$. Also, the NB, GBT, and C4.5 systems have reached somewhat enhanced values of F_{score} and $kappa$. At the same time, the DT and ANN methods have tried to accomplish reasonable values of F_{score} and $kappa$. Eventually, the presented CI-DBN model has portrayed the other techniques with the superior F_{score} and $kappa$ values of 98.67% and 91.16% correspondingly.

4. Conclusion

In this study, a new CI-DBN model was devised for the prediction of WQ. The proposed CI-DBN model incorporates a set of processes they are pre-processing, ADASYN based CI handling, DBN based prediction, and ARA based hyperparameter tuning. The DBN model receives the class balanced input data and performs the prediction process. Finally, the ARA is used to optimally choose the hyperparameter values of the DBN model. An extensive range of simulation analysis was executed and the experimental outcomes showcased the significate of the CI-DBN model compared to the recent WQI predictive approaches interms of various measures. Therefore, the CI-DBN model can be applied as an effective tool for WQI prediction. In future, the predictive efficiency will be enhanced with the help of hybrid DL techniques.

References

[1] Mukate, S., Wagh, V., Panaskar, D., Jacobs, J.A. and Sawant, A., 2019. Development of new integrated water quality index (IWQI) model to evaluate the drinking suitability of water. *Ecological Indicators*, *101*, pp.348-354.

[2] Ewaid, S.H., Abed, S.A., Al-Ansari, N. and Salih, R.M., 2020. Development and evaluation of a water quality index for the Iraqi rivers. *Hydrology*, *7*(3), p.67.

[3] Nong, X., Shao, D., Zhong, H. and Liang, J., 2020. Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water research*, *178*, p.115781.

[4] Abbasnia, A., Yousefi, N., Mahvi, A.H., Nabizadeh, R., Radfard, M., Yousefi, M. and Alimohammadi, M., 2019. Evaluation of groundwater quality using water quality index and its suitability for assessing water for drinking and irrigation purposes: Case study of Sistan and Baluchistan province (Iran). *Human and Ecological Risk Assessment: An International Journal*, *25*(4), pp.988-1005.

[5] Tyagi, S., Sharma, B., Singh, P. and Dobhal, R., 2013. Water quality assessment in terms of water quality index. *American Journal of water resources*, *1*(3), pp.34-38.

[6] Lumb, A., Halliwell, D. and Sharma, T., 2006. Application of CCME Water Quality Index to monitor water quality: A case study of the Mackenzie River basin, Canada. *Environmental Monitoring and assessment*, *113*(1), pp.411-429.

[7] Avvannavar, S.M. and Shrihari, S.J.E.M., 2008. Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. *Environmental monitoring and assessment*, *143*(1), pp.279-290.

[8] Noori, R., Berndtsson, R., Hosseinzadeh, M., Adamowski, J.F. and Abyaneh, M.R., 2019. A critical review on the application of the National Sanitation Foundation Water Quality Index. *Environmental Pollution*, *244*, pp.575-587.

[9] Fathi, E., Zamani-Ahmadmahmoodi, R. and Zare-Bidaki, R., 2018. Water quality evaluation using water quality index and multivariate methods, Beheshtabad River, Iran. *Applied Water Science*, *8*(7), pp.1-6.

[10] Kumar, A. and Dua, A., 2009. Water quality index for assessment of water quality of river Ravi at Madhopur (India). *Global journal of environmental sciences*, *8*(1).

[11] Charles, J., Vinodhini, G. and Nagarajan, R., 2020. Isolation Forest With Optimal Adaptive Neuro-Fuzzy Inference System Based Water Quality Prediction and Classification Model. International Journal of Advanced Research in Engineering and Technology, 11(11).

[12] Mariammal, M.G., 2021. Efficient IOT based Water Quality Prediction Using Cat Swarm Optimized Neural Network classification. Psychology and Education Journal, 58(1), pp.4279-4282.

[13] Haghiabi, A.H., Nasrolahi, A.H. and Parsaie, A., 2018. Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), pp.3-13.

[14] Chen, Z., Xu, H., Jiang, P., Yu, S., Lin, G., Bychkov, I., Hmelnov, A., Ruzhnikov, G., Zhu, N. and Liu, Z., 2021. A transfer Learning-Based DBN strategy for imputing Large-Scale consecutive missing data and its application in a water quality prediction system. Journal of Hydrology, 602, p.126573.

[15] He H, Bai Y, Garcia EA, Li S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008 (IEEE World Congress on Computational Intelligence). IEEE; 2008. p. 1322–8.

[16] Abraham, B. and Nair, M.S., 2018. Computer-aided diagnosis of clinically significant prostate cancer from MRI images using sparse autoencoder and random forest classifier. *Biocybernetics and Biomedical Engineering*, *38*(3), pp.733-744.

[17] Q. Y. Jiang, L. Wang, X. H. Hei, Parameter identification of chaotic systems using artificial raindrop algorithm, Journal of Computational Science, vol. 8, pp. 20-31, May. 2015.

[18] Jiang, Q., Wang, L., Hei, X., Yu, G. and Lin, Y., 2016. The performance comparison of a new version of artificial raindrop algorithm on global numerical optimization. *Neurocomputing*, *179*, pp.1-25.

[19] Lerios, J.L. and Villarica, M.V., 2019. Pattern Extraction of Water Quality Prediction Using Machine Learning Algorithms of Water Reservoir. *International Journal of Mechanical Engineering and Robotics Research*, *8*(6).

[20] Muhammad, S.Y., Makhtar, M., Rozaimee, A., Aziz, A.A. and Jamal, A.A., 2015. Classification model for water quality using machine learning techniques. *International Journal of software engineering and its applications*, *9*(6), pp.45-52.

[21] Babbar, R. and Babbar, S., 2017. Predicting river water quality index using data mining techniques. *Environmental Earth Sciences*, *76*(14), p.504.

[22] Ahmed, U., Mumtaz, R., Anwar, H., Shah, A.A., Irfan, R. and García-Nieto, J., 2019. Efficient water quality prediction using supervised Machine Learning. *Water*, *11*(11), p.2210.

[23] Liao, Y., Xu, J. and Wang, W., 2011. A method of water quality assessment based on biomonitoring and multiclass support vector machine. *Procedia Environmental Sciences*, *10*, pp.451-457.