Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2023.040103
ISSN 2790-2730 Vol. 4, Issue 1: 19-28

Raman Spectroscopy System for Water Pollution Control based on Artificial Neural Network

Dadang Rustandi*

University of Freiburg, 79110 Freiburg, Germany

**corresponding author*

Keywords: Artificial Neural Network, Water Pollution, Pollution Prevention, Raman Spectroscopy

Abstract: Water resources are an important material resource for living and production. China is a country with poor water resources, and at the same time, the existing water resources in China are all polluted to varying degrees, especially the quality of surface water is closely related to the quality of people's production and life. With the boom in artificial neural network (ANN) research, neural networks have now been used in a number of fields such as graphics processing, expert decision making systems, sound processing, etc. due to the advantages of ANNs themselves, which have achieved amazing results. The theory has turned into a new multifaceted avant-garde discipline associated with multiple fields. In recent years, ANN research has been gradually applied to environmental science, some of which have applied ANN research to areas such as water eutrophication prediction and water quality prediction. The application of ANN technology to surface water quality prediction is at an early stage, and its characteristics make it a great advantage in this field. This paper investigates the use of BP ANNs to predict surface water quality, to make rapid and accurate predictions of surface water pollution, and to provide decisions for the protection of water resources and pollution prevention.

1. Introduction

ANNs are large-scale non-linear dynamic systems, consisting of a large number of neurons connected together, currently working towards further simulation of the human brain that can handle non-structural problems and can solve complex systems modelling simulation systems. Because of the original strengths of ANNs, neural networks have now been used in a number of fields, such as graphics processing, expert decision systems and sound processing, with spectacular results. The theory has turned into a new multifaceted avant-garde discipline associated with

---

multiple fields. In recent years, ANN research has been gradually applied to environmental science, some of which have applied ANN research to areas such as water eutrophication prediction and water quality prediction. The application of ANN technology to surface water quality prediction is at an early stage, and its characteristics make it more advantageous to play in this field [1-2].

In a related study, Cristian et al. describe the application of Raman and RS and its variant, surface-enhanced RS (SERS) techniques to liquid samples contaminated with different classes of substances, including nitrates, phosphates, pesticides and their metabolites [3]. The technique has also been used to detect air pollutants polycyclic aromatic hydrocarbons (PAHs), particularly benzo(a)pyrene, which is considered to be the reference for the carcinogenicity of the entire class of these compounds. In order to pre-concentrate the analytes, a method based on the well-known coffee ring effect was used, which ensures pre-concentration of the analytes without any pretreatment of the sample, providing a versatile method for the rapid and in situ detection of water contaminants. The results obtained allow us to reveal low concentrations of these analytes, close to or below their regulatory limits. a semi-analytical bio-optical model (ABOM) was proposed by Anurag et al. to derive chl-a concentrations in optically complex waters using hyperspectral data [4]. Here it is shown that the large errors associated with chlorophyll estimates in optically complex waters can be greatly reduced by modelling changes in the slope of a cdm. This study shows the development of a new approach for bio-optical modelling on a global scale.

This thesis is based on the detection principle of RS, using the Raman scattering effect of nitrite ions, to design a nitrite detection system based on RS technology. The main work of this paper is to build the optical system of the spectrometer based on the study of the mechanism of the RS detection method for nitrite, and to realise the conditioning from excitation light to Raman scattered light through the rational selection of various optical devices, the use of a good performance excitation light source and the addition of a suitable filter. Surface-enhanced RS was used to measure actual nitrite solutions and the experimental results were analysed. After experiments, it was found that the system can respond differently to different nitrite ion concentrations, allowing for qualitative and quantitative detection of nitrite solutions. The system is easy to use and intuitive due to its convenient host computer operation and real-time data curve display.

2. Design Studies

2.1. Problems in the Study

ANN algorithms are inspired by the human brain and nervous system. ANNs can learn the interrelationships between water environment variables and water quality indicators, using interconnections and 'new' variable values to be able to quantitatively predict changes in water environment quality [5-6].

Although the ANN water quality prediction model can achieve good prediction accuracy, it cannot fully describe the underlying causes of water quality from the model. Because this model does not reflect the internal principles of water quality change, the accuracy of the prediction model is limited [7-8].

When ANN models predict water quality in practice, the network approximation and generalization ability is highly dependent on the training sample, and the selection and analysis of the training set data is a difficult problem to solve. So far, there is no mature theoretical method to determine the number of nodes in a hidden layer network, relying mainly on experience to determine the number of nodes [9-10].

2.2. Qualitative Water Quality Analysis

In this paper, raman spectroscopy (RS) was used for water quality anomaly detection and pollutant class differentiation [11-12], and the main work on water quality characterization based on multispectral methods included the following three aspects.

1) Spectral data pre-processing

UV-Vis spectroscopy acquisitions are subject to noise from the instrument as well as from the environment, which can affect the subsequent data analysis. Similar noise is present during the acquisition of 3D fluorescence spectroscopy data. Also, the fluorescence spectra acquired by fluorescence instruments contain scattered light including Raman scattering and Rayleigh scattering, and these voices can mask the original fluorescence signal. Therefore, certain pre-processing means are needed to remove or attenuate the interference caused by these noises and thus improve the analyzability of the data [13-14].

2) Spectral feature extraction

UV-visible spectra are a kind of high-dimensional data, and the correlation between the data is large. Extracting the information in the specific band or the full band that can characterize the whole spectrum is of great significance for the subsequent discriminant analysis [15-16]. As 3D fluorescence spectral data is also high-dimensional data, direct full-spectrum analysis is less effective and time-efficient. Feature extraction is a critical step for the detection process, and the effectiveness of feature extraction has a great impact on the final qualitative discrimination results. Therefore, effective feature extraction is needed to characterise the information of the spectrum from the huge spectral data [17-18].

3) Anomaly discrimination or classification recognition

After feature extraction, detection needs to be carried out based on the extracted features. Commonly used qualitative detection methods include similarity metrics and building classification models. Similarity metric refers to the selection of suitable similarity metrics to measure the similarity of features between substances. On the one hand, a threshold can be set to discriminate between anomalies and non-anomalies, and on the other hand, the results of similarity metrics can be used to classify samples into the most similar categories. A classification model is a mapping model based on existing data, in which data of the same category are mapped to the same area, and for new samples, after the same processing process, the sample category can be determined based on the model mapping results.

The principle of Raman spectroscopy detection is shown in Figure 1.

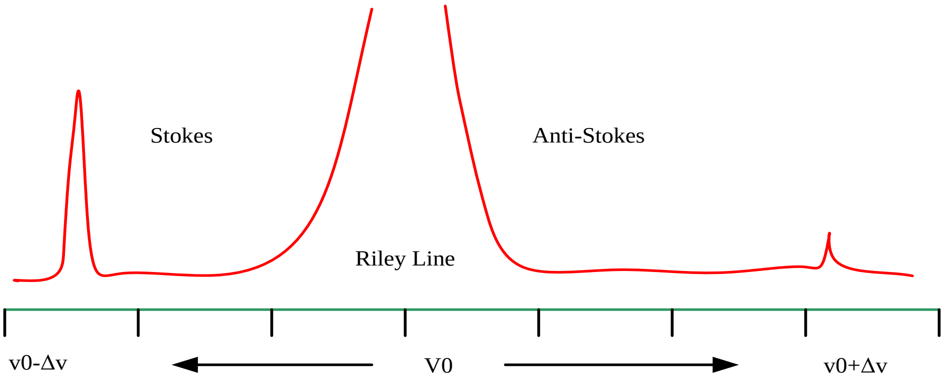

Figure 1. Schematic diagram of the Raman spectrum

3. Experimental Research

3.1. Modelling Methods for Spectral Qualitative Analysis

After pre-processing and feature extraction to obtain features that reflect the sample information, analytical models need to be built to predict the new samples. Commonly used methods include the recognition metric and building classifier models.

1) Similarity metric

The similarity measure is a metric that evaluates the degree of similarity between two samples in a comprehensive way. The purpose of this method is the same as that of the dissimilarity metric. The prediction of a new sample can be made by calculating statistical information about the similarity between its characteristics and those of other categories of samples. Commonly used similarity measures include various distance measures (e.g. Euclidean distance, cosine distance, Chebyshev distance, Manhattan distance, Marxian distance, Hamming distance, etc.), eigenvector angle measures, correlation coefficients, etc. When modelling, it is necessary to select the appropriate similarity metric according to the characteristics of the Xiangben and the characteristics of the features.

2) Classifier model

Classification is an important method in data mining. For new samples, after the same pre-processing and feature extraction, the predicted categories can be obtained by simply feeding the features into a modal I type. Different classifiers have different performance effects in different scenarios.

3.2. Adjustment Rules for BP Network Weight Coefficients

For each sample of the sample set, suppose for each of the samples p, its quadratic error function is

$$J_p = \frac{1}{2}\sum_{k=1}^{L}(t_k^p - o_k^p)^2 \tag{1}$$

Then the total error function of the network for all N samples in the sample set is

$$J = \frac{1}{2}\sum_{p=1}^{N}\sum_{k=1}^{L}(t_k^p - o_k^p)^2 \tag{2}$$

Where N is the number of samples; L is the number of network output nodes, tkp is the desired output under the action of sample p, and okp is the actual output.

Usually, we get different results using Jp for gradient search in the weighted coefficient space than we would using J. As can be seen from the article published by the inventor of the BP network, Lou Melhart, the way he uses is based on Jp, by training the network one by one with samples in the sample space in a certain order, changing parameters such as weights in a way that makes Jp fall most rapidly, and finally obtaining the optimal set of parameters.

3.3. RS Data Processing System

RS data processing is an important component of a Raman spectrometer. The functions that can be achieved with the RS software of this system are

1) Spectral data acquisition, read from the acquired spectral data via the acquired spectral data via the interface.

2) Spectral denoising, taking into account the noise of the spectral data.

3) Smoothing of spectra to improve spectral quality.

4) Compressing the relevant Raman spectral data, retaining the characteristic peaks of the spectra and squeezing out the redundant data.

5) Recognition of spectra, there are many types of Raman pollutants which are quite tedious to identify manually, this system can identify spectra through relevant algorithms.

6) Search function, when the spectrometer detects the data of the relevant pollutant, it has a clear understanding of its molecular formula, molecular weight and other physical and chemical properties, and knows its toxicological information, environmental impact, environmental standards and the corresponding emergency measures.

3.4. RS Water Quality Testing System Design and Implementation

RS water quality testing system mainly includes light source, optical circuit, photoelectric conversion device (spectrometer) and other hardware circuits and has data processing function upper computer software system.

The light source, optical path and photoelectric conversion device (spectrometer) belong to the detection unit of the system, which is the core component of the water quality detection system by RS. The selection and setting of the detection unit plays a decisive role in the accuracy and stability of the whole system, and the normal work of the detection unit is required for the realisation of the system's detection function. Therefore, the detection unit should have the characteristics of high stability, low system noise and simple structure, so the selection of materials in this paper is as follows.

(1) The light source needs to meet the 200-780nm wavelength band with good radiation intensity, can adapt to the long time experimental needs, to ensure stable and continuous output. Model DH-2000 deuterium-tungsten-halogen lamp light source, output wavelength range up to 215-2000nm, with a single ultraviolet, Raman, single visible three output modes, the light source is equipped with a key masking button, can be convenient for dark noise measurement. If the incident light intensity of the light source is too high, an attenuator can be added to reduce the radiation intensity.

(2) The optical path is mainly divided into two parts: the optical holder is a standard holder made by OceanOptics, which can be adjusted according to the needs of different experiments, and can also be fitted with collimating mirrors or attenuators to meet different detection needs. The cuvette and optical fibres are made of UV-resistant material to meet the UV requirements of the optical path, taking into account the fact that the experimental spectrum covers the UV band and the absorption and scattering problems of the optical path.

(3) The photoelectric conversion device, i.e. the spectrometer, is a Japanese Hamamatsu TM series C10082CAH spectrometer with a resolution up to the nanometre level, a spectral response range of 200-800 nm and 2048 measurement points. The spectrometer is highly accurate, stable and suitable for continuous long time experimental measurements.

The principle of this RS water quality testing system is: the light source issued a specific light intensity Raman light, vertical incidence into the sample solution to be measured, after the water sample absorption through the outgoing optical fiber transmission to the spectrometer, the spectrometer completed photoelectric conversion output to the host computer software for display, processing and storage and other operations. The main difference is that the spectrometer is replaced with a different one for testing, but other settings are not adjusted.

4. Experimental Analysis

4.1. Relationship Analysis

The values of ΔI were calculated for different nitrite concentrations and are shown in Table 1, and the relationship between the nitrite concentration and the ΔI value was then calculated by applying a linear regression method.

Table 1. Relationship between nitrite concentration and ΔI

Nitrite (nmol/L)	1	5	10	20	40	60	70
ΔI(cps)	62.4	95.3	135.5	221.3	387.6	556.6	637.8

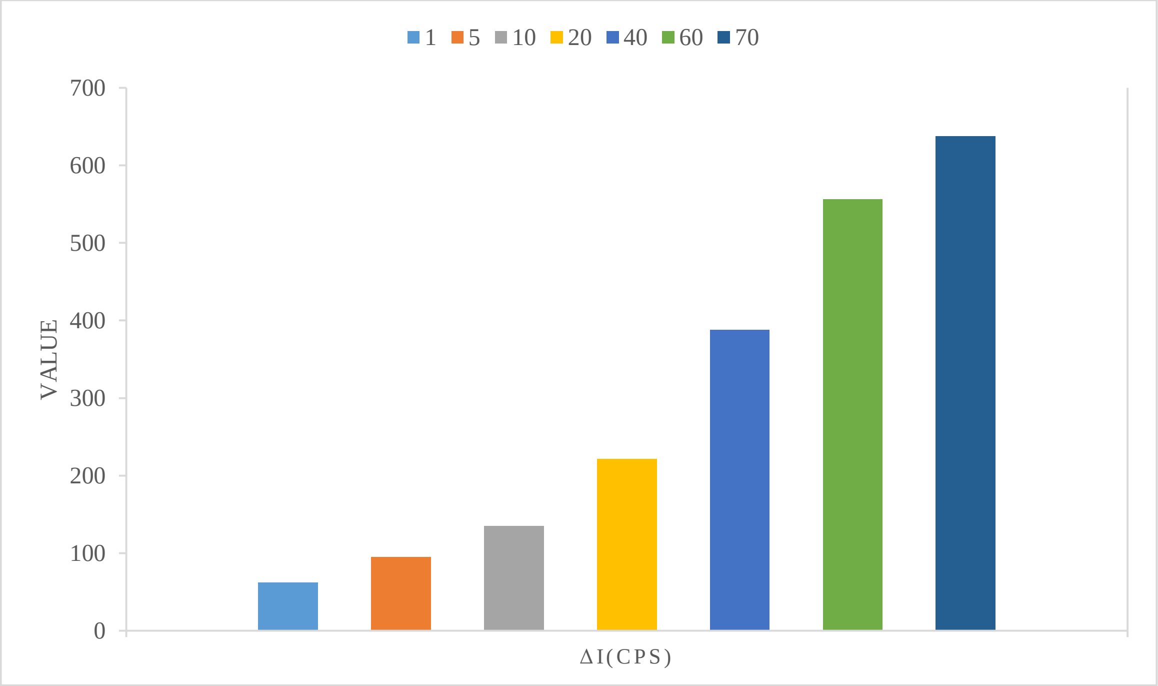

Figure 2. Analytical plot of nitrite concentration versus ΔI

Bringing the data from the Figure 2 into the formula gives

$$\triangle I=8.3C+53.6 \tag{3}$$

where ΔI is the reduction in the peak value of the surface-enhanced Raman scattering peak of Rh6G after the addition of nitrite and C is the concentration of nitrite.

At this point, the concentration of nitrite in the sample to be measured can be calculated from the value of ΔI. The value of ΔI is detected and calculated by the detection system.

In this thesis, nitrite levels in five different water samples were analysed. The existing nitrite levels in each water sample were first tested and then, an amount of nitrite was added to the water samples and measured again to obtain the reproducibility of the test.

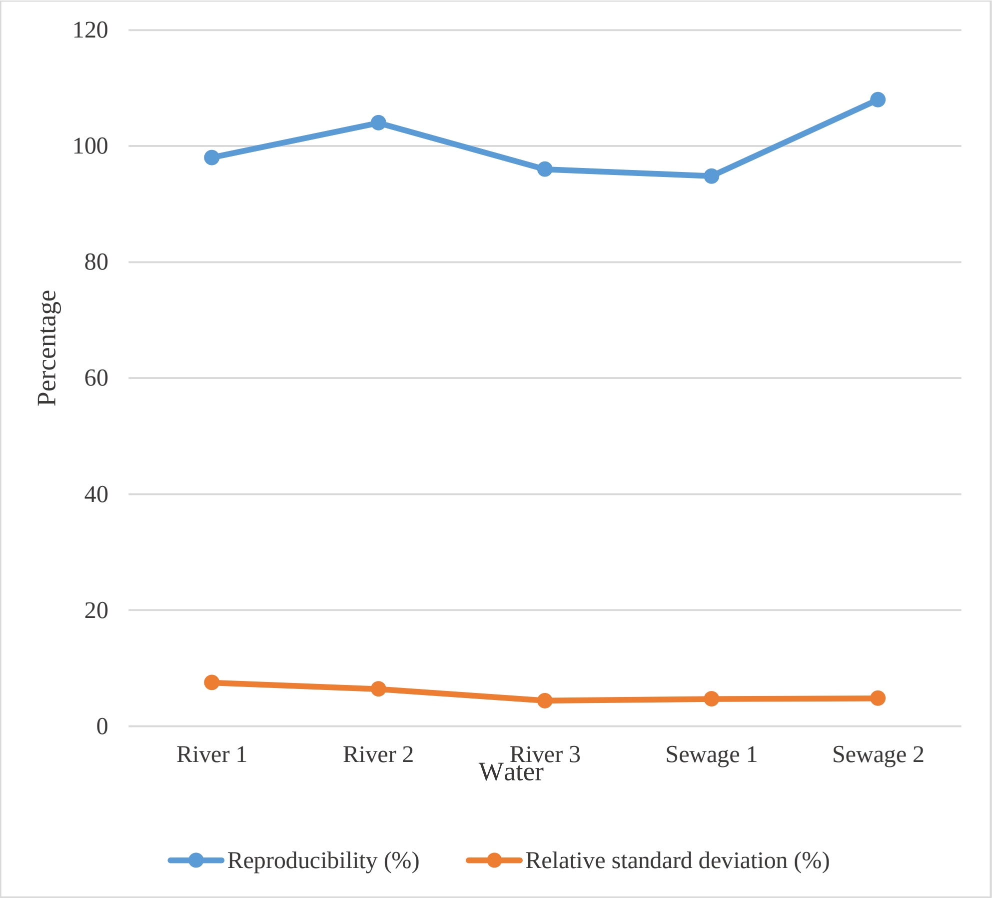

Figure 3. Analysis of results for nitrite content in various water samples

As can be seen from the data in Figure 3, the detection system can be applied to the detection of nitrite in a wide range of water samples with high accuracy. It also exhibits good reproducibility in the range of 94.8-108%.

4.2. Comparison of Results

A comparison of the monospectral and fusion models in terms of false alarm rate, recall and F1 scores is summarised in Table 2 (underlining indicates the worst model for the corresponding metric and bolded font indicates the best model for the corresponding metric).

Table 2. Comparison of anomaly detection results of different models

Models	False Alarm Rate	Recall Rate	F1 Score
UV-visible spectral single-spectrum model	0.083	<u>0.663</u>	0.768
Three-dimensional fluorescence spectroscopy monospectral model	<u>0.25</u>	0.962	0.895
Spectral fusion models	0.083	0.90	0.917

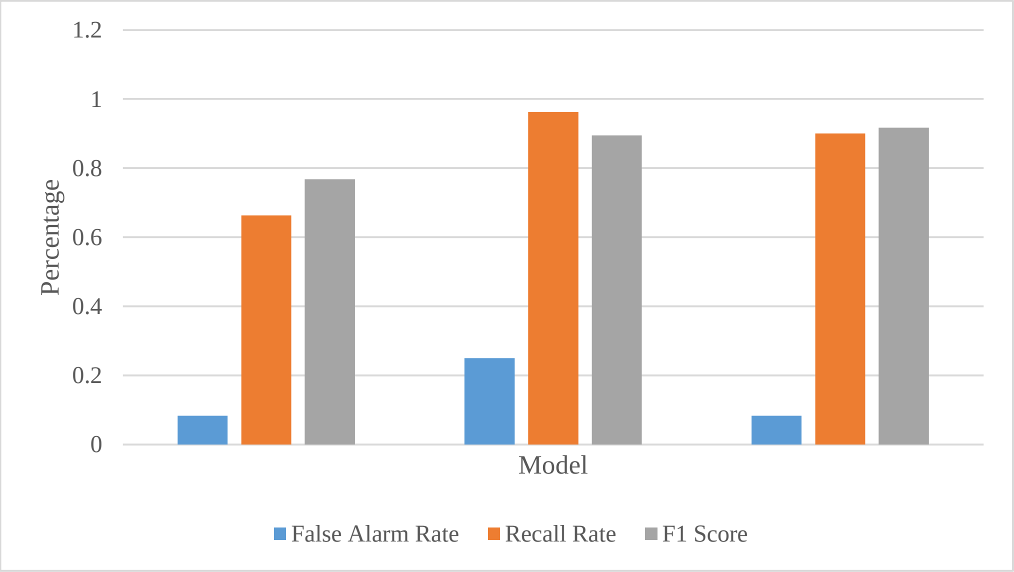

Figure 4. Comparative analysis of anomaly detection results of different models

As can be seen in Figure 4, the fusion model performs better in terms of false alarm rate as well as FI score, with a false alarm rate of 8.3%, a recall rate of 90% and an F1 score of 0.917; the single-spectrum model for fluorescence spectroscopy has a false alarm rate of 25%, a recall rate of 96.2% and an F1 score of 0.895; the single-spectrum model for UV-Vis spectroscopy has a false alarm rate of 8.3%, a recall rate of 66.3% and an FI score of The recall rate of the single-spectrum model was 8.3% and 66.3%, with an FI score of 0.768.

5. Conclusion

As water quality safety is a matter of national importance, it is important to quickly and effectively detect and classify organic pollutants in water bodies. In this paper, the influence law of environmental factors changes on the detection of UV-visible spectra and 3D fluorescence spectra is investigated, and the method of ultraviolet-fluorescence information fusion-based organic pollutant over-limit detection is studied for the problems of high lower limit of organic pollutant detection under the single-spectrum model of UV-visible spectra and high false alarm rate under the single-spectrum model of 3D fluorescence spectra due to the normal fluctuation of raw water quality; for some organic substances in the UV Visible spectroscopy or three-dimensional fluorescence spectroscopy single-spectrum detection response similar, resulting in the UV-visible single-spectrum model or three-dimensional fluorescence spectroscopy single-spectrum model for the poor differentiation of substances, the study of organic pollutant classification and identification method based on RS.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] K. A. Tapia Espinoza, Marcos Leonardo Fuentes Ávila, V. F. Pichardo Oropeza, J. M. Falcón González, J. C. Martínez-Espinosa. Qualitative Analysis of Ciprofloxacin using RS and K-Means: Preliminary Results. Res. Comput. Sci. (2020) 149(2): 23-26.

[2] Imane Zouaneb, Mostefa Belarbi, Abdellah Chouarfia. Converging image processing and data mining for RS analysis. Int. J. Commun. Networks Distributed Syst. (2022) 28(3): 287-311. https://doi.org/10.1504/IJCNDS.2022.122176

[3] Cristian Ciobanu, Katherine J. I. Ember, Balázs J. Nyíri, Sreeraman Rajan, Vinita Chauhan, Frederic Leblond, Sangeeta Murugkar. Potential of RS for Blood-Based Biopsy. IEEE Instrum. Meas. Mag. (2022) 25(1): 62-68. https://doi.org/10.1109/MIM.2022.9693451

[4] Anurag Gupta, Syed Moosa Ali, Aswathy Vijaya Krishna, Arvind Sahay, Mini Raman. Role of Visible Spectroscopy in Bio-Optical Characterization of Coastal Waters. IEEE Geosci. Remote. Sens. Lett. (2021) 18(8): 1327-1331. https://doi.org/10.1109/LGRS.2020.3003663

[5] David Chen. Analysis of Machine Learning Methods for COVID-19 Detection Using Serum RS. Appl. Artif. Intell. (2021) 35(14): 1147-1168. https://doi.org/10.1080/08839514.2021.1975379

[6] Giuseppe Sciortino, Andrea Ragni, Alejandro De la Cadena, Marco Sampietro, Giulio Cerullo, Dario Polli, Giorgio Ferrari. Four-Channel Differential Lock-in Amplifiers with Autobalancing Network for Stimulated RS. IEEE J. Solid State Circuits. (2021) 56(6): 1859-1870. https://doi.org/10.1109/JSSC.2020.3046484

[7] Swati Chopade, Hari Prabhat Gupta, Rahul Mishra, Preti Kumari, Tanima Dutta. An Energy-Efficient River Water Pollution Monitoring System in Internet of Things. IEEE Trans. Green Commun. Netw. (2021) 5(2): 693- 702. https://doi.org/10.1109/TGCN.2021.3062470

[8] Amal Agarwal, Lingzhou Xue. Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis. Technometrics. (2020) 62(2): 161-172. https://doi.org/10.1080/00401706.2019.1623076

[9] K. Kamaraj, B. Lanitha, S. Karthic, P. N. Senthil Prakash, R. Mahaveerakannan. A Hybridized ANN for Automated Software Test Oracle. Comput. Syst. Sci. Eng. (2023) 45(2): 1837-1850. https://doi.org/10.32604/csse.2023.029703

[10] Junaid Rashid, Sumera Kanwal, Muhammad Wasif Nisar, Jungeun Kim, Amir Hussain. An ANN-Based Model for Effective Software Development Effort Estimation. Comput. Syst. Sci. Eng. (2023) 44(2): 1309-1324. https://doi.org/10.32604/csse.2023.026018

[11] Roman Englert, Jörg Muschiol. Numerical Evidence That the Power of ANNs Limits Strong AI. Adv. Artif. Intell. Mach. Learn. (2022) 2(2): 338-346. https://doi.org/10.54364/AAIML.2022.1122

[12] Pablo Negro, Claudia Pons. Artificial Intelligence techniques based on the integration of symbolic logic and deep neural networks: A systematic review of the literature. Inteligencia Artif. (2022) 25(69): 13-41. https://doi.org/10.4114/intartif.vol25iss69pp13-41

[13] Motoaki Hiraga, Kazuhiro Ohkura. Topology and weight evolving ANNs in cooperative transport by a robotic swarm. Artif. Life Robotics. (2022) 27(2): 324-332. https://doi.org/10.1007/s10015-021-00716-9

[14] Riya Aggarwal, Hassan Ugail, Ravi Kumar Jha. A deep ANN architecture for mesh free solutions of nonlinear boundary value problems. Appl. Intell. (2022) 52(1): 916-926. https://doi.org/10.1007/s10489-021-02474-4

[15] Faiyaz Ahmad. Deep image retrieval using ANN interpolation and indexing based on similarity measurement. CAAI Trans. Intell. Technol. (2022) 7(2): 200-218. https://doi.org/10.1049/cit2.12083

[16] G. Gokulkumari. Metaheuristic-Enabled ANN Framework for Multimodal Biometric Recognition with Local Fusion Visual Features. Comput. J. (2022) 65(6): 1586-1597. https://doi.org/10.1093/comjnl/bxab001

[17] Eddy Kwessi, Lloyd J. Edwards. ANNs with a signed-rank objective function and applications. Commun. Stat. Simul. Comput. (2022) 51(6): 3363-3388. https://doi.org/10.1080/03610918.2020.1714659

[18] M. Mohamed Asan Basiri. Versatile Architectures of ANN with Variable Capacity. Circuits Syst. Signal Process. (2022) 41(11): 6333-6353. https://doi.org/10.1007/s00034-022-02087-3