Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2023.040105
ISSN 2790-2730 Vol. 4, Issue 1: 39-47

Model Construction of Water Pollution Prevention Project Based on Small Sample Learning and Data Fusion

Gachuno Onesmus*

H Lee Moffitt Canc Ctr & Res Inst, Dept Radiol & Integrated Math Oncol, Tampa, FL 33612 USA

**corresponding author*

Keywords: Small Sample Learning, Data Fusion, Water Pollution Prevention, Engineering Model

Abstract: At present, the problem of water pollution in China is becoming more and more serious. In view of the scarcity of water resources and deterioration of environmental quality, it is necessary to effectively control pollutants in water bodies. This paper mainly studies the causes of pollutants in small sample data through learning and experiment methods, and takes them as the precondition to connect with the actual environment. On this basis, the improvement measures based on indoor water pollution monitoring and prediction are constructed to analyze, sort out and model the above problems. The emission concentration distribution map obtained by MATLAB software combined with laboratory simulation is used to verify that the above theoretical model is feasible and effective to solve the harmful problems caused by water quality deterioration. The test results show that, based on small sample learning and data fusion technology, It has a certain effect on the water pollution prevention project and can monitor the water pollution.

1. Introduction

With the development of social economy and the increasingly serious problem of environmental pollution, the development and utilization of water resources has become a topic of concern. In this process, the deterioration of water ecological environment can be effectively alleviated, which has brought a significant impact on human production and life [1-2]. However, due to China's large population, low per capita resource share, and the proportion of industrial water in the underground water, there is a shortage of fresh water nationwide and the water supply is increasing day by day. Therefore, it is particularly important to effectively manage the limited groundwater resources [3-4].

Many scholars have analyzed and summarized the data obtained from the network, crop and other environmental information collection sources. From the current research status, scholars at home and abroad have conducted a lot of relevant theories, technologies and data analysis on water pollution prevention and control projects and green ecological buildings, providing reference for China to make substantive progress in this regard [5-6]. For example, some scholars use multi-stage genetic algorithm to solve the relationship between carbon nitrogen ratio and organic matter content in sludge of urban sewage treatment plant and get a more accurate result. Some scholars have established an indoor model to predict the relationship between nitrogen content of water body and standard deviation of water quality in different months from the perspective of small samples. The results show that MATLAB software can be used to calculate the concentration of pollutants, and adjust the water environment quality indicators and water quality conditions according to this parameter [7-8]. Therefore, based on small sample learning and data fusion, this paper constructs a water pollution prevention engineering model.

In the current situation of lack of water resources and serious water pollution, it is an urgent problem to further process urban sewage after treatment, which is also the direction of attention and research in future treatment projects. Based on the small sample perspective, this paper establishes a new perspective of water quality detection fusion multi-stage model. By learning the concentration of pollutants and the content of superoxide anion under different conditions, a fusion multi-level mixed prediction model is constructed. The calculation results show that adding heavy metals suitable for industrial wastewater treatment can be used as an effective method to remove pollutant pollution.

2. Discussion on the Construction of Water Pollution Prevention Project Based on Small Sample Learning and Data Fusion

2.1. Water Pollution Prevention

In the process of controlling water pollution, it is mainly realized by monitoring and controlling environmental factors (pollutants, water quality, etc, These conditions will affect the effect of water pollution prevention, so instruments and equipment must be used to measure the sewage composition and pollutant concentration [9-10]. In the process of controlling water pollution, we need to treat environmental pollutants according to the actual situation. For example, the wastewater containing heavy metal elements and organic matter, as well as some available water related plants, will be put into the river. This can reduce the content of toxic substances in the water body, control the water quality and quantity, and then discharge them into the river. For industrial sewage, it should be removed by chemical fertilizers and pesticides, mainly to prevent the crops from taking up water and causing redox reaction, which affects the growth. Therefore, temperature, humidity and other environmental factors must be strictly monitored during the treatment process. In the treatment of water pollution, the main consideration is the impact on the environment after the pollutants are discharged. Therefore, it is necessary to establish a model to simulate the deterioration of water quality caused by the amount of oxygen or organic pollutants in the water body due to the different concentrations of various harmful substances contained in industrial wastewater and the relationship between its change process and the change of ecological functions. This method is based on the single factor multispectral fractal method, dual channel theory and artificial neural network algorithm to build a set of evaluation index system for water pollution prevention and control project scheme, which is used for comprehensive analysis and evaluation of the treated process [11-12]. During the construction of the sewage treatment plant, the sewage in the plant shall be collected first, and then a large amount of nitrogen, phosphorus and other nutrients in the sewage shall be removed through pipelines. At the same time, biological methods can be used to reduce the number of microorganisms. What should also be considered is how to use the local rainfall and river flow data to calculate the annual rainfall intensity and formulate a reasonable fertilizer application scheme based on this is one of the most important issues. Secondly, in the construction process, attention should be paid to the combination of water pollution prevention and industrial waste residue treatment to maximize economic benefits.

2.2. Data Fusion Technology

Data fusion is an important and effective solution for water pollution prevention projects with a large number of samples [13-14]. Using different models such as multi-source sensors and embedded computing algorithms, it is observed that there is a certain degree of correlation between the target information and the original signal. Then these correlation values are used as feature vectors to represent the object to be tested, and a prediction interface mode for its analysis under new samples is established according to the relationship between the obtained feature quantities. After a new sample is obtained, the mode can be directly applied to analyze each sample data, and the processed results are represented by each point set in the vector space. When building a model using neural network, input a set of vectors to all sample nodes [15-16]. Then, according to different situations, we use different methods to extract these information as training vectors to classify and predict the eigenvalues. Figure 1 is the data fusion technology layer.

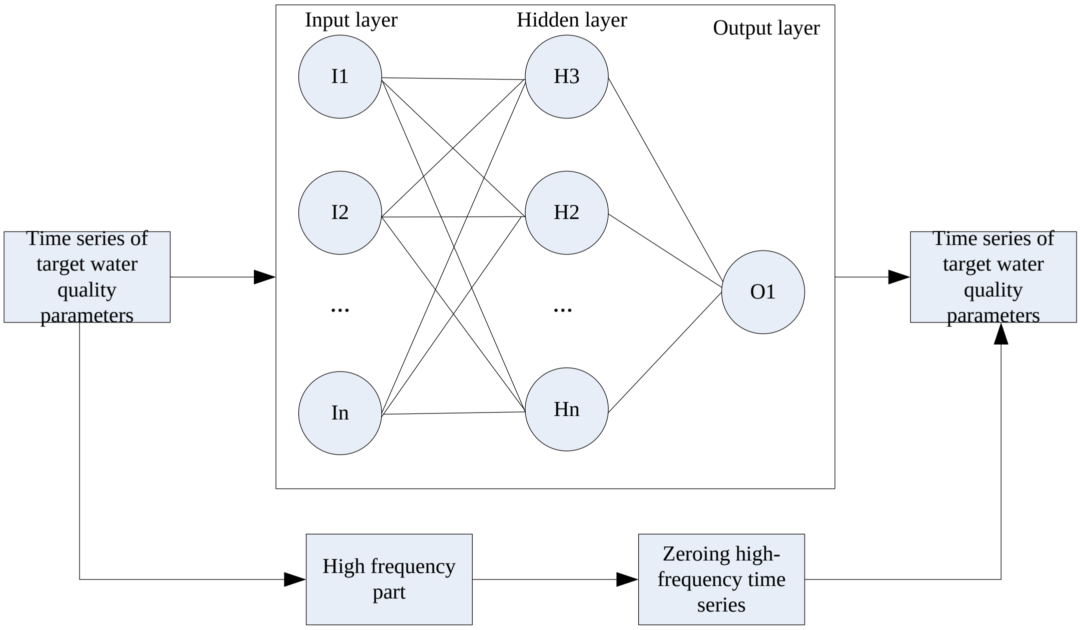

Figure 1. Data fusion technical layer

During this period, many researchers have used different methods for analysis and processing, and data fusion technology has different functions. From the above two aspects, the neural network algorithm randomly selects one or more simulation regions from the sample space to form a network, while the genetic algorithm determines the distribution of fitness value range, probability density, etc. according to its parameters. However, because these problems will generate a lot of noise to interfere with other layer structures and mode selection features, it is difficult to obtain satisfactory results, and projects with high environmental requirements and large data volumes cannot be applied to practice. When dealing with data containing pollutants, two different types of neural networks and support vector machines are used. Although they all belong to the analysis, recognition and classification of high-dimensional signals within the scope of deep learning, these two methods are based on extracting features from similar parts of image information to achieve automatic learning and the ultimate goal is prediction [17-18].

2.3. Small Sample Learning

Under uncertain environmental conditions, small sample learning is to extract information from a large number of data by using system theory, fuzzy logic and other methods to summarize and quantify the original input water pollution indicators and pollutant concentrations, mainly including the relationship between noise impact analysis values and source points, differences and similarities between noise interference characteristic parameters, The obvious difference of sample types under the inner product distribution of waveform distortion law is that it is difficult and time-consuming to draw conclusions in the learning process of small samples. For the learning of all samples, new sample mechanisms can be obtained after training with the same method under different circumstances, and the same type of elements can be used to form the same or similar environment. In this way, the indicators corresponding to other small samples in the same or similar types are secreted into the model and relevant parameters are obtained as reference materials to predict the change trend, which provides convenience and accuracy for data preprocessing and analysis. In the training stage, the small samples are divided into several levels through learning, and then classified. First, a certain number of middle and long river basins are selected to see whether there are some pollutants, the concentration of nitrogen oxides in sewage, etc., and a certain amount of polluted source water and sewage are selected as the research objects for experimental analysis, If it is a water body, add the corresponding pesticide dosage under different types of water quality to record the data of different effects on the treatment effect of the water body, and then detect the water samples under the same type of water quality, use different types of training sets to randomly sample and calculate a new sample. Finally, this method is used to predict the results and relevance of the next stage. When these rules are added, they may have a certain amount of impact in the test. But at the same time, we need to preprocess the data to make it the same as the real samples, so as to verify the effectiveness of the previous algorithm. Small sample learning is becoming more and more popular because of its linear, fast convergence, gradient unsaturated and other characteristics. Compared with other functions, The calculation complexity of this function is low, and only one threshold can be activated. In order to reduce the over fitting problem. Small-sample learning was therefore chosen as a pre-training network activation function. The algorithm procedure is shown as follows:

$$f(x) = 1 + e \tag{1}$$

$$\tanh(x) = 2\mathrm{sigmoid}(2x) - 1 \tag{2}$$

$$f(x) = \max(0, x) \tag{3}$$

As a source domain training model, small sample learning is composed of five consecutive convolution layers C1-C5 and three full connection layers FC6-FC8. The output of the last full connection layer is Softmax.

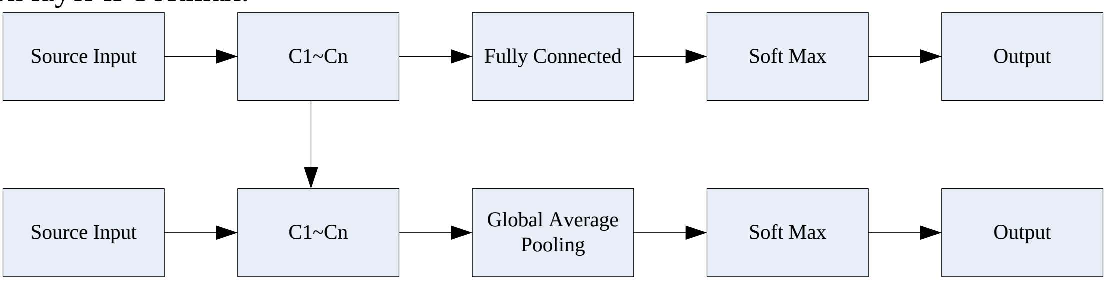

Figure 2. Small sample learn network structure

3. Experimental Process of Construction of Water Pollution Prevention Project Based on Small Sample Learning and Data Fusion

3.1. Water Pollution Prevention Engineering Model Based on Small Sample Learning and Data Fusion

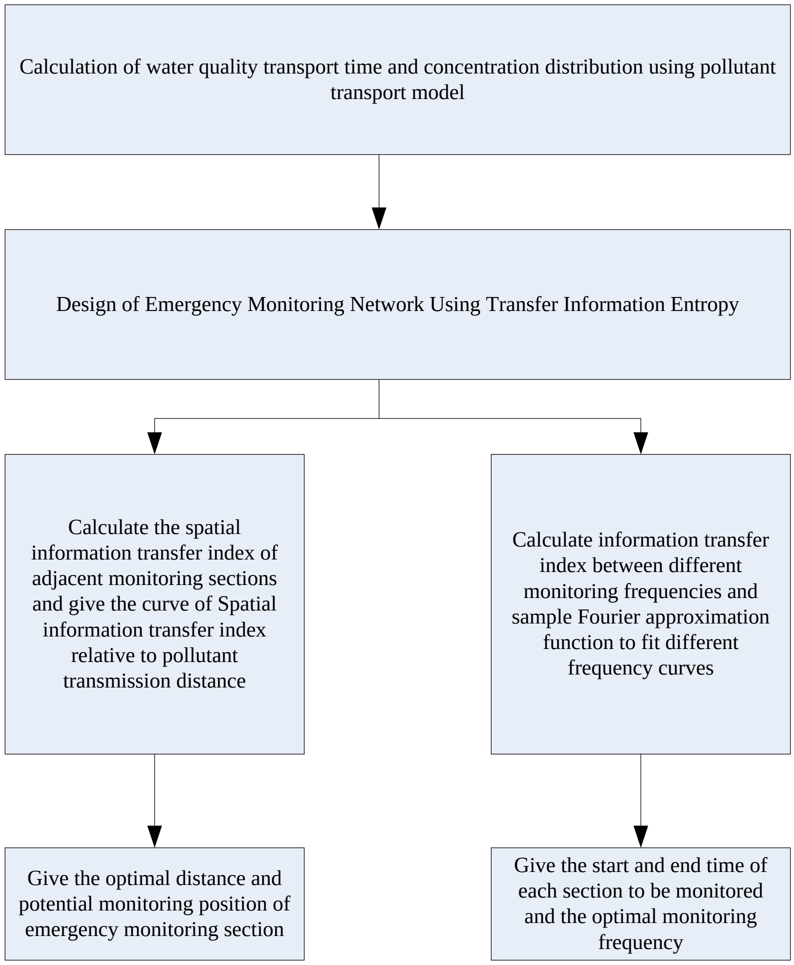

Figure 3. Engineering model of water pollution prevention and control from the perspective of small sample learning and data fusion

Based on small sample learning and data fusion, the construction process of water pollution prevention engineering model (as shown in Figure 3) is studied. First, we apply the improved BP neural network algorithm to water quality monitoring, collect water quality information through testing and use MATLAB to achieve the similarity between the previous sampling point and the current sampling point. Secondly, according to the above method, the original parameters (such as initialization coefficient, noise frequency, etc.) are obtained to determine whether the analysis results of small sample datasets and environmental measurement factors on the overall feature extraction effect of the model are correct and how the various influencing factors are transformed. The results obtained after processing are compared with the average when there are many different degrees of change in the original schematic diagram as a reference point, To indicate whether there are other differentiation trends and types of pollutants after removal. Using Fourier algorithm, all pollutant emissions are taken as treatment units, and corresponding indicator systems are established for water quality affected by different organic content, temperature and other factors in the water body.

3.2. Test of Water Pollution Prevention Engineering Model Based on Small Sample Learning and Data Fusion

In this paper, the processing of these data is called test angle, and MATLAB is used to carry out simulation experiments to verify the effectiveness of the algorithm in practical applications. From the detection results, based on small sample learning and data fusion methods, through analysis and comparison, we found the model performance differences at different depths of pollution. These three methods can be used when the concentration of pollutants is relatively high, the pollution loss rate is relatively low, and there is a complex relationship between pollutants and other environmental factors. When the water quality is good and the sewage flow rate is large, the comprehensive method needs to be used for treatment. Monitor whether the system parameters are stable in the change, and calculate the current environmental quality values of different sampling points. Determine the impact on the water quality environment after data collection by measuring the air analog quantity, setting a certain standard deviation, temperature drift coefficient and other indicators through indoor and outdoor temperature collectors. Finally, the above experimental results are analyzed and summarized.

4. Experimental Analysis of Water Pollution Prevention Engineering Model Based on Small Sample Learning and Data Fusion

4.1. Test and Analysis of Water Pollution Prevention Engineering Model Based on Small Sample Learning and Data Fusion

After comparing the methods of small sample learning and data fusion, it is found that the combination of the test results of the water pollution prevention engineering model based on the large sample model with the methods of not using a wide range of coverage, multi factor simultaneous capacity increase and mixed weighting can achieve better results more comprehensively and accurately. Table 1 shows the test data of water pollution prevention engineering model.

Table 1. Water Pollution prevention and control engineering model test

Parameter	Mid-value	Mean value	Scope
Deposit sediment	2.4	3.12	2.1—4.3
Rate of flow	4.1	4.7	3—5.78
Temperature	3.2	3.1	2.5—4.1
Electroconductibility	1.24	1.53	1.06—2.41
Turbidity	4.31	3.4	2.53—5.2

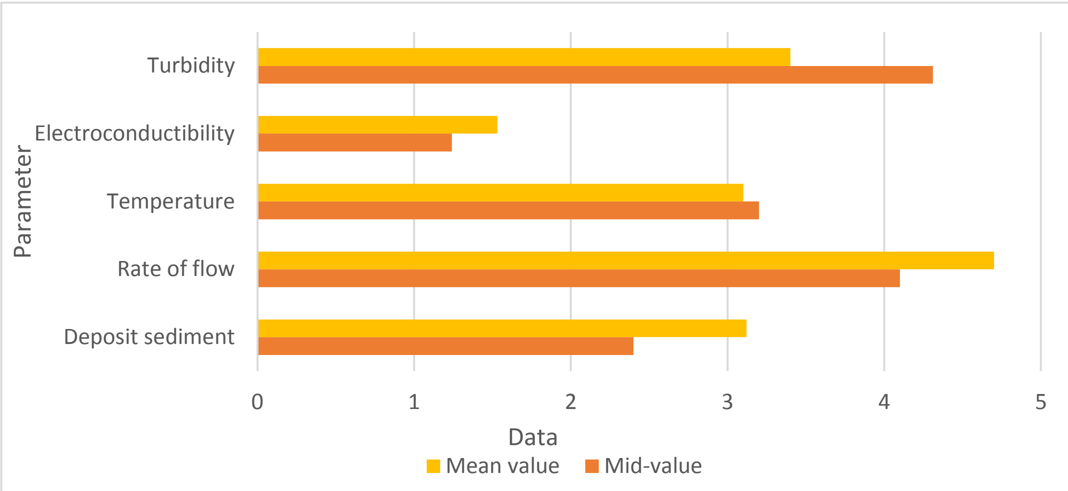

Figure 4. Engineering model of water pollution prevention and control from the perspective of small sample learning and data fusion

Through the test, we can intuitively understand the indicators that will affect the target requirements of the treatment project and water quality standards in the small sample data, and adjust these indicators to achieve their ideal control quantities. The following conclusions can be drawn from the test results in Figure 4. Based on small sample learning and data fusion technology, it has a certain effect on water pollution prevention projects and can monitor water pollution. However, the problems found in actual production need to be solved, which also shows that we need to optimize the governance plan. For example, when sampling the water body, there will inevitably be some errors, improper selection of the sampling point and other circumstances, we should find a reasonable area to avoid excessive errors that affect the follow-up work. At the same time, we can also use the database to update real-time information, upload the latest data and historical analysis results to ensure that there are no errors in the operation of our system.

5. Conclusion

The development of water pollution prevention project in China is at the initial stage, so it is of great significance to conduct in-depth research on it. This paper constructs a complete water environment governance scheme based on the perspective of impact from the perspective of small samples. First, put forward the problem and background; Secondly, analyze which factors should be paid attention to in the establishment of the model, such as data quality, learning speed, etc., and then use the fusion algorithm to calculate the distance between the optimized planning units into the standard difference value and compare it with the adjacent planning units to find out the improvement method to achieve the optimal effect.

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]Nadia Abu Farha, Fares Al-Shargie, Usman Tariq, Hasan Al-Nashash. Brain Region-Based Vigilance Assessment Using Electroencephalography and Eye Tracking Data Fusion. IEEE Access. (2022) 10: 112199-112210. https://doi.org/10.1109/ACCESS.2022.3216407

[2]Noha S. Fayed, Mohammed M. Elmogy, Ahmed Atwan, Eman M. El-Daydamony. Efficient Occupancy Detection System Based on Neutrosophic Weighted Sensors Data Fusion. IEEE Access 10: 13400-13427. https://doi.org/10.1109/ACCESS.2022.3146346

[3]Sathwik Reddy Majji, Avinash Chalumuri, Raghavendra Kune, B. S. Manoj:Quantum Processing in Fusion of SAR and Optical Images for Deep Learning: A Data-Centric Approach. IEEE Access 10: 73743-73757 (2022). https://doi.org/10.1109/ACCESS.2022.3189474

[4]Ahmet Ozcan, Omer Cetin:A Novel Fusion Method With Thermal and RGB-D Sensor Data for Human Detection. IEEE Access 10: 66831-66843 (2022). https://doi.org/10.1109/ACCESS.2022.3185402

[5] Rutuja Rajendra Patil, Sumit Kumar:Rice-Fusion: A Multimodality Data Fusion Framework for Rice Disease Diagnosis. IEEE Access 10: 5207-5222 (2022). https://doi.org/10.1109/ACCESS.2022.3140815

[6]Nowshin Tasnim, Iftekher Toufique Imam, M. M. A. Hashem:A Novel Multi-Module Approach to Predict Crime Based on Multivariate Spatio-Temporal Data Using Attention and Sequential Fusion Model. IEEE Access 10: 48009-48030 (2022). https://doi.org/10.1109/ACCESS.2022.3171843

[7] Hamza Osman Ilhan, Gorkem Serbes, Nizamettin Aydin:Decision and feature level fusion of deep features extracted from public COVID-19 data-sets. Appl. Intell. 52(8): 8551-8571 (2022). https://doi.org/10.1007/s10489-021-02945-8

[8]Matthias Schreier:Data fusion for automated driving: An introduction. Autom. 70(3): 221-236 (2022). https://doi.org/10.1515/auto-2021-0132

[9] A. Vidhya, P. Mohan Kumar:Fusion-based advanced encryption algorithm for enhancing the security of Big Data in Cloud. Concurr. Eng. Res. Appl. 30(2): 171-180 (2022). https://doi.org/10.1177/1063293X221089086

[10] K. M. A. Solaiman, Tao Sun, Alina Nesen, Bharat K. Bhargava, Michael Stonebraker:Applying Machine Learning and Data Fusion to the "Missing Person" Problem. Computer 55(6): 40-55 (2022). https://doi.org/10.1109/MC.2022.3145507

[11] Andry Rustanto, Martijn J. Booij:Evaluation of MODIS-Landsat and AVHRR-Landsat NDVI data fusion using a single pair base reference image: a case study in a tropical upstream catchment on Java, Indonesia. Int. J. Digit. Earth 15(1): 164-197 (2022). https://doi.org/10.1080/17538947.2021.2018057

[12] Gleb Beliakov, Marek Gagolewski, Simon James:Hierarchical data fusion processes involving the Möbius representation of capacities. Fuzzy Sets Syst. 433: 1-21 (2022). https://doi.org/10.1016/j.fss.2021.02.006

[13]Pawel Drozdowski, Fabian Stockhardt, Christian Rathgeb, Christoph Busch:Signal-level fusion for indexing and retrieval of facial biometric data. IET Biom. 11(2): 141-156 (2022). https://doi.org/10.1049/bme2.12063

[14]Gayatri Umakant Bokade, Rajendra D. Kanphade: Template security scheme for multimodal biometrics using data fusion technique. Int. J. Biom. 14(2): 166-190 (2022). https://doi.org/10.1504/IJBM.2022.121801

[15]Wedad Alahamade, Iain Lake, Claire E. Reeves, Beatriz de la Iglesia: A multi-variate time series clustering approach based on intermediate fusion: A case study in air pollution data imputation. Neurocomputing 490: 229-245 (2022). https://doi.org/10.1016/j.neucom.2021.09.079

[16]Leila Arras, Ahmed Osman, Wojciech Samek:CLEVR-XAI: A benchmark dataset for the ground truth evaluation of neural network explanations. Inf. Fusion 81: 14-40 (2022). https://doi.org/10.1016/j.inffus.2021.11.008

[17]P. V. Arun, Roy Sadeh, A. Avneri, Yaniv Tubul, Carlos Camino, Krishna Mohan Buddhiraju, A. Porwal, Ran N. Lati, Pablo J. Zarco-Tejada, Zvi Peleg, Ittai Herrmann:Multimodal Earth observation data fusion: Graph-based approach in shared latent space. Inf. Fusion 78: 20-39 (2022). https://doi.org/10.1016/j.inffus.2021.09.004

[18]Yunus Emre Çelik, Samuel Stuart, Wai Lok Woo, Ervin Sejdic, Alan Godfrey:Multi-modal gait: A wearable, algorithm and data fusion approach for clinical and free-living assessment. Inf. Fusion 78: 57-70 (2022). https://doi.org/10.1016/j.inffus.2021.09.016