Water Quality Prediction using Support Vector Machine in Wireless Sensor Networks

GOPU LAKSHMI ANJANA
M.Tech Student, Department Of CSE,
Gudlavalleru Engineering College, Gudlavalleru,
Andhra Pradesh

CH.SURESH BABU
Associate Professor, M.Tech, Ph.D., Department of CSE, Gudlavalleru Engineering College,
Gudlavalleru, Andhra Pradesh

Abstract

In Aquaculture, the yields of the aquatic organism depend on the quality of water. To collect the data from the pond like temperature, dissolved oxygen, pH level, turbidity, carbon dioxide the sensors are placed inside the pond**.** The detection of water quality can be done using the data classification algorithms. In this project, we have proposed a Support Vector Machine (SVM) classifier to predict the quality of water in wireless sensor networks. The sensor nodes are placed in the pond. Here cluster head based routing protocol is used based on the fractional calculus artificial bee colony algorithm, in which the individual decision trees are merged along the routing path. Then, the results of cluster head-based routing protocol are sent to sink node, in which the proposed SVM classifier is used to classify the water quality parameter. From the results, we proved that, the proposed algorithm achieves the better prediction accuracy as 85%.

Keywords: Aquaculture, Water quality parameters, cluster head routing protocol, fractional calculus artificial bee colony algorithm, wireless sensor node, Support Vector Machine.

INTRODUCTION

Nowadays, aquaculture industry is growing fast because of the huge requirement of sea foods across the globe. The aquaculture plays one of vital role in the economic field. The yields of the aquatic organisms such as fishes, prawns etc. depends on the quality of the water in the aquaculture pond. In order to get the maximum yield, the water quality parameters such as temperature, dissolved oxygen, pH level, turbidity, carbon dioxide, diphosphorus should be maintained in the optimum level. Basically, water sources are easily affected by bacterial formation and other harms. Such tribulation reduces the quality of the water. In order to improve the quality of water, it needs to be monitored frequently. Moreover, the level of the water should be also monitored properly to protect the aqua growth. Water level prediction is also another one significant task in aquaculture, which can be done by various theoretical and deterministic models. These models are used to collect the accurate data based on the following factors, such as rainfall depth, evaporation rate, type of soil, features of land used and so on. In order to predict the quality and level of water, several techniques, such as data mining, artificial intelligence or machine learning techniques are used. These techniques estimate the water level by making use of small number of variables.

In this paper, the quality of water is predicted using Support Vector Machine in Orange tool. Here, the wireless sensor nodes are placed in the water source to sense the water quality parameters, such as temperature, dissolved oxygen, pH level, turbidity, electrical conductivity, nitrate, calcium, total hardness, total alkalinity, ammonia, hydrogen sulphide, biochemical oxygen demand, carbon dioxide, and diphosphorus. After sensing the water quality parameters, the wireless sensor node transmits the sensed data to sink node. While transmitting the data, in order to reduce the transmission loss and delay, the optimal cluster head is generated based on the multi-objective fractional artificial bee colony algorithm. Here, the classification of data is based on Support Vector Machine (SVM). SVM are based on the idea of finding a hyper plane that best divides a data set into two classes. The quality of water can be predicted based on the results of SVM classifier. The main contribution of this paper is given as follows:

- Development of a new support vector machine (SVM) for data classification.
- Application of the SVM to predict the status of aqua pond using WSN.

LITERATURE REVIEW

In the table.1 shows the review and comparison of various data classification methods in wireless sensor networks for predicting the quality of water. Yoann Pitarch *et al.* [7] have proposed sequential pattern-based classification method to identify the temporal behavior of the water. In order to handle the uncertain data, G. Vijay Suresh *et al.* [6] have developed rough set theory-based classification, which has been used to collect the data from the medical field. The derivative-based classification method is proposed by Arthi Simon *et al.* [5], which has been utilizes the combination of reflectance band difference/ratios and derivative signatures for classification. The data classification in wireless sensor network is one of the major challenges in various applications such as wild life monitoring, precision agriculture etc. To achieve the highest classification accuracy with very low storage and communication overhead in wireless sensor networks, Xu Cheng *et al.* [28] implemented the classification method based on the hierarchical distributed data. In order to deal with the sheer volume of data the distributed data mining classifier is used in the wireless sensor networks, which have been proposed by Stasa Vujcic Stankovic *et al.* [26]

Authors	**Methods**	**Advantages**	**Disadvantages**
Stasa Vujcic Stankovic *et al.*[26]	Distributed data mining classifier	used to deal with the sheer volume of data produced by wireless sensor networks	the wireless sensor nodes affected based on the problems of resource constraints such as limited power, CPU storage capacity, and communication bandwidth
Asmaa Fawzy *et al.* [27]	Outlier detection method	dealing with large scale datasets and achieves the high accuracy rate for identifying outliers	It limits the performance of accuracy based on the larger dataset
Xu Cheng *et al.*[28]	Hierarchical distributed data classification	high classification accuracy with very low storage and communication overhead	Generated pseudo data as close to the original as possible with limited side information.
Jong P. Yoon *et al.* [29]	data mining technique for sensor collaboration	It reduces the false negatives, which improves the monitoring and detecting capability of WSNs.	Large datasets affects the quality of classification
Arthi Simon *et al.* [5]	Derivative-based classification	Utilizes the combination of reflectance band difference/ratios and derivative signatures	classification accuracy using MODIS aqua data is reduced by large in-situ data sets
G. Vijay Suresh *et al.* [6]	Rough set theory-based classification	Handling of uncertain data	the large datasets stored in database systems limits the classification accuracy
Yoann Pitarch *et al.* [7]	Sequential pattern-based classification	Consideration of temporal behavior	this classification algorithm does not manage the imbalanced data classes
Ch. Suresh Babu et al.[31]	Functional tangent decision tree	Improves prediction quality	Dataset are imbalanced classes, more missing values in most of fields, uncertainty attributes and values, Low sensitivity, specificity and accuracy
Ch. Suresh Babu et al.[32]	Distributed functional tangent decision tree	Improve the accuracy	Can be expensive to setup

Table1. Comparison of various data classification algorithms

2.1. Challenges:

Maintaining water quality is one of the important processes in farming aquaculture organisms, due to the extensive demand on aquaculture. Basically, the aqua organisms are very sensitive to water condition, so the prediction of water level for the future will be essentially need for the aqua farmers to make the effective decisions.

The water quality classification is based on the nature of datasets available. Mostly, the imbalanced classes of data and more absent values of classes present in the nature of datasets. So, the classification based on the imbalanced and missing attributes is one of the important challenges present in the water quality prediction. Then, another one important problem is data uncertainty which is one of the common problems in real-world applications due to imprecise measurement, network latency, out-dated sources and sampling errors.

The other key challenge of WSNs is the data classification and collection of data from the sensor nodes. The data transmission from the sensor nodes are affected based on the power consumption, computation capability and number of available dead nodes [19].

Basically, the lifetime of a sensor nodes are directly determines the duration of the sensing task, which can be limited by the amount of energy each sensor has [30].

SYSTEM OVERVIEW

From the system overview, the proposed water quality prediction method can be explained as follows: Here, the numbers of sensor nodes are kept inside the particular water source. From that particular water source, various water quality parameters collected by different sensor nodes. For example, the pond 1 contains 14 sensor nodes, which can collect 14 kinds of water quality parameters from the pond 1. Meanwhile, the other sensor nodes also collect the 14 kinds of water quality parameters from the other water sources. Here, we have shown the 5 ponds, in which the 14 kinds of water quality parameters collected by 14 sensors. After collecting the parameters, the sensor nodes send the information to sink node through the cluster head. Then, B is considered as base station, which is used to collect the information from the all cluster heads. At first, the sensor nodes are fixed with the water source to extract the water quality parameters such as temperature (temp), dissolved oxygen(DO), pH, turbidity (Turb), electrical conductivity (EC), nitrate (NO_3^-), calcium (Ca^{2+}), total hardness (TH), total alkalinity (TA), ammonia (NH_3), hydrogen sulphide (H_2S), biochemical oxygen demand (BoD), carbon dioxide (Co_2), diphosphorus (P_2). After extracting the water quality parameters the sensor nodes are used to transfer the information to base station

Basically, while transferring the data from one sensor node to another sensor node, the energy consumption is one of the major problems based on the both amount of transmitting data and the transmitting distance. In order to overcome this challenge, the cluster head routing protocol is used to transfer the data packets from sensor node to base station, in which the data transmission can be occurred with shortest path using better routing protocol. In this paper, different cluster head routing protocols are used transfer the data from sensor nodes to base station to overcome the challenges based on the delay, energy, power consumption. The cluster head is used to collect all the sensed data from the surrounding sensor nodes. After collecting data, the classification of data can be performed based on the SVM classifier to improve the classification performance.

3.1. Network model:

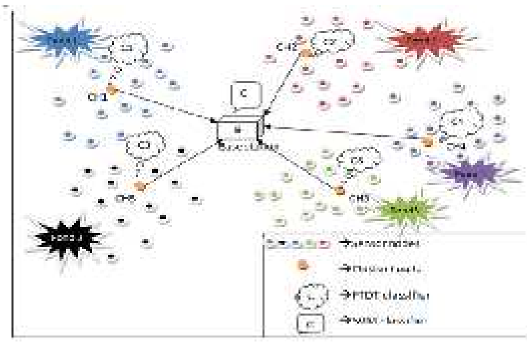


Fig1: Network Model

This section shows the network model [25] of the proposed system, which consists of n number of sensor nodes with only one base station B. Here, the n numbers of sensor nodes are placed within the water source to collect the information. Basically, the communications between the sensor nodes are available within the radio range based on the wireless link representation. In order to attain the maximum communication within the radio range, each and every sensor nodes are homogeneously distributed inside the dimensions of P and Q meters. Then, the sensor nodes are collecting the quality water from the nearest pond. After collecting the water quality parameters, the sensor nodes are grouped into clusters based on the own unique ID of sensor nodes. Then, the base station of sink node is located in the dimensions of 0.5P, 0.5Q, which is fixed nearer to the optimal cluster head location to receive all the data information from the sensor nodes. Then, the cluster head-based routing mechanism is used to transfer the data from every sensor node to base station. Basically, the selection of the cluster head is based on the following conditions: initially, check whether the

Volume No.6, Issue No.2, February - March 2018, 7988-7996.

distance among the cluster members with the particular node is less or not. If the distance is less, that particular node is selected as the cluster head. Furthermore, the energy of the cluster head can be calculated, in which the calculated energy should be as high as possible for cluster head. Moreover, the delay of the cluster head should be less to transmit the data. Based on the above three constraints, the cluster head can be evaluated to perform the better data transmission. Here, CH1 is a cluster head representation of sensor nodes that is used to collect the water quality parameters from the pond 1. Accordingly, CH2, CH3, CH4 and CH5 are the cluster head-based routing protocol for the wireless sensor nodes that contained the water quality information based on the pond 2, pond3, pond4 and pond 5.

Once, the information is collected by the cluster head that is used to send the information to base station. Here, based on the free space and multi-path fading model [31], the data packets are sending from normal sensor node to cluster node with the loss of energy. Here, the transmission of data is performed based on the distance between the transmitter and receiver. Then, the transmitter is used to dissipate energy based on the radio electronics and power amplifier. Based on the data transmission distance and nature of the normal node and header node, the energy can be dissipated from the every packet of the data. The energy dissipation of the normal sensor node can be represented as follows:

$$E_d(N_s) = E_e * B + E_p * B * \left\| N_s^k - CH_j \right\|^4 \quad if \quad \left\| N_s^k - CH_j \right\| \geq d_0$$

$$E_d(N_s) = E_e * B + E_s * B * \left\| N_s^k - CH_j \right\|^2 \quad if \quad \left\| N_s^k - CH_j \right\| < d_0$$

Where, B is defined as data bytes that is send from normal sensor node, E_p is defined as the power amplifier based energy parameter, $\left\| N_s^k - CH_j \right\|$ is defined as the distance between the normal sensor node to cluster head, E_e is defined as electrical energy based on different factors such as modulation, filtering and amplifiers, which can be represented as follows:

$$E_e = E_t + E_{da}$$

Where, E_t are transmitter energy and the data aggregation energy can be represented by E_{da}. When the cluster node CH receives the B bytes of data, the dissipation of energy through the cluster head by receiver can be represented as follows:

$$E_d\left(CH_s^k\right) = E_e * B$$

After sending or receiving B bytes of data, the value of energy in the every node can be updated as follows:

$$E_{t+1}\left(N_s^k\right) = E_t\left(N_s^k\right) - E_d\left(N_s^k\right)$$

$$E_{t+1}\left(CH_s^k\right) = E_t\left(CH_s^k\right) - E_d\left(CH_s^k\right)$$

The above process of data transmission is continuous till the dead node formation. When, the energy of node is less than zero, that particular node is said to be dead node.

3.2. Routing protocol:

This section shows the detailed view of routing protocol based on the FABC-enabled cluster head selection. Here, the optimal selection of cluster head can be performed in three ways such as initialization, FABC for cluster head selection and energy depletion and termination [25].

Step 1: initialization:

The steps involved in the initialization process can be explained as follows:

i) At first, the sensor nodes are distributed within the network area then the energy of the distributed sensor nodes are initialized as E_0.

ii) The centralized clustering based routing strategy is performed based on the location of every sensor nodes within the network area. Here, the location every sensor node is known at the sink node or base station.

iii) Then, the cluster head formation can be performed using the FABC algorithm, which is mainly used to select the optimal cluster heads to decrease the energy consumption and delay with the optimal routing.

Step 2: FABC for cluster head selection

iv) At first, the food source vectors are initialized randomly with the searching area, which can be evaluated using fitness value. Here, the fitness function is based on the three parameters such as distance, energy and delay, which can be shown as follows [25]:

v) $$fitness = \sim * f_i^{loc} + \quad * f_i^{energy} + f_i^{delay}$$

vi) Based on the FABC algorithm, the food sources are updated in employed bee phase. Then, the fitness function is used to evaluate the food sources between newly generated and old one using the fitness function. Then, based on the minimum fitness value, the food source is selected.

vii) The above process is continued until the terminal requirement. After selecting the cluster head, which is used to update the energy of the sensor node.

Step 3: energy depletion and termination

viii) Once the cluster head is identified, the data transmission among the normal sensor nodes and cluster head is executed.

ix) Then, the energy of the every node is updated based on the every byte of data transfer.

The above two processing steps are continuous till every the nodes are considered as dead nodes.

3.3. SVM classification

Support vector machine is a classification method that performs classification tasks by constructing a hyper plane. SVM supports both classification and regression and it can handle multiple continuous and categorical attributes.

The classification of SVM is divided into two types

Step 1: For this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N$$

where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs), index i label the N training cases.

Step 2: The Classification SVM Type 2 model minimizes the error function:

$$\frac{1}{2} w^T w - \nu\rho + \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \ldots, N \text{ and } \rho \geq 0$$

RESULTS AND DISCUSSION

This section shows the experimental results of the proposed Support Vector Machine using Orange tool. Here, the proposed Support Vector Machine is implemented using Orange tool with i3 processor of 4GB RAM, running in windows 8.1.

3.4. Simulation set up of aqua data

This section shows the simulation set up of aqua data to predict the quality of water using various physicochemical parameters such as temperature (temp), dissolved oxygen(DO), pH, turbidity (Turb), electrical conductivity (EC), nitrate (NO3-), calcium (Ca2+), total hardness (TH), total alkalinity (TA), ammonia (NH3), hydrogen sulfide(H2S), biochemical oxygen demand (BoD), carbon dioxide (Co2), diphosphorus (P2) [16-18]. By analyzing the above 14 water quality parameters, we can easily identify the nodes are in danger or not. Here, the 14 water quality parameters are collected from 21 ponds. Then, the sample dataset collected from the 21 ponds are given in table 2. The temperature of ground water is maintained at 25 ^{o}C.

Temp	**DO**	**pH**	**Turb**	**EC**	**NO3-**	**Ca2+**	**TH**	**TA**	**NH3**	**H2S**	**BOD**	**CO2**	**P2**	**WQ**
21	5.1	7.3	80	2209	35.1	213	713	211	0.37	0.2	3	13	0.9	0
20	6.1	7.4	70	1779	8.3	172	516	212	0.45	0.1	4	31	1.5	0
20	4.3	8.4	65	2076	10.2	188	582	210	0.49	0.08	5	12	1.7	0
20	4	9.	54	297	46.	21	80	16	0.1	0.0	3.	5	1.	0

		2		5	2	2	5	2	5	7	6		6	
20	4.2	9.1	564	2381	35.1	181	629	158	0.16	0.06	4	6	2	1

Table 2: Sample dataset

4.2. Metrics for evaluation

This section shows the performance evaluation, based on the danger and not danger nodes for evaluating the performance of the water quality parameters. The accuracy is measured by using confusion matrix. Confusion matrix is used to describe the performance of the classification model on a set of test data for which the true values are known.

$$Accuracy = {(TN + TP)} \Big/ {(TN + FP + FN + FP)}$$

Where, the true negative value can be represented as TN, which is also defined as the correctly rejected data. Then, the incorrectly identified data can be defined as false positive FP.

4.3. Experimental results

This section shows the experimentation results of water quality prediction using proposed SVM in wireless sensor networks. Here SVM is used for classification. SVM are based on the idea of finding a hyper plane that best divides a dataset into two classes. Here, the below figure shows that the dataset will often look more like the jumbled nodes below which represent a linearly non separable dataset.

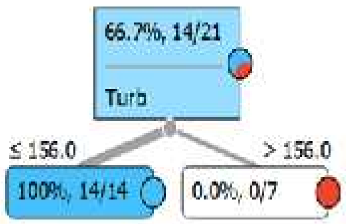
While performing the operation, the dataset is divided into two classes in which the nodes are colored and are separated. By using the scatter plot, here we will compare two attributes. Water quality attribute is considered as a target element. Water quality is compared with other attribute to know whether which nodes are in danger and which nodes are not in danger. Here we have to check which nodes are in danger or not danger. In the below figure blue color nodes i.e., 0.0 represent that the nodes are not in danger and red color i.e., 1.0 nodes represent that the nodes are in danger.

A)

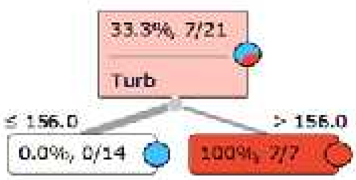
B)

Fig 2: sensor nodes at **a** not danger **b** danger

Tree is a straightforward calculation that parts the information into hubs by class virtue. It is a forerunner to Random Forest. Tree in Orange is outlined in-house and can deal with both discrete and consistent datasets. It can likewise be utilized for both characterization and relapse assignments. The choice tree is built for both risk and not threat hubs. The below figure, represents the decision tree for danger and not danger nodes. After generating the decision tree, the tree results are sent to the sink node. Basically, the sink node is used to merge all the trees, which have been collected from the cluster head using the proposed SVM. Finally the tree generated from the sink node is used to predict the quality of water for aquaculture.



a) Decision tree for 0.0



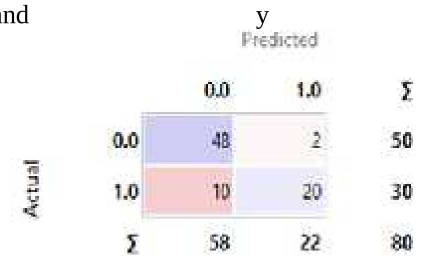
b) Decision tree for 1.0

Fig 3: Decision tree for not danger (0.0) and danger nodes (1.0)

To find accuracy the proposed SVM uses test score and accuracy. Test & Score gets the data from File and SVM. It performs cross-validation or some other train-and-test procedures to get class predictions by both algorithms for all (or some) data instances. The test outcomes are nourished into the Confusion Matrix.

The Confusion Matrix gives the number/extent of occurrences between the anticipated and the real class. Confusion matrix contains the information about the actual and prediction class based on classification problem. A test dataset is required with the expected outcome. Prediction is done for each row in the test dataset.

The Scatter plot widget provides 2-dimensional scatter plot visualization for both continuous and discrete-valued attributes. The data is displayed as a collection of points, the x-axis determines the position on the horizontal axis and the y-axis determines the position on the vertical axis. Properties of the graph such as, like color, size and shape of the points, axis titles, maximum point size and jittering are on the left side of the widget. The Scatter plot gets two sets of data. The file widget sends the complete information to the scatter plot, while the confusion matrix sends only the selected data, misclassifications for instance. The scatter plot shows all the data with symbols. Water quality (WQ) and pH are considered on x and y axis.



Actual \ Predicted	0.0	1.0	Σ
0.0	48	2	50
1.0	10	20	30
Σ	58	22	80

Fig 4: Accuracy using confusion matrix

The below figure, shows the comparison between DFTDT and SVM. From the fig, we can conclude, SVM provides more accuracy than DFTDT.

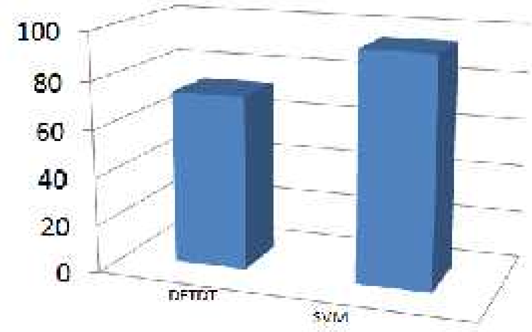


Fig 5: Comparison of DFTDT and SVM

CONCLUSION

In this paper, we presented the Support Vector Machine in wireless sensor networks to predict the quality water. A first, the number of sensors can be located in ponds to collect the required water quality parameters such as temperature (temp), dissolved oxygen(DO), pH, turbidity (Turb), electrical conductivity (EC), nitrate (NO3-), calcium (Ca2+), total hardness (TH), total alkalinity (TA), ammonia (NH3), hydrogen sulfide(H2S), biochemical oxygen demand (BoD), carbon dioxide (Co2), diphosphorus (P2). Here, we have considered 14 water quality parameters, which can be collected from the 21 ponds by research experts. After sensing the water quality parameters, the functional tangent decision tree is created and the merging process can be done in cluster head-based routing protocols. Then, the results of cluster head-based routing protocols is send to sink node, in which the proposed SVM algorithm is used to predict the water quality parameters with 85% accuracy. Then, the networking performance can be evaluated using normalized energy consumption with the existing works. Future studies can also discover the solutions to deployment of sensor node, data distribution, and routing in WSNs.

REFERENCES

[1] Yi-Chung Hu, "Rough sets for pattern classification using pairwise-comparison-based tables", Applied Mathematical Modelling, vol. 37, pp. 7330–7337, 2013.

[2] Khadija Al-Aidaroos, Azuraliza Abu Bakar, and Zalinda Othman, "Data Classification Using Rough Sets and Naïve Bayes", In proceedings of international conference on Rough set and knowledge technology pp. 134–142, 2010.

[3] Riaj Uddin Mazumder, Shahin Ara Begum and Devajyoti Biswas, "Rough Fuzzy Classification for Class Imbalanced Data", in Proceedings of Fourth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing, 2015.

[4] K. Srinivas, G. Raghavendra Rao, A. Govardhan, "Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories", Arabian Journal for Science and Engineering, vol. 39, pp. 2857–2868, 2014.

[5] Arthi Simon, Palanisamy Shanmugam, "An Algorithm for Classification of Algal Blooms Using MODIS-Aqua Data in Oceanic Waters around India", Advances in Remote Sensing, vol. 1, pp. 35-51, 2012.

[6] G. Vijay Suresh, E. Venkateswara Reddy, and E. Srinivasa Reddy, "Uncertain Data Classification Using Rough Set Theory ", In Proceedings of International Conference on Information Systems Design and Intelligent Applications, pp. 869–877, 2012.

[7] Yoann Pitarch, Dino Ienco, Elodie Vintrou, Agnès Bégué, Anne Laurent, Pascal Poncelet, Michel Sala, Maguelonne Teisseire, "Spatio-temporal data classification through multidimensional sequential patterns: Application to crop mapping in complex landscape", Engineering Applications of Artificial Intelligence, vol. 37, pp. 91–102, 2015.

[8] Xu E, Shuang Lin and Lulu Jin,"Data Recovery Method for Seafood Quality Safety System Based on Rough Set Theory ", International Journal of Security and Its Applications Vol.8, No.5 (2014), pp.195-202.

[9] C. Li and Y. Huang, "Food Safety Monitoring and Early Warning System", Chemical Industry Press, Beijing, 2006.

[10] Bunchingiv Bazartseren, Gerald Hildebrandt, K.P. Holz, "Short-term water level prediction using neural networks andneuro-fuzzy approach", Neurocomputing, vol. 55, pp. 439 – 450, 2003

[11] Nazleeni Samiha Haron, Mohd Khuzaimi B Mahamad, Izzatdin Abdul Aziz, Mazlina Mehat, "A System Architecture for Water Quality Monitoring System Using Wired Sensors", In Proceedings of International Symposium on Information Technology, vol. 4, pp. 1 - 7, 2008.

[12] Li Nan, Wang Ruimei, Zhang Jian, Fu Zetian, Zhang XiaoShuan,"Developing a knowledge-based early warning system for fish disease/health via water quality management", Expert Systems with Applications, vol. 36, pp. 6500–6511, 2009.

[13] Fabrice Rossi, Nathalie Villa, "Support Vector Machine For Functional Data Classification", Neurocomputing, Vol. 69, no. 7–9, pp.730–742, March 2006.

[14] Pawlak, Z.: "Rough Sets and Intelligent Data Analysis", Information Sciences, vol. 147, pp. 1–12, 2002.

[15] Thangavel, K., Pethalakshmi, "A.: Dimensionality Reduction based on Rough Set Theory: A Review", Applied Soft Computing, vol. 9, pp. 1–12, 2009.

[16] Sami G. Daraigan, Ahmed S. Wahdain, Ahmed S. Ba- Mosa, Manal H. Obid, "Linear correlation analysis study of drinking water quality data for AlMukalla City, Hadhramout, Yemen", international journal of environmental sciences, vol. 1, no 7, 2011.

[17] Anita Bhatnagar, Pooja Devi, "Water quality guidelines for the management of pond fish culture", Environmental Sciences, vol. 3, no. 6, pp, 1980-2009, 2013.

[18] Mohammad Mehdi Heydari, Ali Abasi, Seyed Mohammad Rohani and Seyed Mohammad Ali Hosseini, "Correlation Study and Regression Analysis of Drinking Water Quality in Kashan City, Iran", Middle-East Journal of Scientific Research, vol. 13, no. 9, pp. 238-1244, 2013

[19] Dilek Karabudak,"Security Optimization and Data Classification in Wireless Sensor Networks",http://www.cs.bilkent.edu.tr/~guvenir/courses/CS550/Workshop/Dilek_Karabuda k.pdf

[20] Tsang, S., Ben Kao ; Yip, K.Y. ; Wai-Shing Ho ; Sau Dan Lee, "Decision Trees for Uncertain Data ", IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 1, pp. 64-78, 2009.

[21] Sasikanth Avancha, Amit Baxi, David Kotz, "Privacy in mobile technology for personal healthcare", ACM Computing Surveys, vol. 45, no. 1, 2012.

[22] Jong P. Yoon, "Three-tiered data mining for big data patters of wireless sensor networks in medical and healthcare domains", In proceedings of international conference on Internet and Web Applications and Services, pp. 18 to 24 , 2013.

[23] Luca rizzon, Maurizio Rossi, Roberto Passerone, Davide Brunelli, "Wireless sensor networks for environmental monitoring powered by microprocessors heat dissipation", Energy Neutral Sensing Systems, no.8, 2013.

[24] Vladimir Dyo, Stephen ellwood, David Macdonald, Andrew Markham, Niki Trigoni,et. al., Wildsending: Design and deployment of a sustainable sensor network for wildlife monitoring, ACM Transactions on Sensor Networks, vol.8, no. 4, 2012.

[25] Rajeev Kumar, Dilip Kumar, "Multi-objective fractional artificial bee colony algorithm to energy aware routing protocol in wireless sensor network", Wireless Networks, pp. 1-14, 2015.

[26] Stasa Vujcic Stankovic, Ooran Rakocevic, Nemanja Kojic, Dragan Milicev, "A Classification and Comparison of Data Mining Algorithms for Wireless Sensor Networks", In Proceedings of International Conference on Industrial Technology, pp. 265-270, 2012.

[27] Asmaa Fawzy, Hoda M.O. Mokhtar, Osman Hegazy, "Outliers detection and classification in wireless sensor networks", Egyptian Informatics, vol. 14, no. 2, pp.157–164, 2013.

[28] Xu Cheng, Ji Xu, Jian Pei, Jiangchuan Liu, "Hierarchical distributed data classification in wireless sensor networks", Computer Communications, In Proceedings of International Conference on Mobile Adhoc and Sensor Systems, pp. 10 - 19, 2009.

[29] Jong P. Yoon, Juan Ortiz, "Data Mining Approach to Situation-aware Sensor Actuation in Wireless Sensor Networks", In proceedings of International Conference on Future Generation Communication Technologies, pp. 1 - 6, 2015.

[30] Enrique J. Duarte-Melo, Mingyan Liu, "Analysis of energy consumption and lifetime of heterogeneous wireless sensor networks", In proceedings of international conference on IEEE Global Telecommunications, vol. 21, pp. 21 - 25, 2002.

[31] CH. Suresh Babu "FTDT: Rough set integrated functional tangent decision tree for finding the status of aqua pond in aquaculture", In proceedings of Journal of intelligent and Fuzzy Systems

[32] CH. Suresh Babu "DFTDT: distributed functional tangent decision tree for aqua status prediction in wireless sensor network**",** In proceedings of International Journal of Machine Learning and Cybernetics

