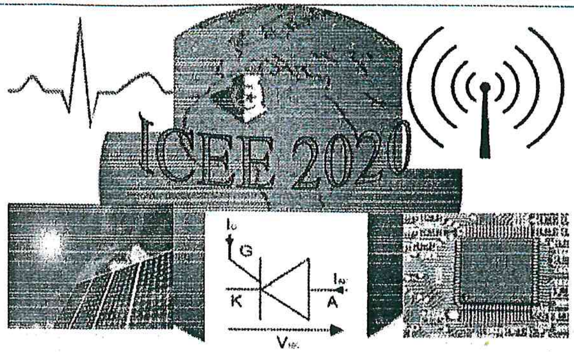

IEEE

Signals & Systems Laboratory
LSS
Institute of Electrical and Electronic Engineering
University of Boumerdes

The 6th International Conference on Electrical Engineering- ICEE'2020
September 25th to 27th, 2020, Istanbul, Turkey

Attendance and Presentation Certificate

This is to certify that **Mohamed Ladjal** has attended the sixth International Conference on Electrical Engineering (ICEE2020) held online on September 25-27, 2020 and presented the paper entitled: **Optimization of SVM parameters with hybrid PCA-PSO methods for water quality monitoring**

Co-authors: **Ouali Mohammed Assam; Mohamed Djerioui**

The sixth International Conference on Electrical Engineering
ICEE-2020

Conference General Chair
Prof.Dr. Abdelmadjid RECIOUI

2020 International Conference on Electrical Engineering (ICEE)
September 25-27, 2020, Istanbul, Turkey

Optimization of SVM parameters with hybrid PCA-PSO methods for water quality monitoring

Mohamed LADJAL
LASS, Laboratoire d'Analyse des Signaux et Systèmes
Department of Electronics, University of M'sila,
M'sila, Algeria
mohamed.ladjal@univ-msila.dz

Mohammed Assam OUALI and Mohamed DJERIOUI
LASS, Laboratoire d'Analyse des Signaux et Systèmes
Department of Electronics, University of M'sila,
M'sila, Algeria
mohamedassam.ouali@univ-msila.dz,
mohamed.djerioui@univ-msila.dz

Abstract— **For the development of a water quality modeling classification, parameter optimization is important. In this research, in order to enhance the strength of the used approach, we propose a hybrid approach that combines SVM classifiers with PSO and PCA selection features. This is used for classifying the status of water quality with the Radial Basis Function (RBF) SVM kernel. To enhance the classification accuracy, PSO selects the best parameter for SVM. The problem of irrelevant data in the space of functions can be solved by PCA. A binary classification based on two water quality classes (Class I: upper, Class II: lower) is considered to be the problem. Datasets were obtained for training and testing over 5 years (2014-2018) from many samples in Tilsdit dam-Algeria, and are used in this situation. A simulation of the training time and recognition rate will be carried out in order to verify the efficiency of the method. The results obtained demonstrate that the proposed method had great potential for classifying water quality.**

Keywords-component; Water quality monitoring, PCA, Features selection, PSO, Optimization, SVM, Classification.

I. INTRODUCTION

An important framework and requirement for water environmental planning and assessment, pollution control, and treatment are to use water quality models to characterize and simulate water quality. The impact of human activities on the natural environment has steadily increased with the rapid expansion of industry and agriculture that occurred after the Industrial Revolution. A significant amount of research has been carried out on the development of water quality models to study water monitoring. Water pollution has become one of humanity's most serious problems. The essential to understanding water quality problems related to regional water protection, food security, and environmental security is an accurate evaluation of regional water quality. The proper assessment and treatment of problems in water quality have a major practical significance for the sustainable use of water resources, clean production, and green development. Surface water quality assessments can reduce decision risks due to water quality uncertainty and may also provide scientific advice on green agriculture and sustainable development. The quality determination of water is a highly dimensional, complex, and variable problem. Several variables are influencing the quality of water. The Water Quality Index (WQI) [1], health risk assessments [2], the Technique for Order Preferences by Similarity to Ideal Solution (TOPSIS)[3], the Fuzzy Comprehensive Assessment[4], and the Gray Relational Analysis [5] are primarily conventional water quality assessment approaches. The approach of machine learning is often considered an eventual solution to the problems of modeling nonlinear processes. Indeed, several applications for water quality assessment and classification based on such methods have been developed [6]. To solve classification problems, methods using Support Vector Machines (SVM) have been widely used. This approach has been documented successfully in various applications [7].

The goal of this study is to evaluate the application of support vector machines (SVM) to the water quality classification of the Tilesdit dam in Algeria in combination with the technique of particle swarm optimization (PSO). SVM is a set of related methods of supervised learning used for classification. The particle swarm optimization (PSO) technique was successfully used here to carry out the optimization process corresponding to the kernel optimal hyperparameters set in SVM training. PSO is one of the oldest bio-inspired algorithms based on swarm intelligence (SI) that was introduced by Kennedy and Eberhart [8]. It is a population-based search algorithm based on the bird flocking simulation [9]. Hybrid PSO Optimized SVM models [10] were used as automated learning tools for the above-mentioned purpose, training them to classify water quality. In order to achieve a good result, the preparation of data inputs involves special treatment to ensure that the classifier makes a good decision. To build the best preparation for data inputs, several methods have been developed. Recently, significant attention has been given to the use of feature selection for data preparation previous to input into the classifier [11]. Therefore, we need features selection to avoid the redundancy. Most of these approaches, like Principal Component Analysis (PCA), are based on linear approaches. The technique can solve the problem of irrelevant data in the features space. The application of PSO and PCA combined with SVM is proposed in the present study to carry out the classification of water quality. The PCA is used here as features selection technique. Then, to search for the optimal parameters, the PSO is used. Also, the SVM model is used to define the relationship between the water environmental variables based on the measured data to understand the internal rules of the water environments and integrate prior knowledge into the

978-1-7281-8292-6/20/$31.00 ©2020 IEEE

optimization process. For the selection of control parameters, the results may provide guidance. Then, to search for the optimal parameters, the PSO is used. Also, the SVM model is used to define the relationship between the water environmental variables based on the measured data to understand the internal rules of the water environments and integrate prior knowledge into the optimization process. For the selection of control parameters, the results may provide guidance.

II. Study Area And Data Description

The data used for this study is were collected over 5 years (2014–2018) from several samples in Tilsdit dam, which is located at 122 km east of Algiers (as shown in Fig. 1). It is situated 20 km southeast of the department of Bouira (Algeria) in the town of Bechloul.This hypothetical study site was selected to assess the performance of a simulation-optimization model used to design a monitoring system for water qualiy. The Tilisdit dam lies between 35° 13’ 22’’north latitude and 4° 14’ 23’’ east longitude. It has a semi-arid climate and an average rainfall of approximately 440-660 mm / year.

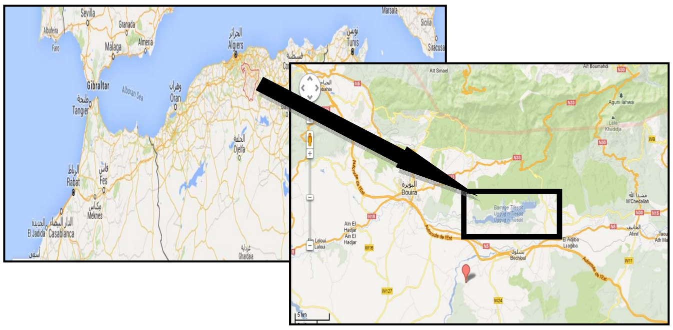

Figure 1. Map showing the study area [Google Maps].

TABLE I. Summary Analysis Of The Physicochemical Variables.

Variables	Min	Max	Mean	Standard deviation
pH	7,15	8,30	7,567	0,25
EC	414,00	624,00	585,393	36,278
T°	9,70	24,20	16,13	3,483
TU	1,320	23,81	3,835	2,392
Mg	7,290	47,628	22,268	4,931
B	158,620	289,14	222,497	23,213
TH	0,00	168,00	32,287	23,029
FTA	130,00	237,00	181,845	18,703

The water quality indicators monitored, including temperature (T °), pH, electrical conductivity (EC), and turbidity (TU), is measured in the field using sensors mounted directly at the station. Subsequently, the samples for their chemical constituents such as calcium, magnesium, chloride, sulfate, bicarbonate, total hardness, permanent hardness, full title alkaline, ammonia, nitrite, color, and free residual chlorine are examined in the laboratory every week. The above-measured data was used to evaluate the relationship between these parameters and to check the monitoring model for water quality. Table 1 presents summary analysis of the selected parameters (physicochemical variables) of surface water.

III. Suggested Methods

In this research, monitoring of water quality may be regarded as a problem of pattern recognition. It usually consists of data collection, signal processing, features selection, and water quality decision. Our proposed approach is based first by preparation of dataset using PCA algorithm to features selection before inputting into classifiers. This phase is done to delete the unnecessary obsolete features. A PSO-optimized SVM classification technique is used to perform the water quality classification process. This new hybrid model based on PCA-PSO-SVM built in this study is shown in Fig.2.

The goal is to classify a water quality in two distinct classes from independent variables. In the following sections, a brief description of the main techniques used in this work, is provided.

A. Principal Component Analysis

PCA is a frequently used approach for reducing the dimensions of multivariate problems. It has found application in fields such as features extraction and selection, cluster analysis, visualization of high dimensionality data, data compression, regression and pattern recognition. Minimum losses in Principal Components (PCs) will be present in the information of input variables in this method.

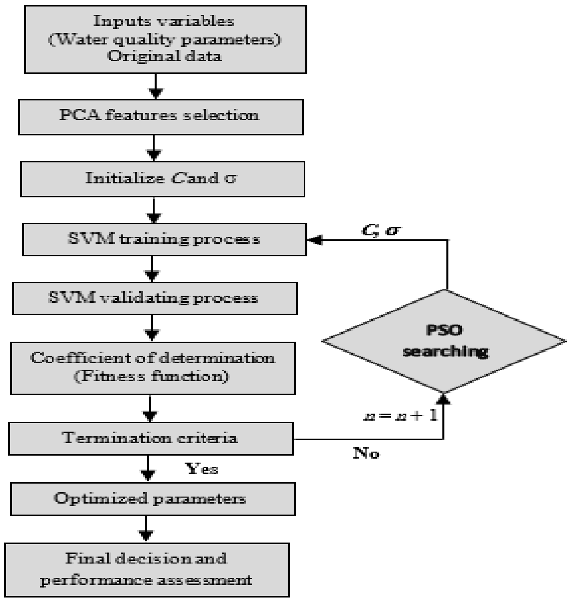

Figure 2. The flowchart of a model based on the hybrid PCA-PSO-SVM.

Several probably correlated variables are transformed into a smaller number of uncorrelated variables called PCs [6] by PCA. All these are orthogonal to each other, so no redundant data is available. PCs can be specified by the following equation [12]:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \ldots + a_{im}x_{mj} \quad (1)$$

Where z_{ij} represent PCs, a_{im} the related eigenvectors and x_{mj} input variables, i is the component number, j is the sample number and m is the total number of variables. This information is achieved by solving the equation [13]:

$$|R - I\lambda| = 0 \tag{2}$$

Where I is the unit matrix, R is the variance-covariance matrix, and λ is the eigenvector.

PCA results are typically discussed in terms of component scores, also referred to as factor scores (the values of transformed variables corresponding to a specific data point) and loadings (the weight by which each original standardized variable should be multiplied to achieve the component score).

The PCs also have the following properties [12]:

- They are uncorrelated.
- They have sequentially maximum variance.
- In the representation of original inputs by the first several PCs, the mean-squared approximation error in this step is minimal.

B. The particle swarm optimization (PSO) algorithm

The PSO Algorithm is an evolutionary optimizing algorithm classified in the bio-inspired algorithms based on the Swarm Intelligence Community (SI) [8], where Np Particle populations or solutions grow each time, moving towards an optimum solution to the issue. Ant Colony's Optimization Method is also a biomedical algorithm of swarm intelligent (SI) [9] and genetic and differential algorithms [10] are bio-inspired algorithms (not Si-based) for optimization where the population's growth is stochastic. In fact, in each iteration, a new population inside the PSO algorithm changes its positions. Every person is affected by his or her trajectory in his or her movement.

C. Support Vector Machines (SVM)

SVM appraoch has been thoroughly evaluated for classification, regression, and estimation of density [14, 15]. The SVM-classifier is now a well-known tool for mapping data to a typically higher dimensional space using kernel methods. Classification is then carried out in this space by creating an ideal linear hyperplane separation (Figure 3). We construct the solution to the problem of quadratic SVM optimization provided by the following data set with a two-class problem that assumes in the feature area the optimal hyperplane:

$$(x_i, y_i), y_i \in \{-1,+1\},\ i = 1,...,n \tag{3}$$

n is the number of observations, x is a distribution in space $\Re^d$, and y_i is the corresponding class label.

The vector w and a constant b are given to determine the optimal separating hyperplane [16]:

$$w.x + b = 0 \tag{4}$$

Under the following conditions:

$$\begin{aligned} (w.x_i) + b &\geq +1, \quad if\ \ y_i = +1 \\ (w.x_i) + b &\leq -1, \quad if\ \ y_i = -1 \end{aligned} \tag{5}$$

The main idea behind the SVM is to optimize the hyperplane margin to achieve good efficiency in the classification. In other words, a quadratic optimization problem that relates to parameters w, b is resolved [17]:

$$\begin{cases} \min\limits_{w} & \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{n}\xi_i \\ \text{with} & y_i(wx+b) \geq 1 - \xi_i \quad \xi_i \geq 0, \quad i = 1,...,n \end{cases} \tag{6}$$

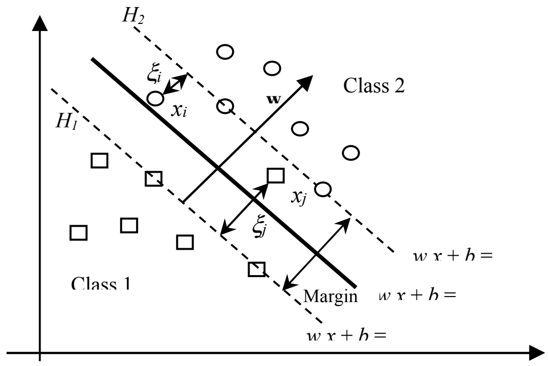

Figure 3. The structure of a basic SVM.

The problem changes with the following dual problem by using the multipliers of Lagrange αi [17, 18]:

$$\begin{cases} \underset{\alpha_i}{Max} & L(\alpha) = \sum\limits_{i=1}^{n}\alpha_i - \frac{1}{2}\sum\limits_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i x_j \\ \text{with} & \sum\limits_{i=1}^{n}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad i = 1,..., n \end{cases} \tag{7}$$

C is a parameter that adjusts the level of classification errors.

The transformation from non-linear to even more dimensional space proceeds using a kernel function [19], defining non-linear mapping from the input space. A dual problem of the Lagrangian SVM [18, 20] is increasing [18, 20]:

$$\begin{cases} \max\limits_{\alpha_i} & L(\alpha) = \sum\limits_{i=1}^{n}\alpha_i - \frac{1}{2}\sum\limits_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \\ \text{with} & \sum\limits_{i=1}^{n}\alpha_i y_i = 0, \ \ 0 \leq \alpha_i \leq C \end{cases} \tag{8}$$

Karush-Kuhn-Tucker's (KKT) rule suggests that an optimal solution α is a necessary and sufficient condition [20]:

$$\alpha_i^0\{y_i[(w_0 x_i) + b_0] - 1\} = 0, i = 1,...,n \tag{9}$$

This means $\alpha_i^0 = 0$ or $y_i[(w_0 x_i) + b_0] = 1$, the latter corresponds to the Support Vectors (SVs), which is equivalent to [20]:

$$SVs = \{x_i \ \text{that} \ \alpha_i \succ 0\} \tag{10}$$

The function of decision is given by [17, 20]:

$$f(x) = sign(\sum_{SVs} \alpha_i y_i K(x_i.x) + b) \tag{11}$$

If *f(x)* is less than 0, then *x* is of class -1; if not, it is of class 1, such as *b* is the solution of the equation.

The selection of the appropriate kernel function is very important.The most used kernel functions are [14, 15, 21] :

The polynomial function:

$$k(x,x') = (\gamma.x^T.x' + c)^d \tag{12}$$

with $c \geq 0$ and $d \in N$

The radial basis function (RBF):

$$k(x,x') = \exp(-\frac{\|x - x'\|^2}{2\sigma^2}) \tag{13}$$

IV. Results And Analysis

A. Features selection

PCA method is used for features selection that contains 80-90% variation of eigenvalues [22]. Also, we may understand that there is a change from data features to uncorrelated components. In this step, the totals of 1200 samples are obtained from eight variables of water quality. These are selected for the analysis because of their continuity of measurement in time scale (Figure 4).

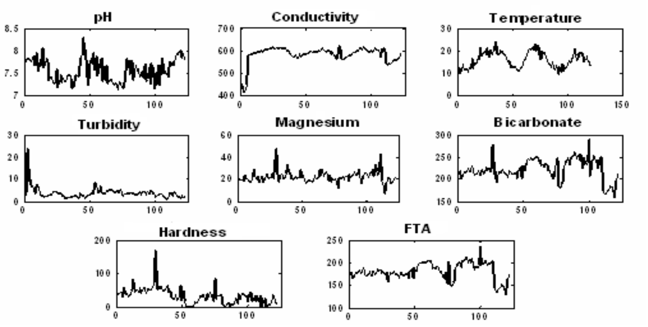

Figure 4. Evolution of the variables of water quality.

By the application of PCA on input variables, a variance-covariance matrix is formed. After solving equation (2), eight eigenvalues are obtained. The results of PCA are presented and summarized in Table 2. In this table, eigenvalues, variance proportion, and cumulative variance proportion are presented. According to these results, it is clear that the four PCs (PC1-PC4) indicate 84.682% of total variance proportion of input variables and the others components are eliminated. This means that the majority of the original data variance was accounted by these PCs.

TABLE II. Eignvalues, Proportion (%), And Coefficients In Four Principal Components

Variables	Principal components			
	PC1	PC2	PC3	PC4
pH	-0.572	-0.464	-0.253	**0.516**
EC	**0.736**	0.30	-0.374	0.081
T°	-0.037	**0.78**	-0.10	-0.446
TU	-0.26	-0.473	**0.729**	-0.275
Mg	0.381	0.479	0.43	0.478
B	**0.867**	-0.411	0.107	-0.028
TH	-0.032	**0.704**	0.419	0.272
FTA	**0.854**	-0.45	0.086	-0.002
Eigenvalues	2.566	2.235	1.124	0.850
Variance proportion (%)	32.07	27.933	14.048	10.63
Cumulative variance proportion (%)	32.07	60.004	74.052	**84.682**

In addition, eigenvectors which asses the coefficients for formation of PCs, are obtained through PCA application. Table 2 presents the correlations of each variable with the principal components acquired. Most effective variables in PCs formation are shown by bold font. The first four principal components together account for 84.682% of the total variance in the dataset, in which the first principal component is 32.07%, the second principal component is 27.933%, the third principal component is 14.048 %, and the fourth principal component is 10.63 % of the total variance. Generally, it is clear that the *Electrical Conductivity*, *Bicarbonate* and *Full Title Alkaline* have the most effect on the PC1 that it includes more than 32% of input variables variance proportions. In addition, the *Temperature* and *Total Hardness* have the most effect on the second component (PC2), which includes more than 27% of input variables variance proportions. Furthermore, PC3 and PC4 are affected by the *turbidity* and the *pH* respectively. We can notice the rapid decay of eigenvalues presented in table 3. The eigenvalues of the first fourth principal components (more than one) can be used to assess the dominant physicochemical processes [39, 40]. The concentrations of *EC*, *B*, and *FTA* show high positive loadings (0.736 - 0.867) whereas concentrations of *Mg* have low positive loadings (0.381) for the first principal component. In the second principal component, *T°* and *TH* have high positive loadings (0.704 – 0.78) and the others concentrations show low positive loading (0.3 - 0.479). In the third principal component, the concentrations of *TU* show high positive loadings (0.729) whereas concentrations of *FTA* have low positive loadings (0.086). For the fourth principal component, the concentrations of *pH* show high positive loadings (0.516), and the *Mg* show moderate positive loadings (0.478), *TU* and *TH* show low positive loadings (0.081 - 0.272).

As the input for the proposed multi-class models, the first four PCs are identified. The retained variables are electrical conductivity (EC), temperature (T °), turbidity (TU), and pH, which are easier to measure continuously and which are the same parameters used in the field using sensors mounted directly at each phase of the production plant. This solution is in any case not final, it will probably be necessary to perform a periodic re-learning system to take account of situations likely to be encountered and to allow continuous adaptation of it to any changes in the water quality.

B. Parameter optimization estimation and classification process

The PSO algorithm has been used to optimize the water quality model's control parameters. It is an evolutionary algorithm based on a population that has been used in different fields for optimization problems [23]. PSO check the best parameter combination like other evolutionary algorithms, by attempting to boost the solution to a specific quality measure. The goal optimization function has shown the precision of the simulated results. For water quality monitoring (WQM) model, the selected input data by PCA are: Temperature (T°), pH, Electrical Conductivity (EC) and Turbidity (TU) selected as input of the proposed classification models. The decision on the status of water quality depends on some variables, which are easier to continually measure. Two classes of water quality (Class I: upper, Class II: lower, etc.) were considered, according to the Environmental Water Quality Standards [24]. To carry out the experiments and test sets, real data concerning different qualitative water conditions is used for training and research. We used a real 1200 sample dataset. The corresponding results are obtained by running the WQM model with different parameter values. To obtain the input-output response relationships between the water quality model parameters and outputs, the data can be used to train the SVM model. Then the parameter optimization can be carried out by the SVM model embedded in the PSO algorithm. Furthermore, it is well known that the SVM approach is highly dependent on the SVM hyperparameters (see Equal (9)), and Ś is a kernel parameter if RBF functions are chosen (see Equal (13)); the choice of hyperparameters for SVMs exists in a vast array of literature [25, 26]. Certain techniques are typically used in determining appropriate parameters: cross-validation (25-27), grid search, random search, Nelder-Mead search, pattern search (25-27).

In other words, a novel hybrid PSO–SVM-based model was applied to classify water quality status (output variable) from the four variables selected by PCA (input variables) in the Tilisdit dam, studying their influence in order to optimize its calculation through the analysis of the MSE with success. Fig. 2 shows the flowchart of this new hybrid PCA-PSO–SVM-based model developed in this study.

As mentioned above, a robust 10-folds cross-validation algorithm was used to ensure the classification capability of the PSO-SVM model [28]. The algorithm mentioned is to divide the datasets into 10 sections and use nine of them to train and the rest to analyze. This experiment has been done 10 folds using the average mistake to assess and measure each group in 10 divisions. Therefore, all possible PSO-SVM parameter variability was evaluated to obtain optimal results and to look for parameters that minimize the average error. The error criterion was determined using 90% of the sample using these best hyperparameters and evaluated using the remaining 10% of the built-in model. This helps one to imagine the actual conditions under which the model is to be constructed so that it can subsequently match new observational data that are unrelated to the construction of the model as much as possible.

SVM has been used for classification modeling using MATLAB code. The search in the space of the parameter was carried out considering that the SVM algorithm significantly changes its results as its parameters increase or reduce its power by 10. We have worked with power ten. The exponents were the 2-dimensional search field [−6, 4] and the searching parameter [−6, 4]. Table 3 indicates the limits (initial areas) of the solution space used in PSO method.

TABLE III. THE INITIAL RANGES OF THE TWO HYPERPARAMETERS OF HYBRID MODEL.

SVM hyperparameters	Lower limit	Upper limit
C	10^{-6}	10^{4}
σ	10^{-6}	10^{4}

There have been 20 Np particles included. If no improvement is achieved in the 10 iterations and the maximum number of iterations equal to 1000, the stop criterion shall be met. In this instance, there were 273 iterations to satisfy convergence. A DELL PC with an Intel Core Processor i7 @ 2,6 GHz, 8 GB RAM was used for this application. For the 297 iterations, the total CPU time used was 736 sec.

The PSO module is used for optimizing the SVM parameters. By evaluating the forecast error on each iteration, the BSO searches for best C and Ţ parameters. There is a two-dimensional search field, one for each parameter. Average error (MSE) is the main fitness factor. Table 4 displays the optimal hyperparameters for the RBF – SVM-based model that has been found using the PSO technique.

TABLE IV. TABLE IV. HYPERPARAMETERS OF THE OPTIMIZED PSO-RBF–SVM-BASED MODEL.

Kernel	Values of optimal hyperparameters	Rec_rate
RBF	Regularization factor C = 19.7080, σ = 0.9650	92%

According to this result, the RBF – Kernel function SVM is the best model for the Tilesdit water quality classification since the RBF – Kernel provided SVM has a recognition rate ratio (Rec rate) of 92 percent. This analysis showed that SVM has a more nonlinear ability to learn. The SVM model has only two control parameters however and the model is easy to build. For now, for small problems with samples, there is a good generalization. The application of this model and the interaction with the PSO the algorithm will reduce computational time for parameter optimization and selection of features. Finally, using the RBF – Kernel SVM model to effectively approach the nonlinearities present in the classification issue. These results explicitly agree again with the outcome criterion for 'goodness of fit' (MSE), so the RBF-Kernel SVM model is the best fit. This technique can also be used in other areas, but the particular features of each area must still be taken into account. These results correspond to the results of several works. This finding is significant because the overall costs of the monitoring system (less time of training and decreased number of sensors) have been economically affected.

V. CONCLUSION

In this paper, we have presented a performance evaluation SVM classification model for intelligent water quality monitoring. The study area is the Tilesdit dam from Algeria. An appropriate intelligent procedure based on the physicochemical parameters of surface water was proposed. It included PCA features selection, PSO parameters optimization and SVM. Particularly in this study, we utilized a cross-validation procedure that can prevent the overfitting problem by the random selection of subsamples used for testing and training datasets. The main findings from this investigation can be summarized as follows based on the numerical results: The water quality monitoring in the Tilesdit dam can first of all be correctly modeled on the hybrid PSO – RBF – SVM model. Second, a hybrid PCA-PSO – RBF – SVM model for the projection of water quality from other indicators of quality variables has been successfully developed to reduce the expense of the quality evaluation of water systems. Third, the hybrid PSO – RBF – SVM-based model for an experimental dataset was achieved with a high recognition rate. Fourth, the influence of the kernel parameters setting of the SVMs on water quality classification performance was established. Finally, it has been shown that the hybrid PCA-PSO – SVM classification method increases the generalization ability that can only be achieved with the SVM-based classification. In general, this advanced approach may well be used effectively in another field but the specific characteristics of each location must always be considered. The accuracy of the system decision can be enhanced by using new input parameters or with soft sensors that are not measured continuously in the presence of chemical parameters. The robustness of the proposed noise solution must also be carried out. It is important to note that the sensitivity of the field and unforeseen threats require a more inclusive effort to increase the system's immunity and to develop it further to reduce the risks to public health. Also, the results of this study would be useful for promoting more future studies.

ACKNOWLEDGMENT

The authors thank the editor and anonymous reviewers for numerous suggestions and their helpful and constructive comments on an earlier draft of this article that significantly improved the quality of the paper Authors wish to acknowledge Tilesdit as well as the free access to databases and for their valuable guidance in the field sampling.

REFERENCES

[1] Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Rajiv Ganthi, R., Chidambaram, S., Anandhan, P., Manivannan, R., Vasudevan, S., 2010. Application of water quality index for groundwater quality assessment: Thirumanimuttar subbasin, Tamilnadu, India. Environ. Monit. Assess. 171 (1), 595‑609.

[2] Haque, M.M., Attas, H.A.A., Hassan, M.A., 2016. Health risk assessment of trace elements in drinking water from Najran City, southwestern Saudi Arabia. Arab. J. Geosci. 9 (6), 1‑12.

[3] Li, P., Wu, J., Qian, H., 2012. Groundwater quality assessment based on rough sets attribute reduction and TOPSIS method in a semi-arid area, China. Environ. Monit. Assess. 184 (8), 4841‑4854.

[4] Tao, L., Cai, S.M., Yang, H.D., Wang, X.L., Wu, S.J., Ren, X.Y., 2009. Fuzzy comprehensive- quantifying assessment in analysis of water quality: a case study in Lake Honghu, China. Environ. Eng. Sci. 26 (2), 451‑458.

[5] Li, J., Li, X., Lv, N., et al., 2015. Quantitative assessment of groundwater pollution intensity on typical contaminated sites in China using grey relational analysis and numerical simulation. Environ. Earth Sci. 74 (5), 3955‑3968.

[6] Sirilak Areerachakul and Siripun Sanguansintukul, "Classification and regression trees and MLP neural network to classify water quality of canals in Bangkok, Thailand," International Journal of Intelligent Computing Research (IJICR), vol 1 N°(1/2), pp. 43-50, 2010.

[7] Yue Liao, Jianyu Xu, Wenjing Wang, "A method of water quality assessment based on biomonitoring and multiclass support vector machine", Procedia Environmental Sciences, Vol. 10, pp. 451-457, 2011.

[8] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the Fourth IEEE International Conference on Neural Networks, IEEE Service Center, vol.4, Perth, Australia, 1995, pp.1942–1948.

[9] R. C. Eberhart, Y. Shi, J. Kennedy, Swarm Intelligence, Morgan Kaufmann, San Francisco, 2001.

[10] D. Simon, Evolutionary Optimization Algorithms, Wiley, New York, 2013.

[11] Achmad Widodo, Bo-Suk Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors", Expert Systems with Applications, Vol. 33, pp. 241-250, 2007.

[12] Jolliffe, I. T, "Principal component analysis", Springer.New York , 2nd ed, 2002.

[13] J. L. Semmlow, "Biosignal and biomedical image processing", Marcel Dekker, New York, 2004.

[14] V. N. Vapnik, The nature of statistical learning theory, 2nd ed, Springer, New York, 2000.

[15] B. Schölkopf and A. Smola , Learning with kernels, support vector machines, regularization, optimization and beyond, MIT Press, Cambridge MA, 2002.

[16] Min Hyeok Bae, Teresa Wu, Rong Pan, " Mix-ratio sampling: Classifying multiclass imbalanced mouse brain images using support vector machine", Expert Systems with Applications, Vol. 37, pp. 4955-4965, 2010.

[17] Kunwar P. Singha, Nikita Basantb, Shikha Guptaa, "Support vector machines in water quality management", Analytica Chimica Acta, Vol. 703, pp. 152– 162, 2011.

[18] Hend S. Al-Khalifa and Amani A. Al-Ajlan, "Automatic readability measurements of the arabic text: an exploratory study", The Arabian Journal for Science and Engineering, Vol. 35 N° 2C, pp. 103-124, 2010.

[19] Chih-Hung Wu, Gwo-Hshiung Tzeng, Yeong-Jia Goo, Wen-Chang Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy", Expert Systems with Applications, Vol. 32, pp. 397-408, 2007.

[20] Ming-Huwi Horng, " Multi-class support vector machine for classification of the ultrasonic images of supraspinatus", Expert Systems with Applications, Vol. 36, pp. 8124-8133, 2009.

[21] Maysam Abedi, Gholam-HossainNorouzi,AbbasBahroudi,"Support vectormachineformulti-classificationofmineral prospectivityareas", Computers & Geosciences, Vol. 46, pp. 272–283, 2012.

[22] Héctor Ricardo Hernández De León, " Supervision et diagnostic des procédés de production d'eau Potable", Thèse de Doctorat, Institut National des Sciences Appliquées de Toulouse, 2006.

[23] Poli, R., 2008. Analysis of the publications on the applications of particle swarmopt imisation. J. Artificial Evol. Appl., 1–10.

[24] Décret exécutif N° 11-125 du 22 Mars 2011 relatif à la qualité de l'eau de consommation humaine, Journal officiel de la Republique Algerienne N° 18.

[25] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, New York, 2000.

[26] I. Steinwart, A. Christmann, Support Vector Machines, Springer, New York, 2008.