New Learning Approach for Unsupervised Neural Networks Model with Application to Agriculture Field

Belattar Sara[1]*, Abdoun Otman[2], El khatir Haimoudi[3]

Computer Science Department, Advanced Science and
Technologies Laboratory, Polydisciplinary Faculty
University of Abdelmalek Essaadi (UAE), Larache, Morocco

Abstract—An accurate and lower cost hybrid machine learning algorithm based on a combination of Kohonen-Self Organizing Map (SOM) and Gram-Schmidt (GSHM) algorithm was proposed, to enhance the crop yield prediction and to increase the agricultural production. The combination of GSHM and SOM allows to withdraw the most informative components about our data, by overcoming correlation issues between input data prior to the training process. The improved hybrid algorithm was trained firstly on data that have a correlation problem, and it was compared with another hybrid model based on SOM and Principal Component Analysis (PCA), secondly, it was trained using selected soil parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall). A comparative study with the standard SOM was conducted. The improved Kohonen-Self Organizing Map when applied to correlated data, demonstrated better results in terms of classification accuracy (8/8), and rapidity = 0.015s compared to a classification accuracy (7/8) and a rapidity = 97,828 s using SOM combined with PCA. Moreover, the proposed algorithm resulted in better results for crop prediction in terms of maximum iteration number of 675, mean error ⩽ 0.00022, and rapidity = 18.422s versus an iteration number of 729, mean error ⩽ 0.000916 and rapidity= 23.707s with the standard SOM. The proposed algorithm allowed us to overcome correlation issues, and to improve the classification, learning process, and rapidity, with the potential to apply for predicting crop yield in the agricultural field.

Keywords—*Kohonen-self organizing map; gram-schmidt algorithm; principal component analysis; agriculture field; crop yield prediction*

I. INTRODUCTION

Preserving balanced agriculture is of paramount importance as it is considered the main source of nutrition for maintaining human life. Thus, the consumption of fruits and vegetables increases on a daily basis with the increase of population [1]. Recently, climate changes hamper the accurate prediction of the crop. However, the environmental changes influence the agricultural production leading to the minimization or a decrease of crop yields which is highly dependent on climate and rainfall conditions this decrease lead also to migration problem from developing countries because the agriculture field represent the source life of the most population [2]. For that, many researchers in agriculture fields started to develop machine learning (ML) algorithms to improve agricultural production. Such as; bias-corrected random forest (BRF), multi-layer perceptron neural network (MLP), and support vector machines (SVM), were proposed to improve the estimation of agricultural drought in South-Eastern Australia [3]. Developed an estimation model of crop yield based on stepwise linear regression (SLR) and vegetation indices in order to predict crop yields [4]. Used a modified Self Organizing Map SOM based on Learning Vector Quantization (LVQ) for weather and crop prediction [5]. Another, study applied the boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in a region (Sahel) with vulnerable weather and with very little capacity to adapt [6]. Earlier research [4, 5, 7] that targeted crop prediction using machine learning tools don't account for correlated data and dependence problems, in the preprocessing step which reduces classification rate and leads to an increased cost. Pushpa Mohan et al. [5], developed a coupling of SOM and LVQ to enhance the accuracy of prediction rate, in the case of using the correlated data, a decrease of accuracy and prediction rate will be noticed. The same limitations issues were observed in the cited literature. Therefore, developing an efficient new hybrid model for solving this issue and increase the accuracy of the prediction rate. The reduction of agricultural yield is a major common problem in Africa [8], especially in Morocco. For this purpose, in this work, we decide to create an intelligent system based on a new hybrid of neural networks and the Gram-Schmidt algorithm to combater this phenomenon and also to improve the classification, prediction accuracy and rapidity of our intelligent system. Hence, this intelligent system capable of predicting the crop yield for specific soil and atmosphere based selected parameters such as (pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), which might help to overcome the ecological changes, and to improve the agricultural production. Each vegetable or fruit can be characterized by certain soil parameters to ensure good production. Various models or architectures of artificial neural networks can be found in literature this includes, Kohonen-Self-Organizing Map, Conventional neural network (CNN), and Multilayer perceptron (MLP), etc. Each model or architecture can give very good results effective in a type of task where it was applied. The model proposed in this work, is the SOM this paradigm uses the unsupervised learning algorithm and it is one of the most popular artificial neural networks models, it applied in many industries from which [9,

*Corresponding Author

10, 11, 12, 13], this model is intelligent to learn how to classify quantitative multi-parameter data using an unsupervised learning algorithm which allowed us to make a good prediction system. The goal of combining the SOM and GSHM algorithms is to improve the standard SOM. The GSHM algorithm is one of the fundamental procedures in linear algebra, the purpose of using it is to filter or to withdraw the most informative components for the sake more precise, eliminate inter-data correlation from the pattern, identify each object, and to obtain the new matrix contain novel factorizations or components. The improved SOM was trained by this new matrix, for making this training we used the correlated data from [14] and the crop prediction data collected and preprocessed from [7], the objective of using the correlated data is to test the robustness of our improved SOM in classification with the data that have a correlation problem between inputs. The empirical results showed that our new hybrid model or improved SOM proposed has a significantly better than standard SOM and hybrid model based on SOM and Principal component analysis (PCA) using the data that have a correlation problem because the PCA method can also resolve the problem of correlation and dependence between inputs, and for the crop prediction data, the outcomes showed that the improved SOM gave the best performance in terms of maximum iteration number, rapidity, and with high accuracy of mean error versus standard SOM.

This paper is organized as follows. In Section II, some techniques and methods intelligence applied in the agriculture field to solve the various issues about this domain are portrayed in the Literature review section. All methods and algorithms used in this work are presented and described in Section III. In Section IV, the experimental results of standard SOM and new hybrid or improved SOM. In Section V, we make the discussion about our work in terms of results and limitations. This followed by a section of the conclusion, we conclude the results obtained and give our perspectives for future research.

II. Literature Review

In this section, we present the most current works in the agriculture field with intelligent methods by researchers in crop and weather prediction, cited in our introduction. In this study, some brief important works found in the literature are presented.

Feng, P., Wang, B., Liu, D. L., & Yu, Q, [3] examined three advanced methods of machine learning, i.e. Bias-Corrected Random Forest, Support Vector Machine, and Multi-layer Perceptron Neural Network for remote sensing drought indices, their objective were to determine numerous remotely-sensed drought factors might be used for evaluation agricultural drought in south-eastern Australia, they supported on thirty remotely-sensed drought factors from the Tropical Rainfall Measuring Mission (TRMM) and therefore Moderate Resolution Imaging Spectroradiometer (MODIS) to breed a ground-based drought index of Standardized Precipitation Evapotranspiration Index (SPEI) at some stage in 2001-2017. The results showed that the proposed methodology of bias-corrected random forest gave a far better performance than SVM and MLP for SPEI prediction.

Pushpa Mohan, Kiran Kumari Patil, [5] have made a combination between Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ), this combination named Weighted-Self Organizing Map (W-SOM). In this work, the purpose of using the Weighted-Self Organizing Map is to predict the weather and crop for the Mysore region and to increase the prediction accuracy for rice production. In their experimental results, they found that the proposed approach of W-SOM improved the accuracy term in weather and crop prediction up to 0.5% than standard Self-Organizing Map, Ensemble Neural Network (ENN), and Kernel-Nearest Neighbors (KNN).

Zhang, L., Traore, S., Ge, J., Li, Y., Wang, S., Zhu, G. … Fipps, G, [6] employed tree regression and artificial neural networks to modeled upland rice yield under climate change within the Sahel, for artificial neural networks, they used multilayer perceptron, probabilistic neural network, generalized feedforward, and linear regression, so the probabilistic neural network followed by boosted tree regression gives the best performance of the calibrated rice yield models.

Snehal S. Dahikar, Sandeep V. Rode, Pramod Deshmukh, [7] used artificial neural networks and in particular the paradigm of feed-forward backpropagation network to combat the climate change because it has a direct effect on crop production. In this paper, in order to make an intelligent prediction system, they used soil parameters and also the parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), for predicting a decent crop.

Han, J.-C., Huang, Y., Li, Z., Zhao, C., Cheng, G., & Huang, P, [9] proposed the Self-Organizing Map in groundwater level prediction (GWL), this prediction can contribute to maintaining reliable water supply in the various field (e.g. agriculture, domestic and especially in arid and semi-arid regions). The goal of this work was to apply the SOM methodology to determine spatially homogeneous clusters of GWL piezometers in order to make a good prediction. In this literature, the proposed modeling system has the capacity to take decision-making in order to inform the use of groundwater resources control, particularly in arid regions.

III. Methods

A. Data Pre-Processing / Source

This work uses data from the literature [7] since our region from loukkos of Morocco does not have for this moment a data of crop-specific. Especially, this data are collected from Shri Shivaji Agriculture College, Amravati of Vidarbha region, India. This region is agricultural, and it is rich in a forest, among the important crop yields of this region are, cotton, oranges, and soybeans. Each crop has specific parameters of soil and weather to give a good agricultural production (e.g. ph, nitrogen, phosphate, potassium, depth, temperature, and rainfall) (see parameters in Table I). Another data contained the correlation problem from [14] is used in this work for testing the capability of our improved model in

classification. Hence, preprocessing it utilized in order to analyze the collected data.

B. Kohonen-Self Organinzing Map

The proposed methodology in this work named as standard Self-Organizing Map or Kohonen network is based on the artificial neural network. This method consisting of two layers, an input layer, and an output layer. An input layer or any pattern to be classified is represented by a multidimensional vector qualified as input vectors, and each pattern has an input neuron. An output layer is also called a competition layer where neurons compete. In SOM, the values from input neurons are passed on to all neurons in the competition layer, at the same time to approximate values passed by input neurons, and can imagine the output layer as a grid, (usually one or two-dimensional). In each node in the grid there is a "neuron" and each neuron is bound to a weight vector of the same dimension as input data vectors responsible for an area in the data space (again called input space).

In a Kohonen-Self-Organizing Map, weight vectors provide a discrete representation of the input space. They are positioned in such a way that they maintain the topological shape of the input space. By keeping neighborhood relations in the grid, they allow easy indexing (via coordinates in the grid). It can be seen from (Fig. 1) that each entry neuron has a link to all competition neurons.

The most famous learning types found in the artificial neural network are, unsupervised learning and supervised learning. SOM is a type of unsupervised learning. This type of learning is based on competitive and/or cooperative [15]. The algorithms based on unsupervised learning, are represented in particular to clustering algorithms, these latter used as an important data mining technique to group information set into a group of a similar object [16]. For that in this scenario, we propose it in order to improve it in terms of classification accuracy, learning process results, and make a good intelligent prediction system.

Before starting the learning procedures of SOM, the input data is stored as a matrix of (N-rows X M-columns) (see Fig. 2).
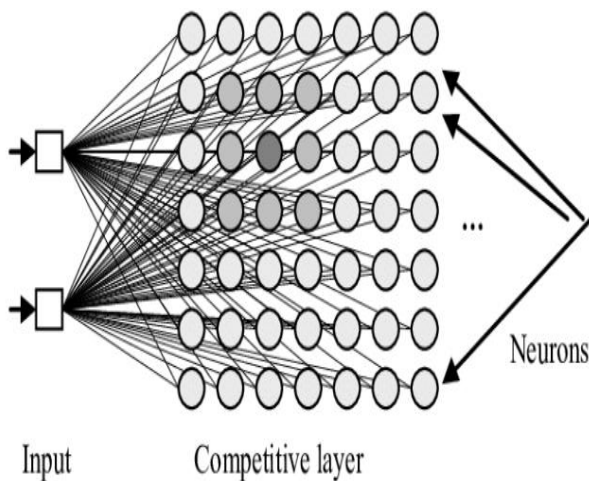

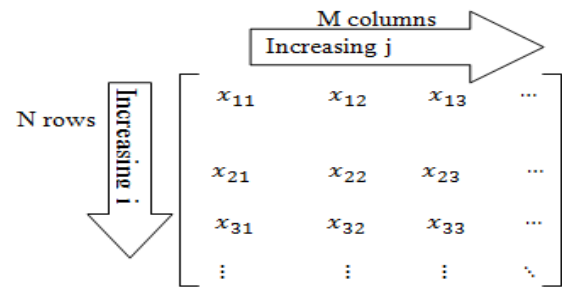Fig. 1.   Kohonen-Self Organizing Map [17].

Fig. 2.   Input Data Matrix.

Where, the rows are the objects and the columns are the components.

The inputs have to be average not too far from zero and variance not too far from 1. The input values should also not have too much weight. We can make some monotonous non-linear transformations that reduce the large values and minimize the number of training epochs of the neural networks, which leads to a better predictor and speed up learning. For that, in the first step, we add the algebraic formula of normalization [18] depicted in (1),

$$v_{ij} = x_{ij} \left/ \sqrt{\sum_{i=1}^{N} x_{ij}^2} \right. \tag{1}$$

Where, x_{ij} is the input objects and N is the number of variables inside the vector x.

Input vectors are connected with all nodes or neurons presented in the grid or map by weight vectors, the outputs of SOM methodology named the winning node. There are numerous functions for finding this winning node $j(x)$, (e.g. Euclidean distance, Product scalar, and Manhattan distance). In this work, we used the Euclidean distance function. The distance between the inputs vector and the weight vector for each node by applying the Euclidean distance is calculated in (2),

$$j(x) = \arg\min_j \; \| \, x(n) - w_j(n) \, \| \tag{2}$$

Where, j is the winner unit and $j=1, 2,....,N$, n represent the number of node, w is the synaptic weight vector of the winner unit j and x is the input vector.

Hence, for updating the weights of winning node and neighborhood region, we have to apply the formula presented in (3) [19],

$$w_j(n+1) = w_j(n) + \alpha_j(n+1)\, h_{j,i}(n+1)(x(n) - w_j(n)) \tag{3}$$

Where, index i stands for the winning node and index j stands for other nodes presented in the grid, $w_j(n+1)$ and $w_j(n)$ are the synaptic weight vectors after and before updating in n and $n+1$ iteration, the learning rate $\alpha_j = \alpha_0 \, exp(-t/\theta)$ in iteration t, α_0 is the initial learning rate and neighborhood function $h_{j,i}(n) = exp(-(j-i)^2/\sigma_i^2(n))$. This neighborhood function is centered at the wining node i.

C. Improved Kohonen-Self Organizing Map based on Gram-Schmidt Algorithm

To achieve the improvement of Kohonen-Self Organizing Map, we make the hybridization of two well-known algorithms of Gram-Schmidt and standard SOM. Thus, the development steps of the new hybrid model are described in the flowchart (Fig. 3).

Firstly, our data are passed through the GSHM processing block before it is processed with the SOM, in order to eliminate the correlation, dependence between inputs or objects, and to withdraw the most informative data. These capabilities of this new implementation will affect classification accuracy in the data that have the problem of correlation from the pattern and will affect on iteration number, mean error, and rapidity, for crop prediction data.

The Gram-Schmidt algorithm is an orthogonal space projection algorithm for finding orthogonal and orthogonal families in a Hilbert space, and it most widely used representative of a broad class of orthogonalization techniques and strategies we refer to [20-22]. In this work, we used a modified GSHM algorithm to identify each input object by A, B factorizations or components.

To illustrate the GSHM algorithm, we consider the matrix (M) which includes the learning data as follows (see Fig. 4).

Each row of the original matrix (M) (Fig. 4) corresponds to a point in an n-dimensional space (the following Fig. 5- for n = 3). In this figure, points (1, 2, and m) in space correspond to the 1st, 2nd, and mth lines of the matrix (see Fig. 5).
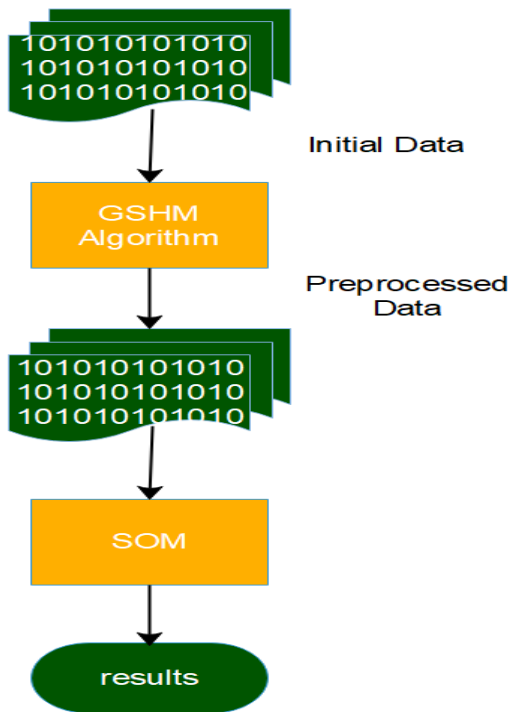

Fig. 3. Flowchart Design of our New Hybrid.

$$M = \begin{Vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \hdotsfor{3} \\ x_{k1} & x_{k2} & x_{k3} \\ \hdotsfor{3} \\ x_{m1} & x_{m2} & x_{m3} \end{Vmatrix}.$$

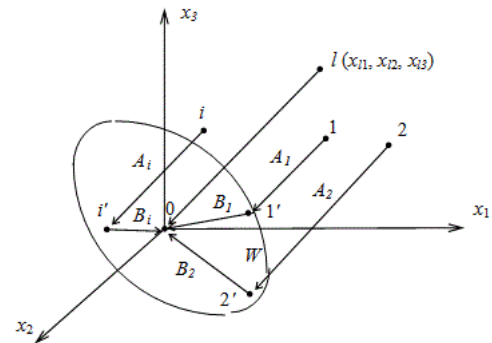Fig. 4. Organization of our Original Data in a Matrix.

Fig. 5. Representation of Matrix Lines in 3D Space.

The procedure for calculating additional components begins with the choice of the point furthest from the initial coordinates (Fig. 5), that is, the row (S) of the matrix where the square sum of its elements is maximum. We draw a vector from the selected point (1) to the initial coordinates, and perpendicular to it passing through the origin of the coordinates. We draw a plan whose equation is as follows:

$$x_{l1} * x_1 + x_{l2} * x_2 + x_{l3} * x_3 = 0 \tag{4}$$

The distances of an arbitrary point in a given space relative to the drawn plan are equal to the absolute value of (d) see the formula below:

$$d = \frac{x_{l1} * x_1 + x_{l2} * x_2 + \ldots + x_{l3} * x_3}{\sqrt{x_{l1}^2 + x_{l2}^2 + \ldots + x_{l3}}} \tag{5}$$

The values of the resulting expression for (d) correspond to the value of the component (A_l) calculated from the following function:

$$A_l^{(j)} = \sum_{p=1}^{m} (M_{l,P}^{(j)} \cdot M_{S,P}^{(j)}) / \sum_{p=1}^{m} (M_{S,P}^{(j)})^2 \tag{6}$$

As |d| it is the distances between the bases points (1) and the initial coordinates, it can be demonstrated that the component (A_l), is the relative distance between each point and the drawn plan see formula below:

$$M_{l,i}^{(2)} = M_{l,i}^{(1)} - A_l^{(1)} \cdot M_{s,i}^{(1)} \tag{7}$$

After calculating the component (A_l), we will apply the previous formula only once to calculate the lines of the matrix that make up the projected points of the original matrix, the Gram-Schmidt orthogonalization algorithm agrees with the constructed plan.

The second component (B_l) that will be calculated is the distance on the plan between the points projected on the plan

and the point of the initial coordinates. The formula for calculating the component (B_l) is depicted in (8).

$$B_l = \sqrt{\sum_{i=1}^{m} (M_{l,i}^{(2)})^2} \qquad (8)$$

Hence, in order to obtain the most informative components (A_l and B_l), we have to develop all the previous steps above. For more details, following the steps presented in Fig. 6.

So, the new matrix M' obtained by GSHM algorithm is portrayed in Fig. 7.

// The procedure steps for generating the new matrix based on the Gram-Schmidt algorithm are,

1. Read the original data (matrix *M)*;
2. Divide the number of columns / 2;
3. With the GSHM algorithm, calculate the value of the component (*A*) and calculate the distances on the plan, between the points projected on the plan and the point of the initial coordinates (*B*);
4. Add the components (*A*) and (*B*) in original data and also keep them in a new matrix M';
5. Use the finally original data contained the first components (*A*) and (*B*) for calculating other components.
6. Repeat the step 3, 4 and 5 until to get new matrix M' contained the same columns numbers of matrix *M*;
7. If the matrix *M* contained an odd number of columns, in this case, we can take only component (*A*).

Fig. 6.   The Steps for Generating the New Matrix by Gram-Schmidt Algorithm.

$$M' = \begin{Vmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{k1} & y_{k2} & y_{k3} \\ y_{m1} & y_{m2} & y_{m3} \end{Vmatrix}.$$

Fig. 7.   The New Matrix Obtained by GSHM Algorithm.

D. Principal Component Analysis

The principal component analysis (PCA) is a well-known and useful technique among the data analysis methods; it often applied in many applications [23, 24, 25]. The ultimate goal of PCA is to reduce and extract relevant information from original data, these proprieties allow resolving the correlation problem between inputs in data. For that PCA represent a robust tool especially in data mining and machine learning [26] in order to make the best analyze and taking the decision based upon the data.

IV. RESULTS

A. Experimental Results with Standard Kohonen-Self Organizing Map

1) Simulation using correlated data

In this section, we make our experimental test by using the data that have a correlation and regularities problem between the pattern and also the crop prediction data. Firstly, the correlated data are collected from [14], in this work they developed an improved SOM based on PCA for solving the problem of correlation. Their data consist of 8 inputs objects or vectors contained (similar objects, objects have regularities between input and normal objects). So, we going, to training our standard SOM from these data. The results of the classification SOM between us and the last cited paper, are depicted in Fig. 8.

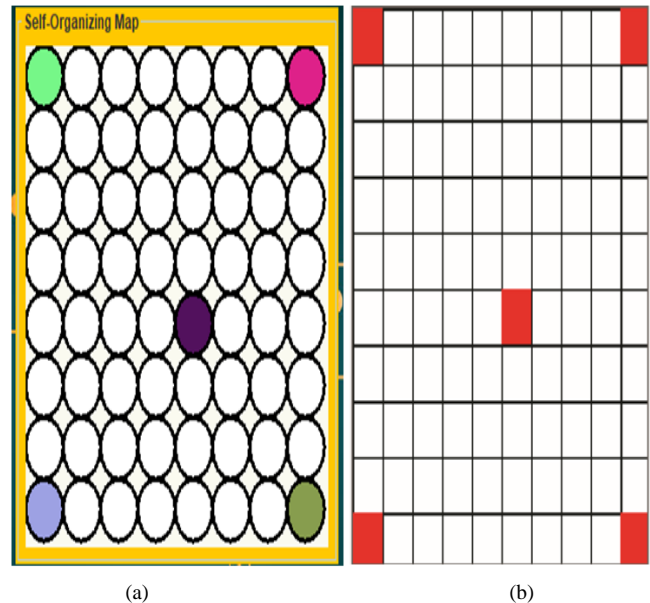(a)   (b)

Fig. 8.   (a) Our Classification Results, (b) Their Classification Results.

The figure above (Fig. 8) shows the results of the classification between our standard SOM and their standard SOM. The experimental results show that the standard Kohonen-Self Organizing Map can't perform a better classification of the data, it gave just five winning nodes in map for both of us, so our wining nodes are colored by a different color. We can conclude, that if we use data that have a problem of correlation in prediction, we can't get the best predictor system due to bad classification, this is among the goals for which we proposed our new hybrid or improved SOM for solving the classification problem in this type of data and to make the good prediction and to improve its accuracy.

2) Simulation using crop prediction data (non-correlated inputs).

Secondly, we make another test using the crop prediction data, these data consist of 9 crops (cotton, sugarcane, jowar, bajra, soybeans, corn, rice, wheat, groundnut), and each crop has specific parameters to give a good prediction (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall). The essential parameters of the crop are presented in Table I.

In order to train our standard Kohonen-Self Organizing Map, we based on the training parameters (Table I). Now, we give the dimension for creating the map of SOM (see Fig. 9).

TABLE I.   THE PARAMETERS RELATED TO THE ATMOSPHERE AND SOIL FOR CROPS (INPUT VECTORS)

Crops	Essential Parameters of Crop						
	PH	*N*	*P*	*K*	*Depth*	*Temp*	*Rainfall*
Cotton	8	100	50	50	30	33	900
Sugarcane	7	175	100	100	60	30	750
Jowar	8	80	40	40	50	30	1000
Bajra	7	40	20	25	15	28	600
Soybeans	7	30	75	15	19	27	800
Corn	8	100	25	0	40	20	500
Rice	7	100	50	50	15	22	90
Wheat	5	100	50	50	30	24	1400
Groundnut	6	25	50	30	20	24	1250

In the figure (Fig. 9) we gave (8X8) for map dimension, in next step we will insert the iteration rate (first stop condition), learning error (second stop condition), learning rate, and input data file (S_matrice) for initializing and starting the learning step of SOM (see Fig. 10).

After finishing the learning, the proposed methodology of the standard Kohonen-Self Organizing Map gave nine winning nodes in the map (each input vector has a winning node).The outputs of the standard SOM or winning nodes are presented in Fig. 11.

The learning is completed when the two conditions mean error and the maximum iteration number is realized. Then, the winning neurons or nodes are colored with different colors. Thus, the coordinates of these winning nodes in the map are depicted in Fig. 12.

We noticed that the standard SOM classified correctly our data, which allowed us to make a good prediction system in this situation. So, the learning process results of the standard SOM are presented in Table II.

In the table II, we can conclude that the standard SOM is finished it the learning by maximum iteration number = 729, mean error = 0.000916, learning number = 0.72900.
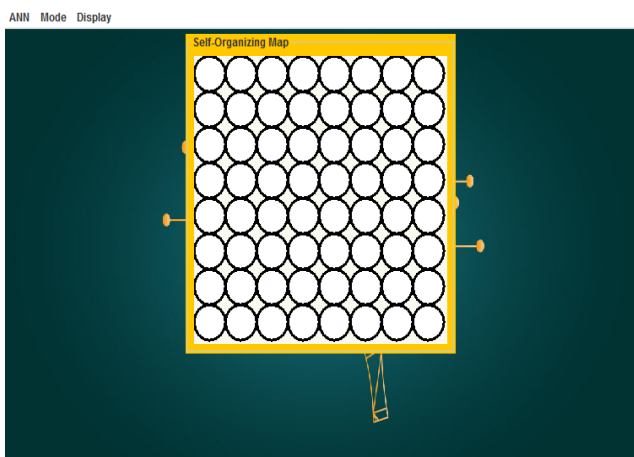

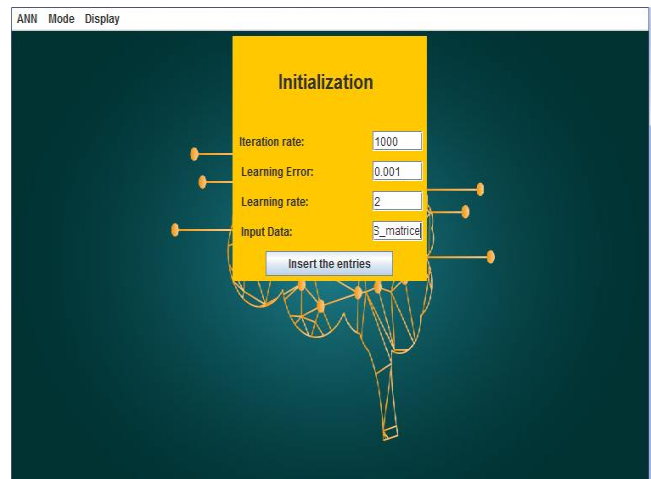Fig. 9.   Standard SOM Creation.

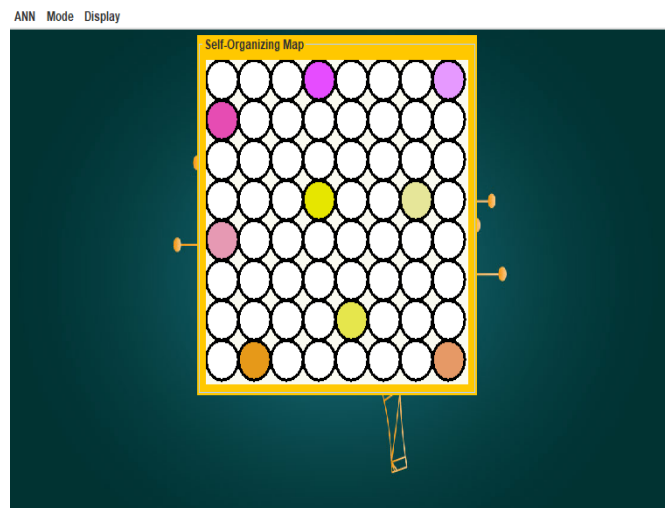Fig. 10.   Standard SOM Initialization.

Fig. 11.   Winning Nodes of Standard SOM.
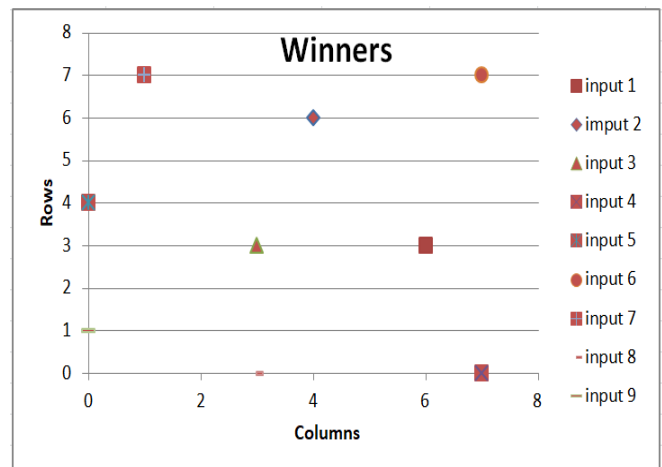

Fig. 12.   Winning Nodes Coordinates of Standard SOM.

TABLE II.   LEARNING PROCESS RESULTS OF STANDARD SOM

Iteration rate	1000
Maximum iteration number	729
Learning error	0.001
Mean error	0.000916
Learning rate	2
Learning number	0.72900
Rapidity	23.707s

Hence, our application is ready to predict the suitable crop associated with various parameters of soil and parameters related to the atmosphere (e.g. pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall), now we will insert these seven parameters for predicting the decent crop. The results of the prediction are presented in Fig. 13.
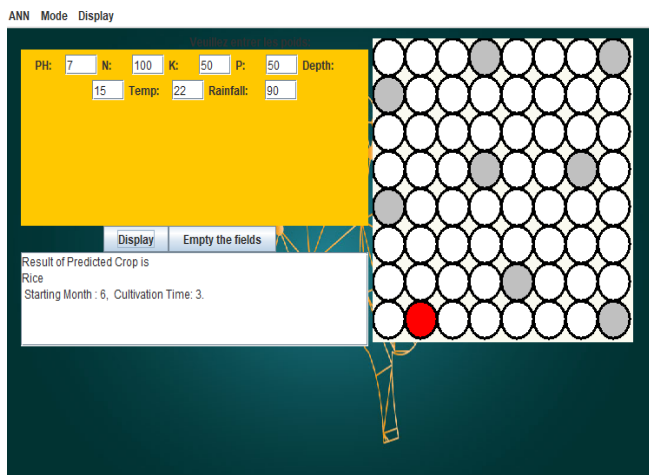

Fig. 13. The Predicted Suitable Crop is Rice.

The predicted crop based on these parameters pH=7, N=100, K=50, P=50, Depth=15, Temperature=22, Rainfall=90 is rice and the winning node of this input is colored by red color in the map. Additionally, our intelligent system gave a suitable starting month and cultivation time for this predicted crop.

In this situation, the crop data used in this test by SOM gave the best results in classification and prediction but with a big max iteration number and max error. Therefore, the objective of our improved SOM is to improve the standard SOM by minimizing iteration number, mean error, and also the rapidity.

B. Experimental Results with Improved Kohonen-Self Organizing Map

1) Simulation using correlated data obtained by PCA and GSHM

Now, we use correlated data obtained by the Gram-Schmidt algorithm and PCA for training the improved SOM based on PCA and GSHM, in this test we use the correlated date transformed by PCA from [14], in order to make the comparison in term of classification accuracy. The results of classification between SOM-PCA and GSHM-PCA as shown in Fig. 14.
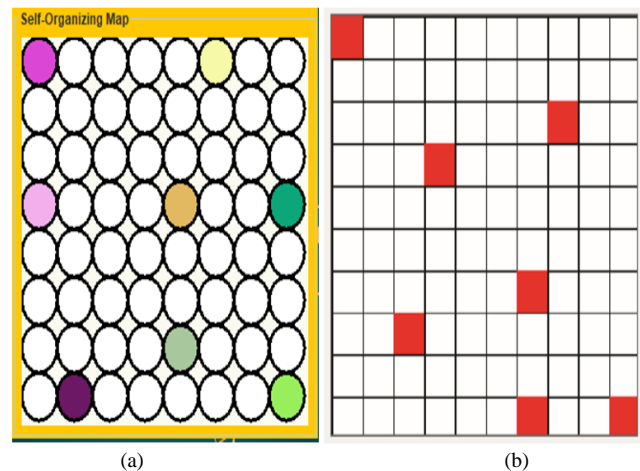

Fig. 14. (a) Improved SOM by Gram-Schmidt algorithm (classification results), (b) Improved SOM by Principal Component Analysis (classification Results).

In Fig. 14, our new hybrid of GSHM-SOM and their proposed model of SOM-PCA gave the best performance in classification accuracy, they can classify correctly the data and resolve the correlation problem by eliminating the dependence between inputs, but our new improved SOM based on the GSHM algorithm gave a high accuracy of classification (8/8) than SOM-PCA (7/8). Moreover, the rapidity of the Gram-Schmidt algorithm is 0.015s compared to Principal component analysis is 97,828 s, because the PCA is an iterative algorithm, for that it takes more time for giving the results against the GSHM is a linear algorithm that allows us to obtain the results in fast time.

2) Simulation using crop prediction data obtained by GSHM

The Gram-Schmidt algorithm is powerful than principal component analysis in classification and rapidity. For that, we proposed it in this work. Now, we will test the crop prediction data obtained by the GSHM algorithm for analyzing the experimental results. Although these data don't contain the correlation issues between inputs, our improved SOM will have an influence on the maximum iteration number, mean error, and rapidity. The data obtained by the GSHM algorithm are presented in Table III.

TABLE III.   INPUT DATA OBTAINED BY GSHM ALGORITHM

Input Vector Nº	Input Vectors Obtained by Gram-Schmidt Algorithm
1	[0, 907.51, 0, 914.97, 0, 922.37, 0]
2	[0.848, 107.85, 0.144, 188.58, 0.230, 178.65, 0.217]
3	[2.410, 6.869, 0.162, 18.222, 0.173, 18.799, 0.183]
4	[2.044, 9.666, 0.692, 7.801, 0.624, 7.830, 0.627]
5	[4.042, 3.558 E-12, 1.196, 1.663, 1.193, 1.860, 1.196]
6	[3.192, 57.992, 5.300, 5.641E-12, 4.520, 8.266, 4.615]
7	[0.841, 34.294, 6.944, 2.480, 6.582, 2.462E-12, 6.547]
8	[17.438, 52.244, 8.380, 26.950, 7.966, 13.357, 7.822]
9	[16.055, 30.509, 3.756, 44.583, 3.808, 28.923, 3.621]

After GSHM we got the most informative components about original data of 9 crops (Table III). Thus, we going to training our improved Kohonen-Self Organizing Map, through the new data that we got by the Gram-Schmidt algorithm. Hence, for starting the learning procedure of SOM, we used the same conditions of the first test (i.e. iteration rate, learning error, learning rate, and map dimension number) and input file data (M_matrice) (see Fig. 15).

As a result, the improved SOM gave also nine winning nodes in the map (each input vector has a winning node), so the wining nodes of improved SOM are depicted in Fig. 16.

The winning nodes of improved SOM are colored by different colors. The coordinates of these winning nodes in the map are shown in Fig. 17.

As a conclusion, we noticed that the improved SOM classified correctly our data transformed by the Gram-Schmidt algorithm and gave a good classification. The learning process results of the improved SOM are displayed in Table IV.

Table IV shows that the new hybrid of improved SOM is stopped by 675 for maximum iteration number and by 0.00022 for means error versus standard SOM it stopped by 729 for maximum iteration number and by 0.000916 for mean error. For that, we can conclude that the coupling of the Gram-Schmidt algorithm and standard SOM has been applied successfully, in terms of maximum iteration number, mean error, rapidity, and precision of classification. These capabilities allow us to make a good prediction system in a fast time.

Our improved SOM is ready to predict the suitable crop by using various soil parameters and parameters related to the atmosphere obtained by the GSHM algorithm, now we will insert these parameters for predicting the decent or suitable crop. The operation for predicting the crop yield is shown in Fig. 18.

The predicted crop based on these parameters pH=0.841, N= 34.294, K= 6.944, P= 2.840, Depth= 6.582, Temperature = $2.462 \cdot 10^{-12}$, Rainfall= 6.547 is rice and the winning node of this input object is colored by red color. In addition, our intelligent system gave a suitable starting month and cultivation time for this predicted crop (Fig. 18).
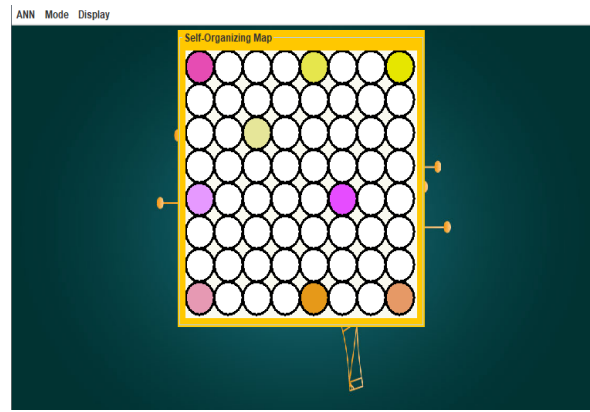

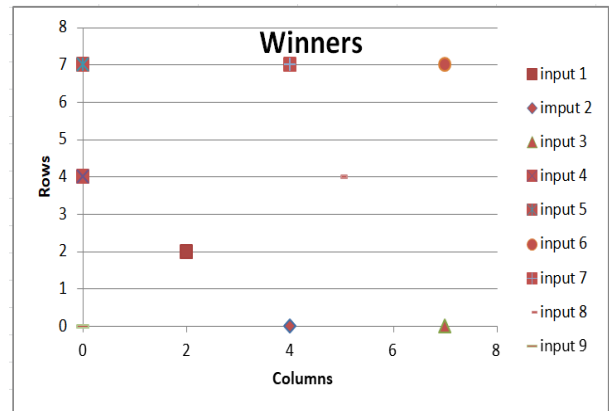Fig. 16. Winning Nodes of Improved SOM.

Fig. 17. Winning Nodes Coordinates of Improved SOM.

TABLE IV. LEARNING PROCESS RESULTS OF IMPROVED SOM

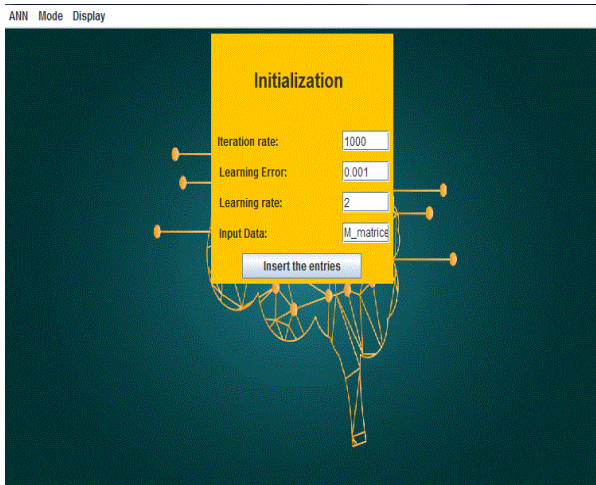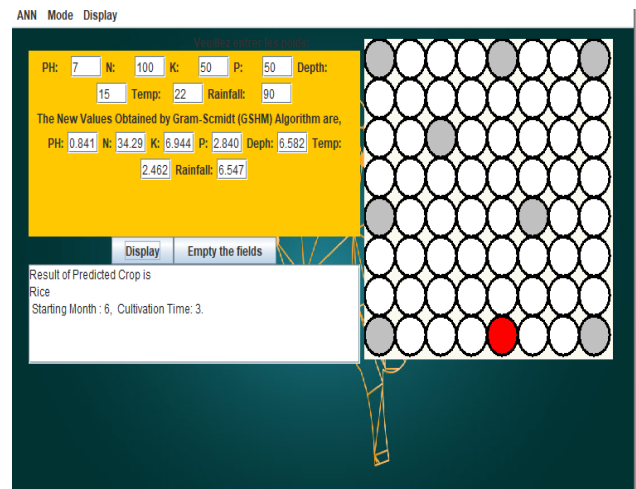|	
Iteration rate	1000
Maximum iteration number	675
Learning error	0.001
Mean error	0.00022
Learning rate	2
Learning number	0.78567
Rapidity	18.422s

Fig. 15. Improved SOM Initialization.

Fig. 18. The Predicted Suitable Crop is Rice.

C. A Comparison between Standard SOM and Improved SOM

In this section, the maximum iteration number and mean error of standard SOM are compared with the maximum iteration number and mean error of improved SOM.

The comparison between standard SOM and improved based on crop prediction data in terms of maximum iteration number and mean error for every two columns from the original matrix M and new matrix M', is displayed in Fig. 19 and Fig. 20.

Fig. 19 determines the maximum iteration numbers graph comparison through our original matrix M of standard SOM and new Matrix M' obtained by the Gram-Schmidt algorithm of the improved SOM in all different two columns. The result shows, that we got an increasing and decreasing in iteration number between columns 2, 4, 6, 7 for original matrix M and new matrix M', but generally, the new matrix M' of 7 columns gave the best maximum iteration number by 675 versus 729 of maximum iteration number in original matrix of 7 columns. In conclusion, we can say that our proposed new hybrid is not only based on iteration numbers for columns but is based to solve the correlation problem in the multitude of learning like the linear dependence pattern components. Knowing that our new hybrid, allowed us to develop a linear algorithm in contrary to the algorithm published in [27].
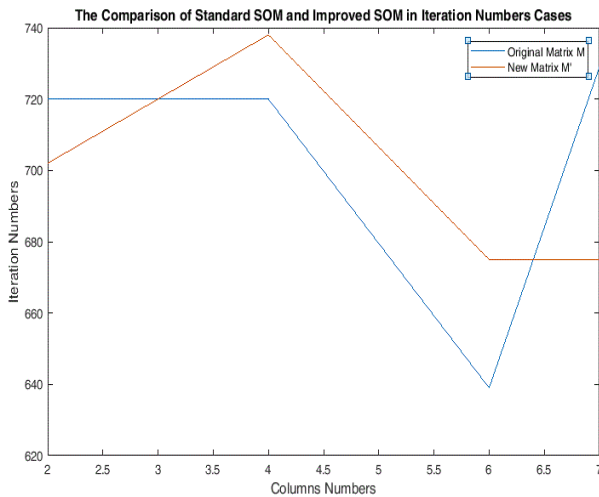

Fig. 19. Comparison in Terms of Iteration Numbers.

Fig. 20 specifies that the improved SOM has better in terms of mean error compared to standard SOM because we have significantly a decreasing of mean error in all two columns for new matrix M' of improved SOM than original matrix M of standard SOM, this means demonstrated that the proposed new hybrid has high accuracy.
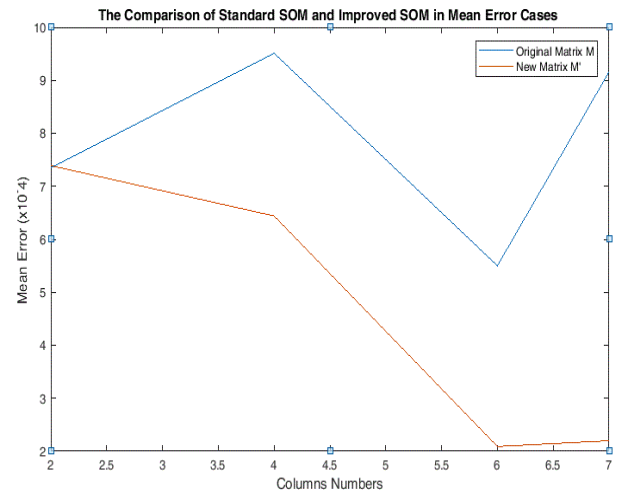

Fig. 20. Comparison in Terms of Mean Error.

V. DISCUSSION

Artificial neural networks offer very powerful and effective methods, capable of performing complicated tasks, besides, it has the properties not found in the classical methods (e.g. learning ability, adaptation, classification, and generalization of results). Thus, these capabilities allow us to solve various issues and tasks requiring intelligence. For that we proposed the new approach of ANN to improve it, in order to be more powerful to apply it in numerous problems facing several fields, especially in the prediction system.

This work proposed a new hybrid model of Kohonen-Self Organizing Map and Gram-Schmidt algorithm in order to improve it at the level of classification accuracy, maximum iteration number, mean error, and rapidity, this study showed that the improved model gave the high accuracy of classification when we used the correlated data and also it minimizes the number of iteration, mean error with high accuracy, and speed up our intelligent system when we use the crop prediction data compared to standard SOM. These powerful of improved SOM can solve the most of problems that we may encounter in the prediction field, which allow us to make a good predictor system for predicting the suitable crop by using various parameters of soil and parameters related to the atmosphere. We can say that the GSHM algorithm represents a powerful tool for machine learning, especially for SOM.

The results of this study couldn't be directly compared to previous literature, as we use different metrics to evaluate the quality of the developed model and very limited studies focused on correlations issues. However, some of the results such as classification and rapidity could be indirectly compared to a previous study by our group [14]. The author targeted correlation problems in the input data using SOM combined with PCA. The proposed model presented here demonstrated better results in terms of classification and computational cost.

However, they are some limitations that need to be acknowledged, this includes the problem of collecting real data in morocco related to crop prediction. In addition, it is difficult to find big data with correlated inputs.

VI. CONCLUSION

This paper illustrated a prediction of the crop by various parameters of soil for the Vidarbha region, this prediction system is realized by our new proposed methodology of Kohonen-Self Organizing Map. In this study, the improved SOM based on the Gram-Schmidt algorithm was implemented to avoid problems related to the correlation issues and speed up the learning process results with high accuracy. The improved SOM gave the best performance than standard SOM in terms of classification, maximum iteration number, mean error, rapidity and it was efficient in prediction. In future work, for further improving the results obtained by improved SOM and also we will use the data of our region loukkos, Morocco; this region is rich in fruits and vegetables (e.g red fruits, potatoes, tomatoes, etc.). And to solve several issues that face farmers.

REFERENCES

[1] Tanha Talaviya, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, Manan Shah, Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides, Artificial Intelligence in Agriculture, 2020, ISSN 2589-7217, https://doi.org/10.1016/j.aiia.2020.04.002.

[2] Falco, C., Galeotti, M., & Olper, A. (2019). Climate change and migration: Is agriculture the main channel? Global Environmental Change, 59, 101995. doi:10.1016/j.gloenvcha.2019.101995.

[3] Feng, P., Wang, B., Liu, D. L., & Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, 303–316.

[4] O. Satir and S. Berberoglu, "Crop yield prediction under soil salinity using satellite derived vegetation indices", Field Crops Research, Vol.192 , pp.134-143, 2016.

[5] Pushpa Mohan, Kiran Kumari Patil, Weather and Crop Prediction Using Modified Self Organizing Map for Mysore Region, REVA University, Computer Engineering, India, December 23-2017, International Journal of Engineering and Intelligent Systems, Vol.11, No .2, 2018.

[6] Zhang, L., Traore, S., Ge, J., Li, Y., Wang, S., Zhu, G., … Fipps, G. (2019). Using boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in Sahel. Computers and Electronics in Agriculture, 166, 105031.

[7] Snehal S. Dahikar, Sandeep V. Rode, Pramod Deshmukh An Artificial Neural Network Approach for Agricultural Crop Yield Prediction Based on Various Parameters, International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 1, January 2015.

[8] Sultan, B., Defrance, D. & Iizumi, T. Evidence of crop production losses in West Africa due to historical global warming in two crop models. Sci Rep 9, 12834 (2019). https://doi.org/10.1038/s41598-019-49167-0.

[9] Han, J.-C., Huang, Y., Li, Z., Zhao, C., Cheng, G., & Huang, P. (2016). Groundwater level prediction using a SOM-aided stepwise cluster inference model. Journal of Environmental Management, 182, 308–321. doi:10.1016/j.jenvman.2016.07.069.

[10] Li, Y., Wright, A., Liu, H., Wang, J., Wang, G., Wu, Y., & Dai, L. (2019). Land use pattern, irrigation, and fertilization effects of rice-wheat rotation on water quality of ponds by using self-organizing map in agricultural watersheds. Agriculture, Ecosystems & Environment, 272, 155–164.

[11] Wang Pan, Yimin Sun, Michela Turrin, Christian Louter, Sevil Sariyildiz, Design exploration of quantitative performance and geometry typology for indoor arena based on self-organizing map and multi-layered perceptron neural network, Automation in Construction, Volume 144, 2020, 103163, ISSN 0926-5805, https://doi.org/10.1016/j.autcon 2020.103163.

[12] Mehmet Ozcalici, Mete Bumin,An integrated multi-criteria decision making model with Self-Organizing Maps for the assessment of the performance of publicly traded banks in Borsa Istanbul,Applied Soft Computing, Volume 90,2020,106166, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106166.

[13] Shin Nagai, Hiroshi Morimoto, Taku M. Saitoh,A simpler way to predict flowering and full bloom dates of cherry blossoms by self-organizing maps,Ecological Informatics,Volume 56,2020,101040,ISSN 1574-9541,https://doi.org/10.1016/j.ecoinf.2019.101040.

[14] El Khatir HAIMOUDI, Loubna CHERRAT, Hanane FAKHOURI & Mostafa Ezziyyani. (2016). Towards a New Approach to Improve the Classification Accuracy of the Kohonen's SelfOrganizing Map During Learning Process. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3.

[15] Lee, M., Song, T.-G., & Lee, J.-H. (2020). Heartbeat classification using local transform pattern feature and hybrid neural fuzzy-logic system based on self-organizing map. Biomedical Signal Processing and Control, 57, 101690. doi:10.1016/j.bspc.2019.101690.

[16] Wang, B., Kong, Y., Zhang, Y., Liu, D., & Ning, L. (2019). Integrationof Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment. Expert Systems with Applications.

[17] https://www.researchgate.net/figure/Kohonen-network-architecture_fig1 _224355225

[18] Bazrafkan, A., & Pakravan, M. R. (2017). An MADM network selection approach for next generation heterogeneous networks. 2017 Iranian Conference on Electrical Engineering (ICEE).doi:10/1109/iraniancee. 2017.7985361.

[19] S. Hosseini, "Binary Tree Time Adaptive Self-Organizing Map", Neurocomputing, Vol.74, No.11, pp.1823–1839, May 2011.

[20] A. BjSrck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, PA, (1996)

[21] G.H. Golub and C.F. VanLoan, Matrix Computations, Third Edition, John Hopkins University Press, Baltimore, MD, (1996).

[22] N. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, Philadelphia, PA, (1996).

[23] Christine Borgen Linander, Rune Haubo Bojesen Christensen, Graham Cleaver, Per Bruun Brockhoff, Principal component analysis of d-prime values from sensory discrimination tests using binary paired comparisons, Food Quality and Preference,Volume 81, 2020, 103864, ISSN 0950-3293,https://doi.org/10.1016/j.foodqual.2019.103864

[24] Seyed Ali Ghorashi, Farhang Honarvar, Morteza Tabatabaeipour,Automated extraction of local defect resonance using the principal component analysis in lock-in ultrasonic vibrothermography, Infrared Physics & Technology,Volume 105,2020, 103204,ISSN 1350-4495,https://doi.org/10.1016/j.infrared.2020.103204

[25] Yu Zhang, Xiaobo Tian, Rongguang Liang,Phase extraction from two randomly phase shifted interferograms by combining advanced principal component analysis and Lissajous ellipse fitting, Optics and Lasers in Engineering,Volume 132,2020,106134,ISSN 0143- 8166,https://doi.org//10.1016/j.optlaseng.2020.106134.

[26] Qianqian Wang, QuanXue Gao, Gan Sun, Chris Ding,Double robust principal component analysis,Neurocomputing,Volume 391,2020,Pages 119-128,ISSN 0925-2312,https://doi.org/10.1016/j.neucom.2020.01.097

[27] Haimoudi El Khatir, Tsymbal Yu.V. Method of preprocessing of realization set for counter propagation neural networks, vidbir-zm-2003_19e, p 154-158.