AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE
ISSN: 1995-0748, EISSN: 1998-1074
2016, volume(10), issue(5): pages (1-9)
Published Online 23 October 2016 in http://www.aensiweb.com/AEJSA/

Toxicological Clas`Ifier By K-Means Clustering For Pesticides Organochlorine Organophosphates And Carbamate

[1]Angélica Palacios, [2]Olga Ramos, [3]Darío Amaya

[1,2,3]Universidad Militar Nueva Granada

Received 12 July 2016; Accepted 20 October 2016

Address For Correspondence:
Angélica Palacios, Universidad Militar Nueva Granada

ABSTRACT

Summary- The use of pesticides has been implemented in recent times as efficient solution for pest eradication and control of crops. These chemicals become lethal to different organisms, depending on the dose, type and duration of exposure. Its lethality lies mainly in its toxicity, which differs depending on the type of pesticide by its structure, group and chemical properties. Analysis of toxicity of chemical compounds, is a complex process that requires time and cost mainly. For this, we turn to computer algorithms these can reduce time and optimize the process.In this paper, the result of the development of an algorithm toxicological classification by K-means clustering is presented, from the entity relationship variables associated with pesticides as Lethal Dose 50, Effective Concentration 50, Bio concentration Factor and water solubility, these parameters were selected according to a first regression to identify their relationship with the Lethal dose. Resulting in the classification of a 38.46% of organochlorine pesticides, 89.47% of organophosphates and 75% of carbamates in the group of lower toxicity, while 46.15% of organochlorine pesticides and 7.89% of organophosphate pesticides were classified in the group of higher toxicity.

KEY WORDS

K-Means clustering, Pesticides, lethal dose, effective concentration, BCF.

INTRODUCTION

Chemicals such as pesticides are first choice resources for agriculture, because they represent an undeniable benefit to this activity, guaranteeing better and more profitable production and increased the quantity and quality of products [1]. Despite being a benefit can large proportion affect the human population by exposure to domestic use, proximity to agricultural activities where it is handled, and consumption through water and food waste [2,21,22,23,24,25].

Pesticides can become extremely dangerous because of their intrinsic toxicity [3], [4]. The toxicity associated with these chemicals is specifically related to their chemical structure, body metabolism, degree of decomposition, route of administration, concentration, among others [5,21,22,23,24,25].

Toxicity can be described as specific toxic activity and a substance due to its chemical structure, and is related to the ability to produce damage to an organism. Pesticides are classified according to their toxicity through concentration, either the Lethal Dose or Lethal Concentration [6], [7], [8].

For the classification of pesticides is always successful grouping by its chemical properties, this will establish correlation between structure and activity, structure and toxicity, and its structure and various mechanisms of degradation, due to limitations in toxicity studies has been the need to develop new methods to reduce time and costs of tests [9,21,22,23,24,25].].

With the analysis and measurement of the toxicity of chemicals using methodologies such as (QSAR), seeks to obtain a mathematical relationship between quantized biological activity of the substance and the physicochemical properties of this substance [10], [11].

For analysis and knowledge extraction database is used association techniques, clustering, and classification, among others, depending on the problem at hand. Statistical, evolutionary, search and probabilistic neural networks grounded in rules, algorithms are mostly implemented in different scientific areas because of its highly effective [12], [13].

In works like [14], the study of quantitative structure-activity relationship of pesticides was performed by classification techniques, the objective of this work was to compare two techniques classification, one of them the classification methods Nearest Neighbor Algorithm(k-NN) and the other consisted of artificial neural networks against propagation (CP-ANN), applied to a model of toxicity of different pesticides. In [15] a study on the assessment of non-conventional insecticides was developed through analysis of lethal effects associated with lethal and sub-lethal concentration. The study of the toxic effect of polluting pesticides was carried out in [16].

Evaluation of induced toxicity pesticides present in water was the subject of research [17], where the results obtained allowed the characterization of damage to biomolecules and proteins in organisms present in the water study.

The development of a toxicological classification algorithm using the K-means clustering technique is presented, for classifying 83 pesticides including organochlorines, organophosphates and Carbamate. The variables related to the biology of these pesticides activity are the main physicochemical properties specific for each pesticide.

Methods:

Initially it uses databases to obtain information related to the main properties and characteristics of organochlorine pesticides, organophosphates and Carbamate.

Selected for each pesticide properties were, (lethal dose 50, maximum inhibitory concentration I50, effective concentration 50, Log P, mass density, BCF, water solubility and melting point).

Some data obtained are presented in

Table ***1***.

Table 1: Preliminary database.

Type	*Name*	*DL50*	*EC50*	*Log p*	*Mass D.*
O-chlorine	*Aldrín*	39	0,028	6,5	1,6
O-chlorine	*Endrín*	7,5	0,0042	3,2	1,84
O-chlorine	*TDE*	113	0,009	6,02	1,385
O-chlorine	*Clordano*	460	0,59	2,78	1,61
O-chlorine	*DDT*	113	0,005	6,91	0,99
O-chlorine	*Endosulfan*	38	0,44	4,75	1,8
O-chlorine	*Lindano*	163	1,6	3,5	1,88
O-chlorine	*Isobenzan*	4,8	0,008	4,51	1,9
O-chlorine	*Metoxicloro*	6000	0,00078	5,83	1,41
O-chlorine	*Heptacloro*	147	0,042	5,44	1,58
O-phosphate	*Butonato*	1100	0,006	1,71	-
O-phosphate	*EPN*	14	0,00006	5,02	1,27
O-phosphate	*Etion*	208	0,000056	5,07	1,22
O-phosphate	*Foxim*	2000	0,0008	3,38	1,18
O-phosphate	*Paration*	2	0,0025	3,82	1,26
O-phosphate	*Bensulida*	270	0,58	4,2	1,25
O-phosphate	*Bromofos*	1600	0,0086	5,24	-
O-phosphate	*Clorpirifos*	64	0,0001	4,7	1,51
O-phosphate	*Diazinon*	1139	0,001	3,69	1,11
O-phosphate	*Forato*	2	0,004	3,86	1,17
Carbamate	*Aldicarb*	0,93	0,42	1,15	1,2
Carbamate	*Ferbam*	4000	0,009	-1,6	0,21
Carbamate	*Carbaryl*	614	0,0064	2,36	1,21
Carbamate	*Methiocarb*	19	0,008	3,18	1,25
Carbamate	*XMC*	542	0,055	2,23	1,16
Carbamate	*Barban*	527	0,3	3,41	1,39
Carbamate	*Carbosulfan*	101	0,0015	7,42	1,04
Carbamate	*Tiram*	1800	0,011	1,73	1,36
Carbamate	*Ziram*	267	0,048	1,65	1,71
Carbamate	*Tiofanoz*	8,5	0,31	2,16	1,05

To select properties that have a direct relationship with the lethal dose, a linear regression was performed to obtain the results of

Table **2**.

Table 2: Regression results.

	DL50
DL50	1
IC50	*-0.2127*
EC50	0.35989
Log P	*-0.0777*
Mass D	0.15691
BCF	0.12918
Water Sol	0.12919
Melting P	0.34064

According to the results obtained in the regression, the IC50 and Log P properties were discarded because presented a negative regression index, although properties like mass density and melting point showed a positive rate of regression were discarded due their not consistency in most pesticides which could affect the convergence of the algorithm.

Finally the database for the classification algorithm is shown in Table 3.

Table 3: Database for the classification algorithm.

Name	*DL50*	*EC50*	*BCF*	*Water S.*
Aldrín	39	0,028	3348	0,027
Alpha-hexa	177	0,37	0	2
Clorbenzilato	2784	0,01	750	10
Clordano	460	0,59	8460	0,1
DDT	113	0,005	3173	0,006
Endosulfan	38	0,44	2755	0,32
Endrín	7,5	0,0042	3970	0,24
Heptacloro	147	0,042	2430	0,056
Isobenzan	4,8	0,008	0	0,095
Lindano	163	1,6	1300	8,52
Metoxicloro	6000	0,00078	1622	0,1
Pentaclorofenato	80	0,45	216	1000
TDE	113	0,009	0	0,09
Azametifos	1180	0,00067	1,56	1100
Azinfos-etil	12	0,0002	101	4,5
Azinfos-metil	9	0,0011	40	28
Bensulida	270	0,58	880	25
Bromofos	1600	0,0086	4,47	40
Bromofos-etil	52	0,0086	10000	2
Butonato	1100	0,006	0	292
Cadusafos	30,1	0,00075	220	245
Cianofos	610	0,097	417	46
Clorfenvinfos	12	0,00025	250	145
Clorpirifos	64	0,0001	1374	1,05
Clorpirifos-metil	2814	0,0006	1800	2,74
Diazinon	1139	0,001	500	60
Diclofention	172	0,0011	3000	0,245
Dimethylvinphos	97,5	0,002	0	130
Dioxation	23	0,00035	0	1,55
Edifenfos	150	0,000032	0	56
EPN	14	0,00006	2346	0,5
Etion	208	0,000056	586	2
Etoprop	32,9	0,2	225	1300
Etrimfos	1800	0,004	336	40
Fenamifos	6	0,0019	110	345
Fenitrotion	330	0,0086	29	19
Fenkaptona	44	0,0013	0	0,044
Fonofos	6,8	0,0023	300	13
Forato	2	0,004	483	50
Fosalon	120	0,00074	180	1,4
Fosmet	113	0,002	79	15,2
Fostiazato	57	0,282	3,2	9000
Foxim	2000	0,00081	1610	1,5
Heptenofos	96	0,0022	0	2200
Iodofenfos	2330	0,0016	0	0,1
Iprobenfos	680	0,2	0	540

Isazofos	40	0,5	75	69
Isofenfos	28	0,0039	240	24
Malation	1778	0,0007	103	148
Metidation	25	0,0064	12,6	240
Paration	2	0,0025	40	24
Alanycarb	330	9,4	164	20
Aldicarb	0,93	0,42	42	4930
Aminocarb	30	0,19	6	915
Barban	527	0,3	2,2	11
Bendiocarb	34	0,03	64,8	280
Benfuracarb	205	0,01	90	8,4
Carbaryl	614	0,0064	44	9,1
Carbofurano	7	0,0094	12	332
Carbosulfan	101	0,0015	990	0,11
Chinometionato	5000	0,12	3000	1
Clororofam	4200	2,6	144	110
Desmedifam	5000	0,45	157	7
Etiofencarb	200	0,22	75	1900
Fenmedifam	8000	0,41	165	1,8
Fenobucarb	620	0,1	0	420
Fenotiocarb	1200	6,7	0	30
Ferbam	4000	0,009	0,95	130
Furatiocarb	53	0,0018	92	11
Isoprocarb	403	0,024	0	270
Methiocarb	19	0,008	75	27
Mexacarbato	14	0,0018	0	100
Pebulato	1120	5,9	359	100
Pirimicarb	142	0,017	24	3100
Propoxur	50	0,15	0	1800
Tiobencarb|	560	1,1	302	16,7
Tiodicarb	50	0,027	6,3	22,2
Tiofanox	8,5	0,31	14	5200
Tiram	1800	0,011	3,9	16,5
Trialato	1100	0,091	1400	4,1
Vernolato	1500	1,8	366	90
XMC	542	0,055	0	470
Ziram	267	0,048	470	0,967

Given the level of toxicity algorithm presume to classify the above pesticides into the following groups, Table 4.

Table 4: Group's classification.

	Group 1	Group 2	Group3
Toxicity	Low	Medium	High

K-Means Clustering:

The algorithm of K-Means classification classifies a group of data elements in a predefined number of groups. Starts with random initial centroids and performs a re-allocation of data in different groups based on the similarity between the data object and the group centroids until a termination criterion is met, either a fixed number of iterations, or the algorithm reaches stability in the circulation of data points between groups [18].

In [19], the K-Means algorithm is summarized in the following steps:

1. Consider the first *k* items in the data set as *k* clusters of a single item.

2. Assign in the order of the data set each of the items to the nearest centroid. After each assignment the new centroid recalculation is done.

3. After all items have been allocated in the previous step, calculating the centroids of the clusters obtained and reassign each item to the nearest centroid.

4. Repeat steps 2 and 3 until a stopping criterion is reached.

In the first step, the idea is to consider the first *k* items in the data set as *k* initial centroids, then assigned each of the items to the nearest centroid, with the characteristic that when performing each assignment coordinates are recalculated new centroid.

For this work the algorithm of K-Means classification was developed as shown below:

A. Standardization of data.

B. Assign each data set a random group.

$$G[i \times 1] = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ g_n \\ 1 \\ 2 \\ \vdots \end{bmatrix}$$

$where\ n = Number\ of\ \ groups;$
$i = Number\ of\ pesticides$

Each data set represents a pesticide and is formed by the values of each of the properties associated with the pesticide.

$$P_i = [X1, X2, X3, \ldots XN]$$

Group Assignment

$$\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_i \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_i \end{bmatrix}$$

C. Perform counting the number of data sets grouped in each of the groups.

D. Calculation of the centroid.

The calculation is performed for n groups regarding the properties of the dataset.

As illustrated in (1), the centroid of a group with respect to an input is the sum of the values of that input that are grouped, divided by the number of sets corresponding to that group.

$$centroid[g, X_n] = \frac{\sum Data[i, X_n]}{N\ Grouping\ g} \quad (1)$$

E. Calculation of Euclidean distance between the point and the centroid (2).

$$Dist = \sqrt[2]{\sum (Data[i, X_n] - centroid[g, X_n])^2} \quad (2)$$

F. Select the minimum distance between the point and the different groups.

G. Assign new groups according to their minimum distance.

H. Calculation of error.

The error is calculated with the theoretical values of the classification in [14].

I. The algorithm ends when the square error is very small or reach the maximum of iterations.

K-Means is the most efficient algorithm in terms of runtime, but has the disadvantage that the results are extremely sensitive to the selection of initial centroid and can converge a local optimal solution [20].

InFig. ***1***, the flow chart of the classification algorithm is presented.

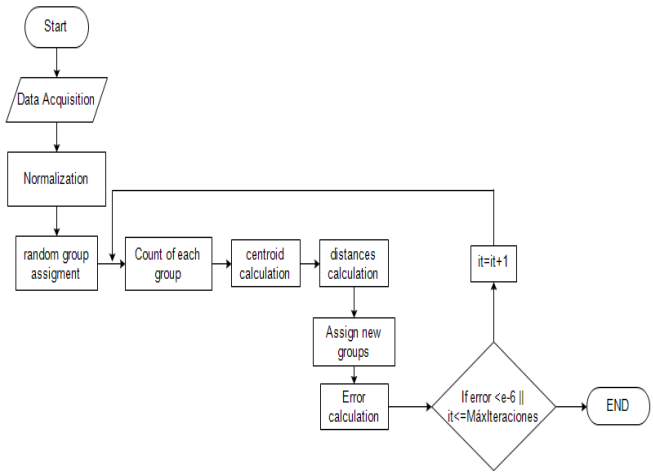

Fig. 1: Flow Chart.

Analysis of Results:

The results obtained in the classification are presented below.
In

Table *5*, the classification obtained by the algorithm after 80 iterations are shown.

Table 5: Classification results.

Name	*Classification*
Aldrín	3
Alpha-hexa	1
Clorbenzilato	2
Clordano	3
DDT	3
Endosulfan	3
Endrín	3
Heptacloro	3
Isobenzan	1
Lindano	1
Metoxicloro	2
Pentaclorofenato	1
TDE	1
Azametifos	1
Azinfos-etil	1
Azinfos-metil	1
Bensulida	1
Bromofos	1
Bromofos-etil	3
Butonato	1
Cadusafos	1
Cianofos	1
Clorfenvinfos	1
Clorpirifos	1
Clorpirifos-metil	2
Diazinon	1
Diclofention	3
Dimethylvinphos	1
Dioxation	1
Edifenfos	1
EPN	3
Etion	1
Etoprop	1
Etrimfos	1
Fenamifos	1
Fenitrotion	1
Fenkaptona	1
Fonofos	1
Fosalon	1
Fosmet	1
Alanycarb	2

Aldicarb	1
Aminocarb	1
Barban	1
Bendiocarb	1
Benfuracarb	1
Carbofurano	1
Carbosulfan	1
Chinometionato	2
Clororofam	2
Desmedifam	2
Fenmedifam	2
Fenobucarb	1
Fenotiocarb	2
Ferbam	2
Furatiocarb	1
Isoprocarb	1
Methiocarb	1
Mexacarbato	1
Pebulato	2

The algorithm presented an error of 8.77e-4, given that the selection of initial groups is random and therefore the possibility that the algorithm falling into a local solution, the error obtained is much lower than expected.

InFig. ***2***, the graph iteration error is shown.

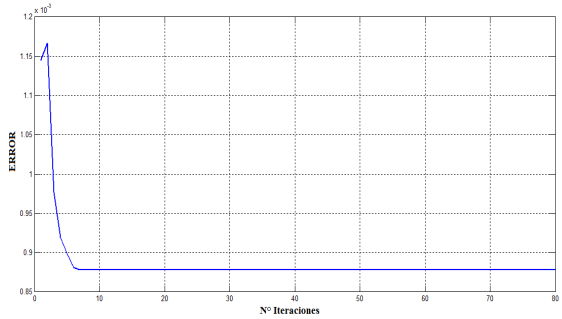

Fig. 2: Graphical error by iterating

In the chart above you can clearly see, the decrease of error in each iteration of the algorithm, which can check the correctness of groups and centroid to get the most optimal.

The final centroids are presented in

Table ***6***.

Table 6: Final centroids.

	DL50	EC50	BCF	Water. S.
Group1	0.0157	0.0219	0.0655	0.0504
Group2	0.2476	0.0742	0.0042	0.4596
Group3	0.0132	0.4387	4.31e-5	0.0145

A total of 83 organochlorine pesticides, organophosphates and Carbamate were classified. Finally as a result were classified according to their main toxicological properties as shown in

Table ***7*** below.

Table 7: Pesticides Classification.

	N Pesticides	Class 1	Class2	Class3
O-chlorines	13	5	2	6
O-phosphates	38	34	1	3
Carbamates	32	24	8	0

Conclusions:

As a result of the regression it is identified that the following properties have a positive correlation with the lethal dose (effective concentration, mass density, BCF and water solubility, whereas the inhibitory concentration maximum and the log P showed a negative correlation, which led to discard them in the execution of the algorithm.

The algorithm allowed to classify 83 pesticides within which 13 organochlorine were, 38 organophosphates and 32 Carbamate were, with an error of 0.00087, the cluster algorithm classify into 3 different classes represented by the level of toxicity of pesticides, being the lowest as a class 1 and class 3 the highest toxicity.

The distribution of pesticides classified in group 1 corresponding to a low toxicity, present 5 of them organochlorine, 34 organophosphates and 24 Carbamate, indicating that 38.46% of organochlorine pesticides, 89.47% of organophosphates and 75% Carbamate belonging to group 1.

In the middle group is the toxicity of pesticides 15.38% organochlorine, organophosphates 2.63% and 25% of Carbamate. Finally the highest toxicity is distributed with the remaining 46.15% of organochlorine pesticides, 7.89% of organophosphates and Carbamate 0%, the only one not classify any of their pesticides likehigh toxicity.

Contribution:

With the development of the work presented, based on computational techniques with the implementation of scheduling algorithms for classifying pesticides according to their toxicity, greatly reducing the cost and time of analysis usually requires a study like this model proposed and that limited by the high costs on existing technology and restrictions thereof. The model is versatile because it relates main variables, so this can be improved and updated with new relationships and / or various parameters associated with the toxicity of these chemicals.

ACKNOWLEDGMENT

The authors would like to thank the Nueva Granada Military University research center for financing this work (research project IMP-ING-1777, 2016).

REFERENCES

[1] Schaaf, A., 2013. «Uso de pesticidas y toxicidad: relevamiento en la zona agrícola de San Vicente, Santa Fe, Argentina,» *Revista Mexicana de ciencias agricolas,* 4(nº 2).

[2] Tsatsakis, A., M. Christakis-Hampsas, Y.J. Liesivuori, 2012. «The increasing significance of biomonitoring for pesticides and organic pollutants.,» *Toxicology Letters,* 210(nº 2).

[3] Azmiy, M., S. Naqvi, 2011. «Pesticide pollution, resistance and health hazards,» *Pesticides-the impacts of pesticide exposure.*

[4] Mora, P.A., J.C. Pampilloy, K.R. Hidalgo, 2014. «Uso del "Pesticide Impact Rating Index"(PIRI) para estimar la peligrosidad del uso de plaguicidas en el cultivo de cebolla (Allium cepa L.).,» *Mundo de la salud,* 38(nº1).

[5] Jaramillo,C.B., I. MarteloEy, E. Duarte R, 2013. «Toxicidad aguda de pesticidas órganO-phosphates y análisis de la relación cuantitativa de estructura actividad (QSAR),» *Biotecnología en el Sector Agropecuario y Agroindustrial,* 11(nº2).

[6] Cajamarca Cango, M.E., 2015. «Aplicación de métodos QSAR y quimiométricos para determinar la relación estructura-actividad de pesticidas en mamíferos (Rata),» Cuenca.

[7] Durán, V.B., E. de la Cruz Malavassi, G.H. Ledezmay, F.R. Muñoz, 2013. «Uso de plaguicidas en cultivos agrícolas como herramienta para el monitoreo de peligros en salud.,» *Uniciencia,* 27(nº 1).

[8] Lowy, C., 2013. «La construcción social de los sistemas agroalimentarios–Los pesticidas.,» de *In X Jornadas de Sociología.*, Buenos Aires.

[9] Cárdenasy, L., P. Tripaldi, 2013. «Modelización de la estructura-toxicidad de diveros pesticidas utilizando métodos QSAR,» de *Memorias del II Congreso Binacional de Investigación, Ciencia y Tecnología de las Universidades.*

[10] Zavala, A., J. Gonzalesy, L. Chacon, 2013. «Estudio QSAR por algoritmos geneticos de reconocedores de surco del DNA,» *Biológicas.*

[11] Colosio, C., F. Rubino, M. Alegakis, E. Ariano, G. Brambilla, S. Mandic-Rajcevic, F. Metruccio, C. Minoia, A. Moretto, C. Somaruga, A. Tsatsakis, R. Turciy, F. Vellere, 2013. «Integration of biological monitoring, environmental monitoring and computational modelling into the interpretation of pesticide exposure data: Introduction to a proposed approach.,» *Toxicology Letters,* 213(nº 1).

[12] Gonzálezy, J.L.G., O.M. Cornelio, 2013. «Propuesta de algoritmo de clasificación genética.,» *Revista Cubana de Ingeniería,* 4(nº 2).

[13] Sánchez, J., H. Contrerasy, M. Figueroa, 2015. «Análisis del desempeño del algoritmo genetico en la clasificación automática de documentos,» Research in Computing Science, 93.

[14] Cárdenas, F., P. Tripaldiy, C. Rojas, 2014. « Estudio de la Relación Cuantitativa Estructura-Actividad de pesticidas mediante técnicas de clasificación,».
[15] Peña, M.J., J.C. Castroy, A. Soto, 2013. «Evaluación de insecticidas no convencionales para el control de aphis gossypii glover (hemiptera: aphididae) en fríjol,» *Rev. U.D.C.A Act. & Div. Cient.,* 16(nº 1).
[16] Julián Duran, V., 2015. «Estudio del efecto tóxico de dos contaminantes en la especie Donax Trunculus mediante el ensayo de retención del rojo neutro,».
[17] Gonzálezy, E., L. Gómez Olivan, 2013. «Evaluación de la toxicidad inducida por antinflamatiorios no esteroides presentes en agua de presa madin sobre Cyprinus Carpio,».
[18] Sainiy, G., H. Kaur, 2014. «A Novel Approach Towards K-Mean Clustering Algorithm With PSO.,» *International Journal of Computer Science & Information Technologies*.
[19] Cleger Tamayo, S., 2013. «Diseño y validación de modelos para sistemas de recomendación.,» Granada.
[20] AhamedShafeeqy, B., K. Hareesha, 2012. «Dynamic Clustering of Data with Modified K-Means Algorithm,» de *International Conference on Information and Computer*, Singapore.
[21] Angélica Palacios, Olga Lucia Ramos Sandoval, Darío Amaya Hurtado,2016. Design And Implementation Of A Neural Network Trained By Particle Swarm To Estimate Retention Times Frequently Used Chemicals In Pesticides. AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE. 10(1); Pages: 47-55.
[22] Jorge Eliecer Buitrago Salazar, Olga Lucia Ramos Sandoval , Dario Amaya Hurtado,2016. Analysis of displacement of compounds used as pesticides in *Annona Muricata* through the simulation of a liquid chromatography column. AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE. 10(2); Pages: 33-42.
[23] Angélica Palacios, Olga Ramos, Darío Amaya,2016. Estimation Of Organochlorine Pesticides In Humans By The Consumption Of Commercial Milk Using Artificial Intelligence (Ai). AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE. 10(2); Pages: 9-15.
[24] *Angélica Palacios, Olga Lucia Ramos Sandoval, Darío Amaya Hurtado*,2016. Prediction Algorithm For Renal System Damage On Human Health By Eating Food With Pesticide Residue**s**. AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE. 10(3); Pages: 9-16.
[25] *Angélica Palacios, Olga Lucia Ramos Sandoval, Darío Amaya Hurtado*,2016. Development Of A Diagnostic Algorithm Of Renal System Damage On Human Health From Exposure To Pesticides. AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE. 10(3); Pages: 26-32