



Prediction of Endocrine System Affection in Fisher 344 Rats by Food Intake Exposed with Malathion, Applying Naïve Bayes Classifier and Genetic Algorithms

Juan David Sandino Mora, Darío Amaya Hurtado, Olga Lucía Ramos Sandoval

Department of Mechatronics Engineering, Faculty of Engineering, Nueva Granada Military University, Bogotá, Colombia

Correspondence to:

Dr. Juan David Sandino, Mechatronics Engineering Program, Faculty of Engineering, Nueva Granada Military University, Bogotá, Colombia.
E-mail: u1801731@unimilitar.edu.co

How to cite this article: Mora JD, Hurtado DA, Sandoval OL. Prediction of endocrine system affection in fisher 344 rats by food intake exposed with malathion, applying naïve bayes classifier and genetic algorithms. *Int J Prev Med* 2016;7:111.

ABSTRACT

Background: Reported cases of uncontrolled use of pesticides and its produced effects by direct or indirect exposition, represent a high risk for human health. Therefore, in this paper, it is shown the results of the development and execution of an algorithm that predicts the possible effects in endocrine system in Fisher 344 (F344) rats, occasioned by ingestion of malathion.

Methods: It was referred to ToxRefDB database in which different case studies in F344 rats exposed to malathion were collected. The experimental data were processed using Naïve Bayes (NB) machine learning classifier, which was subsequently optimized using genetic algorithms (GAs). The model was executed in an application with a graphical user interface programmed in C#.

Results: There was a tendency to suffer bigger alterations, increasing levels in the parathyroid gland in dosages between 4 and 5 mg/kg/day, in contrast to the thyroid gland for doses between 739 and 868 mg/kg/day. It was showed a greater resistance for females to contract effects on the endocrine system by the ingestion of malathion. Females were more susceptible to suffer alterations in the pituitary gland with exposure times between 3 and 6 months.


Conclusions: The prediction model based on NB classifiers allowed to analyze all the possible combinations of the studied variables and improving its accuracy using GAs. Excepting the pituitary gland, females demonstrated better resistance to contract effects by increasing levels on the rest of endocrine system glands.

Keywords: Artificial intelligence, machine learning, organophosphate, rat

INTRODUCTION

Organophosphate pesticides are applied quite frequently in farming applications, in the production of insecticides,

herbicides and fungicides, and especially for achieving the best performance on crops. Therefore, it is common to find clinical profiles with accidental poisonings occasioned by these products.^[1] However, there have been several studies which show the indiscriminate use of pesticides and its effects on human health, through direct or indirect exposure to these chemical compounds.^[2-5] Furthermore,

Access this article online	
Quick Response Code: 	Website: www.ijpvmjournal.net/www.ijpm.ir
	DOI: 10.4103/2008-7802.190611

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

current regulations for pesticide usage lacks on defining good management practices. It generates enough concern from local health agencies to raise awareness to properly regulate the application of pesticides.^[6-8]

Malathion is one of the most globally applied organophosphate compounds (CAS 121-75-5) which is applied to control pests on agricultural crops, public health, and residential pest control. Although the effects on the nervous and endocrine system^[9,10] and its classification in Group 2A as possibly carcinogenic compound,^[11] do not stop being troubling to health agencies to properly regulate their application.^[12-14] It is very relevant for science-based toxicology and public health to carry out researches in animals, considering that they are a primary approach for studying and improving the quality of human health, from eating poisoned food with pesticides and the prediction of health effects.^[15,16]

Therefore, the purpose of this work is to develop an algorithm with the capability to predict the effects in the endocrine system, produced for the ingestion of malathion in Fisher 344 (F344) rats, implementing and optimizing “machine learning” techniques with the application of genetic algorithms (GAs). The algorithm will have the flexibility to work with multiple databases, for the analysis of health effects in various systems of both animals and other living things, such as the study of several pesticides.

METHODS

First, it was referred to ToxRefDB database^[17] in which different studied cases in F344 rats exposed to malathion were collected. Then, according to the type of the obtained data, Naïve Bayes (NB) machine-learning classifier was selected, and it was subsequently optimized using GAs. Finally, the model was executed in an application with a graphical user interface (GUI) programmed in C#, so that the information provided in the application, is sufficiently direct and easy to interpret for the user [Figure 1].

The available information in ToxRefDB database includes several toxicity studies in animals (*in vivo*) of various chemical compounds. For this research, a study with

F344 rats was selected, given its recognized capabilities of being a model organism.^[18] In this study, malathion was supplied to them, with a purity of 97.1%, oral route, for 24 months, in adult rats of both genders.^[19] The collected data were subsequently filtered as “predictor variables,” which includes the gender, the applied dose in mg/kg/day (very low [4–5], low [29–35], medium [359–415], high [739–868]), the dose duration in months and the final alteration, if there was one, classified by the main endocrine system glands [Figure 2].

To carry out the information processing, given the nature of qualitative data, the NB classifier was implemented,^[20] assimilating independence between the predictor variables and their robustness in the application of supervised learning.^[21,22] To perform this task, the original data were classified randomly into two categories: Training data and testing data, with an initial ratio between the sets of 70% and 30%, respectively. The probability calculation (prediction) for a variable E_i , given a set of predictor variables H_n , is defined in Eq. 1.

$$P(E_i | H_1, \dots, H_n) = \frac{PP(E_i)}{\sum_{k=1}^n PP(E_k)} \quad (1)$$

where $PP(\dots)$ corresponds to the partial probability of occurrence E_i . The Eq. 2. defines such probabilities.

$$PP(E_i) = P(E_i) \prod_{k=1}^n P(H_k | E_i) \quad (2)$$

Finally, the probability of an event E_i occurs, given a H_i condition, It is set according to Eq. 3.

$$P(H_i | E_i) = \frac{\sum E_i \rightarrow H_i}{\sum E_i} \quad (3)$$

Analyzing the calculation of the partial probability, if one of the product terms is 0, the entire probability calculation will be affected accordingly. This behavior usually occurs when it is trying to calculate the probability of a variable that does not appear in the data array for training. As a result, Laplace smoothing was implemented in the method,^[23] setting the counters of each one of the unions on one.

Classifier error was divided into training and testing stages separately. To evaluate this variable, the probability for each of the original data is calculated, thus realizing

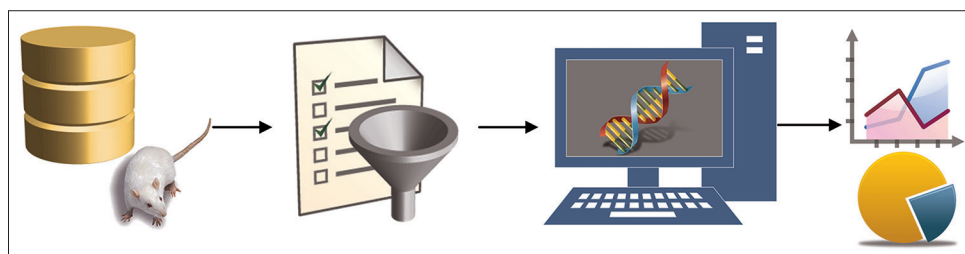


Figure 1: General scheme of performed work

a count of true positives (T_p) and false negatives (F_N), as shown in Eq. 4.

$$e = 1 - \frac{T_p}{T_p + F_N} \quad (4)$$

It is considered a T_p if the likelihood of the analyzed variable is higher than 50%. Finally, the classifier error was calculated in Eq. 5.

$$NB_{error} = \frac{e_{train} + e_{test}}{2} \quad (5)$$

Moreover, NB classifier was optimized through GAs to improve its robustness [Figure 3]. The initial population was constituted with multiple and random NB classifiers. Distribution of data intended for training (P_T) and test ($1 - P_T$) was modified, along with the location of each value data in these categories, altering the value of a constant, known as the random seed value (S_A). The proportion of the training data was set up between 40% and 80%, and the random seed value was comprised between 0 and 5000.

The most relevant parameters in GA settings included to manage an initial population with a random distribution, the calculation of the accuracy as the phenotype for each individual, processed with real numbers thousandths precision. Furthermore, a 14-bit resolution for the management of genotypes, the application of the gray code to perform genotype – phenotype conversion and evolution techniques such as roulette, crossover, elitism, and mutation (altering one gene for each genotype selected). The full development for the selection of the best classifier for each generation of individuals is depicted in Figure 4. The algorithm was programmed and executed for PC with technical specifications such as Intel® Core™ i5-2500 processor (4 cores at 3.3 GHz), 8GB DDR2 RAM, and Windows 8.1 × 64 operating system.

The algorithm was implemented in a GUI, which was developed in C# with the ability to read files from databases automatically, adding the sets of the predictor variables for the diagnosis of interest. In addition, the GUI allows to modify the number of individuals in the population, the proportion of evolution

techniques (elitism, crossover, and mutation), and the stop criteria (number of iterations and tolerance).

RESULTS

According to Table 1, a proportional directly tendency is reflected in the probability rates to suffer alterations by increasing levels in all the studied glands, as long as the dosage and the duration of exposure increases and vice versa. Nonetheless, at comparing these predictor variables while observing the variations, it is visualized that the first one (dosage) is not so prevalent than the second (duration). Among the effects on the endocrine system glands, there will be a trend to suffer bigger alterations in the parathyroid gland for very low dosages (4–5 mg/kg/day) if the duration of exposure increases.

In addition, it is contemplated that the chance of suffering alterations in thyroid gland increase while extending the dosage, followed by parathyroid and pituitary gland. Besides, it is perceived that female gender

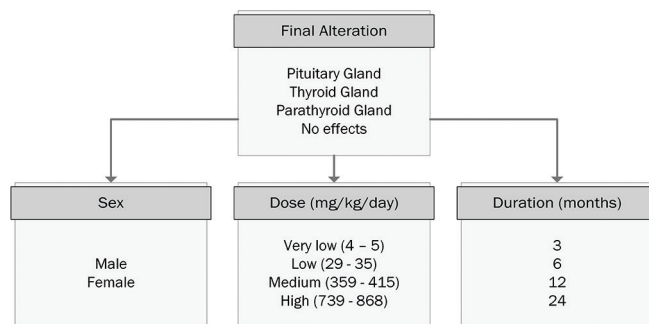


Figure 2: Filtered data for effects prediction by ingestion of Malathion

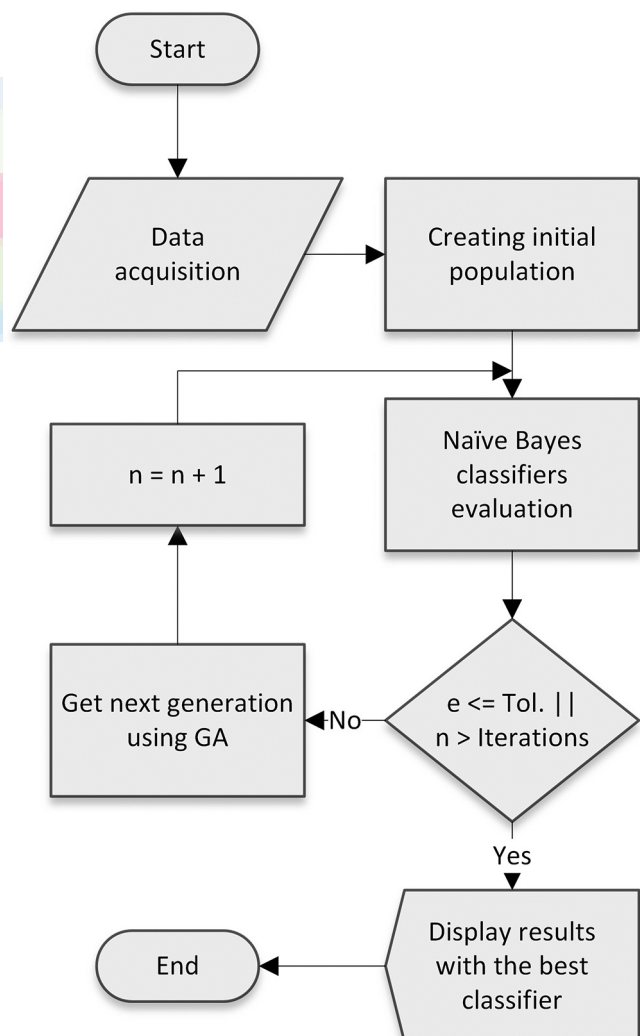


Figure 3: Implemented algorithm for NB classifier optimization

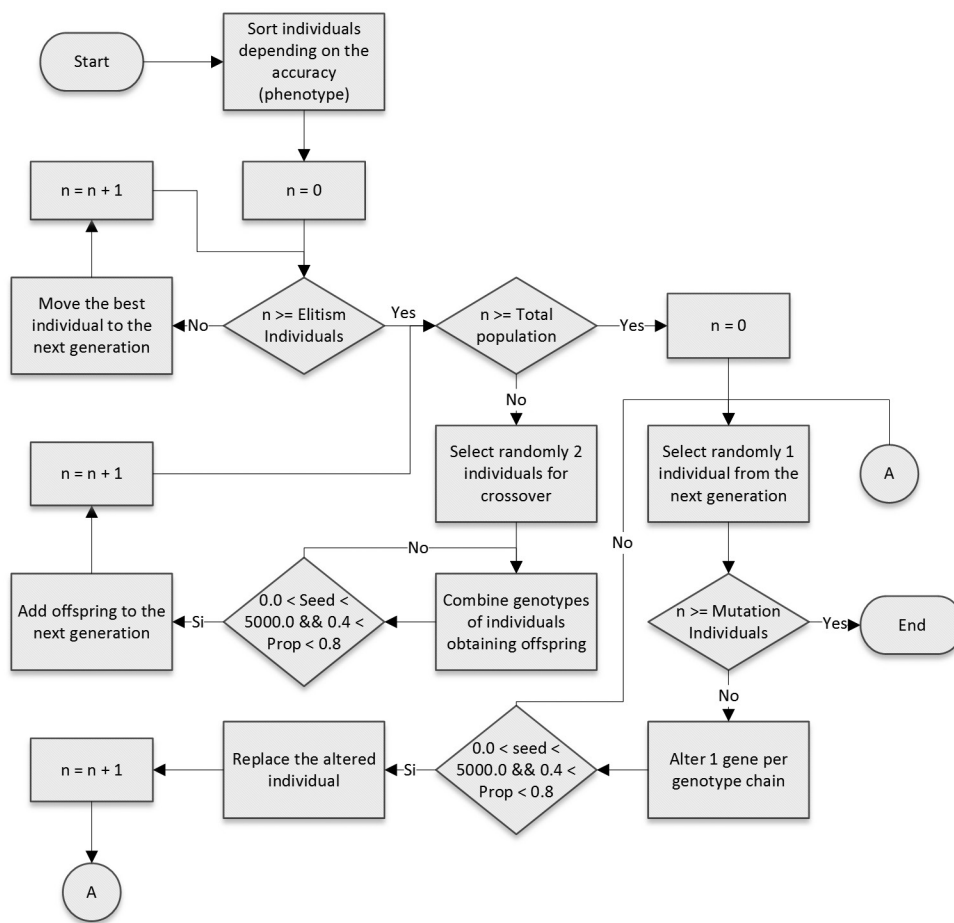


Figure 4: Process to obtain the next generation of individuals (classifiers)

is more likely to suffer alterations in pituitary gland with low exposure times (3–6 months), without taking into account the dosage level. Finally, a faint greater resistance is observed in the female gender to suffer effects on the endocrine system by ingestion of malathion.

DISCUSSION

The evolution of the population (classifiers) was observed, modifying the number of individuals and the proportion of evolution techniques, as depicted in Figure 5. In cases which low population, the evolution of the classifiers was almost uncontrollable, regardless of the proportions of the evolution techniques [Figure 5a]. A small number of individuals greatly increase the genotypic variations that best individuals suffer, affecting the evolution of the population with the passage of various generations. The ideal situation is shown in Figure 5b, which was developed with 25 individuals and proportions of elitism, crossover, and mutation of 15%, 85%, and 30%, respectively. If the number of individuals is remarkably high [Figure 5c], the evolution of the population just gets slight variations. This occurs because there are

better chances of individuals in view of its high density in the initial population; however, the rate of evolution is inadequate, and except for substantially increase the mutation rate (>40%), the population is susceptible not to evolve over generations.

In terms of performance, the proposed algorithm copes adequately while varying population density [Table 2]. It is observed that GA suitably modified ratio data (training – testing); if the number of individuals is low, the proportion of data will be better balanced and vice versa. In addition, the algorithm converges at approximately the same value with different seeds for the random distribution of the original data. Finally, a trend to improve the performance of the algorithm as n increases is seen, due to a lower probability that restrictions may reject individuals for the next generations.

The error introduced on the algorithm is significantly attenuated by the execution of GAs in the structure of NB classifier, unlike the normal execution of classifier, where the magnitude of this variable may be better or worse if the seed distribution and data proportion are chosen manually. Due to the limited data available, which

Table 1: Probability results of the algorithm of predicting increased levels in endocrine system glands

Dose (mg/kg/day)	Duration (months)	Gender	Parathyroid gland (%)	Pituitary gland (%)	Thyroid gland (%)	Number effects (%)
Very low (4-5)	3	Male	11.01	3.60	2.81	82.58
		Female	5.91	9.26	2.11	82.72
	6	Male	15.19	4.96	3.88	75.97
		Female	8.16	12.79	2.91	76.14
	12	Male	38.66	3.16	9.87	48.32
		Female	24.49	9.60	8.75	57.15
24	Male	64.11	6.98	20.00	8.90	
	Female	45.07	23.56	19.68	11.69	
Low (29-35)	3	Male	12.76	4.17	6.51	76.56
		Female	6.90	10.83	4.93	77.33
	6	Male	17.13	5.60	8.74	68.53
		Female	9.30	14.59	6.65	69.46
	12	Male	38.58	3.15	19.69	38.58
		Female	25.17	9.87	17.98	46.98
24	Male	54.23	5.91	33.83	6.03	
	Female	38.41	20.08	33.55	7.97	
Medium (359-415)	3	Male	16.49	6.46	17.68	59.37
		Female	9.01	16.95	13.52	60.53
	6	Male	20.56	8.06	22.04	49.34
		Female	11.28	21.23	16.93	50.55
	12	Male	36.10	3.54	38.70	21.66
		Female	24.44	11.50	36.68	27.38
24	Male	39.88	5.21	52.25	2.66	
	Female	27.89	17.49	51.15	3.47	
High (739-868)	3	Male	19.69	9.65	26.38	44.29
		Female	10.61	24.95	19.90	44.55
	6	Male	23.10	11.32	30.94	34.64
		Female	12.46	29.30	23.37	34.88
	12	Male	35.24	4.32	47.22	13.22
		Female	24.02	14.12	45.05	16.81
24	Male	35.18	5.75	57.61	1.47	
	Female	24.07	18.87	55.19	1.87	

Table 2: Algorithm performance for 50 generations

n	Elapsed time (ms)	Mean (ms)	SD (ms)	SEM 1/√(ms)	Seed of random value	Training proportion (%)	Classifier error (%)
15	553	11.06	14.99	3.87	1476	66.40	11.95
20	70	1.40	0.61	25.69	4314	65.20	13.33
25	166	3.32	6.63	9.71	2829	75.60	11.90
30	123	2.46	1.05	29.22	637	78.90	11.36
35	100	2.00	0.99	35.18	2521	78.20	12.34
40	77	1.54	0.71	47.61	509	78.40	11.99

SD=Standard deviation, SEM=Standard error of mean

combinations can be omitted for the prediction of effects on the endocrine system, the average magnitude of error is considered feasible or appropriate for this study.

Researches focused on identifying and predicting health effects from the consumption of pesticides, such as those

presented by Rayo *et al.*, Altamirano *et al.* and EPA,^[24-26] show the usefulness of the application of NB classifiers and neural networks for such tasks. Nevertheless, it is observed that the accuracy of implemented techniques descends slightly by evaluating a large amount of predictor variables. The current research work improves the accuracy of these previous studies, by adding GA techniques and executing it in the proposed algorithm. Researches based on epidemiological models^[27] are limited only to know the history and effects of poisoning in a particular population in which was presented this phenomenon, which are employed as complementary techniques in such studies. A prediction model based on NB classifiers allows to analyze all possibilities of predictor variables.

The current work can be used to complement the study perform by Campetelli *et al.*,^[28] which relation between a

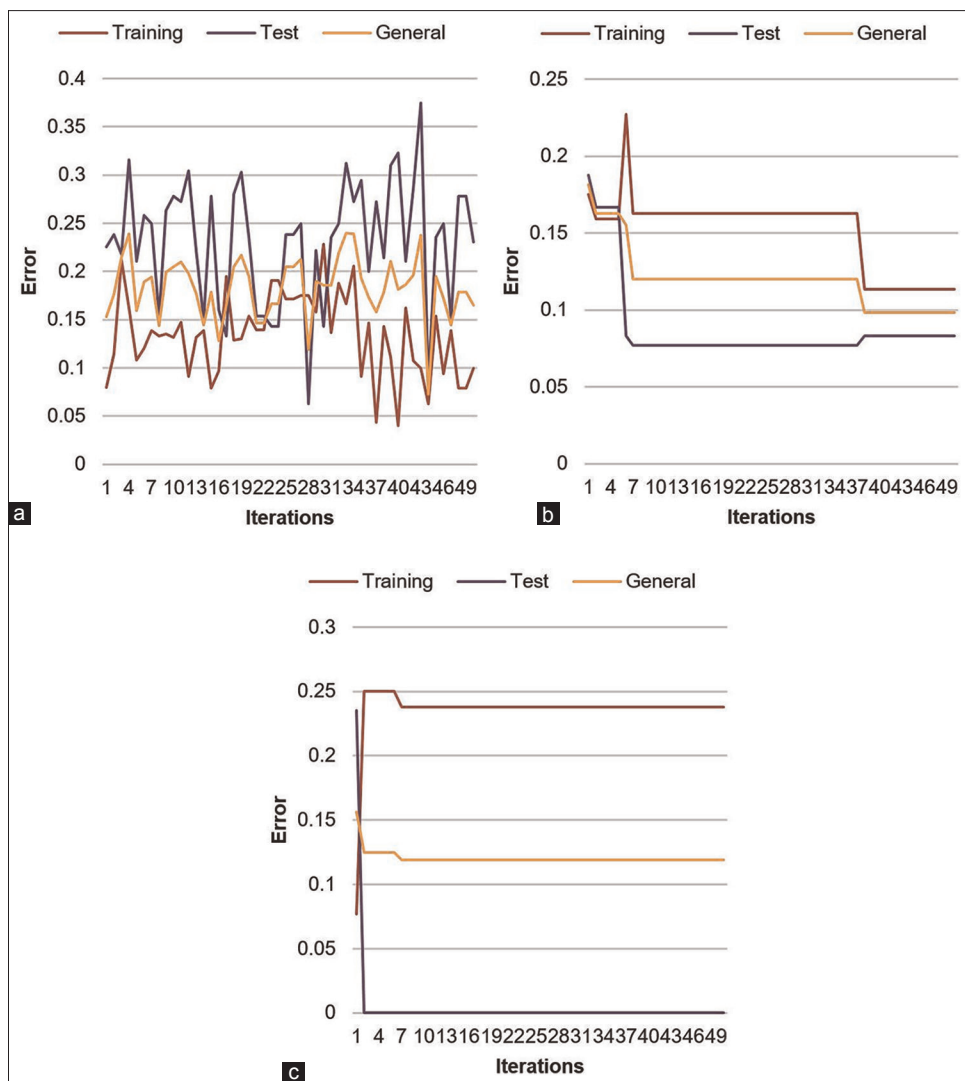


Figure 5: (a) Evolution of the population using various parameters (b) evolution of the population using various parameters (c) evolution of the population using various parameters

rat-human mathematical model for the endocrine system, will be useful to improve and explore more prediction tasks in discovering alterations in glands, with the advantage of decrease test in animals.

CONCLUSIONS

It was found a tendency to suffer bigger alterations, increasing levels in the parathyroid gland in dosages between 4 and 5 mg/kg/day, in contrast to the thyroid gland for doses between 739 and 868 mg/kg/day, with greater resistance for females to contract effects on the endocrine system by the ingestion of malathion. A trend was found in female gender to suffer alterations in the pituitary gland with exposure times between 3 and 6 months, no matter of dosage level. An algorithm to predict effects in the endocrine system by ingestion of malathion was developed and executed, properly combining the NB classifier and GAs to optimize its accuracy.

Acknowledgements

The authors would like to offer their special gratitude to the Research Vice-chancellorship of Nueva Granada Military University, for funding the research project IMP-ING 1777.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Received: 21 Oct 15 **Accepted:** 20 Aug 16

Published: 14 Sep 16

REFERENCES

1. Fernández DG, Mancipe LC, Fernández DC. Organophosphorus poisoning. *Rev Med* 2010;18:84-92.
2. Dich J, Zahm SH, Hanberg A, Adami HO. Pesticides and cancer. *Cancer Causes Control* 1997;8:420-43.
3. González-Álvarez YC. Reported chemical poisoning on the

- Epidemiological Monitoring System - Sivigila. Bogotá District Secretariat of Health; 2011.
4. Gomez-Arroyo S, Martínez-Valenzuela C, Carbajal-López Y, Martínez-Arroyo A, Calderón-Segura ME, Villalobos-Pietrini R, *et al*. Genotoxic risk due to pesticide occupational exposure in Latin America. *Rev Int Contam Ambient* 2013;29:159-80.
 5. Domínguez-Majín LJ. Epidemiological characterization of plaguicides poisoning. *Inform Quincenal Epidemiol Nacl (IQUEN)* 2013;18:243-55.
 6. INS. Chemical poisoning. Bogotá:Vigilance and Control In Public Health; 2014.
 7. González-Vides G. Pesticide poisoning: University Hospital cases of the Caribbean and the University Clinic of San Juan de Dios Cartagena: National University of Colombia; 2010.
 8. Polanco Y, Salazar JC, Curbow B. A quantitative analysis of Colombian campesinos' use of pesticides: Perceived control and confidence in this use. *Rev Fac Nac Salud Pública* 2014;32:373-82.
 9. MedlinePlus. Malathion Poisoning. 2013. Available from: <http://www.nlm.nih.gov/medlineplus/ency/article/002834.htm>. [Last cited on 2015 Aug 11].
 10. EPA. Malathion | Mosquito Control | US EPA. 2015. Available from: <http://www2.epa.gov/mosquitocontrol/malathion>. [Last cited on 2015 Aug 19].
 11. Guyton KZ, Loomis D, Grosse Y, El Ghissassi F, Benbrahim-Tallaa L, Guha N, *et al*. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. *Lancet Oncol* 2015;16:490-1.
 12. Hernández AF, Pla A, Gómez MA, Pena G, Gil F, Pino G, *et al*. Susceptibility to organophosphorus insecticides on greenhouse workers: importance of biochemical markers. In: III conference of the Spanish Society of Ecological Agriculture. Spain: SEAE; 1998. p. 369-77.
 13. Idrovo AJ. Monitoring of pesticide poisoning in Colombia. *Rev Salud Púb* 2000;2:36-46.
 14. Orozco-Cardona RE, Ceballos C. Chemical poisoning. Medellín: Sectional Secretariat of Health and Social Protection; 2014.
 15. Committee on the Use of Laboratory Animals in Biomedical and Behavioral Research. *Use of Laboratory Animals in Biomedical and Behavioral Research*. UK: National Academies Press; 1988.
 16. TRS. *The Use of Non-Human Animals in Research: A Guide for Scientists*. London: The Royal Society; 2004.
 17. EPA. ToxRefDB | Computational Toxicology Research Program (CompTox) | Research & Development | US EPA. 2013. Available from: <http://www.epa.gov/nccct/toxrefdb/>. [Last cited on 2015 Aug 21].
 18. Lannaccone PM, Jaboc HJ. *Rats! Dis Models Mech* 2009;2:206-10.
 19. Daly IW. A 24-Month Oral Toxicity/Oncogenicity Study of Malathion in the Rat via Dietary Administration. Report No.: 90-3641:J-11 90-3641. Lemvig, Denmark: Huntingdon Life Sciences. Cheminova Agro A/S; 1996.
 20. Zhang H. Exploring conditions for the optimality of Naive Bayes. *Intern J Pattern Recognit Artif Intell* 2005;19:183-98.
 21. Caruana R, Niculescu-Mizil A. An Empirical Comparison of Supervised Learning Algorithms. In: 23rd Intl. Conf. Machine Learning; 2006. p. 161-8.
 22. Witten I, Frank E, Hall M. Ch. 4.2. Statistical modeling. *Data Mining: Practical Machine Learning Tools and Techniques*. United States: Morgan Kaufmann; 2011. p. 93.
 23. Manning CD, Raghavan P, Schütze H. *Text classification and Naive Bayes. An Introduction to Information Retrieval*. Ch. 13. Cambridge, England: Cambridge University Press; 2009; p. 260.
 24. Mishra M, Fei H, Huan J. Computational prediction of toxicity. In: *Bioinformatics and Biomedicine (BIBM)*. IEEE International Conference on 2010. IEEE; 2010. p. 686-91.
 25. Mishra M, Potetz B, Jun H. Bayesian classifiers for chemical toxicity prediction. In: *Bioinformatics and Biomedicine (BIBM)*. IEEE International Conference on 2011. Atlanta: IEEE; 2011. p. 595-9.
 26. Rayo R, Espinosa G, Giralt F. Using an ensemble of neural based QSARs for the prediction of toxicological properties of chemical contaminants. *Process Saf Environ Prot* 2005;83:387-92.
 27. Altamirano JE, Franco R, Bovi-Mitre MG. Epidemiological model for the diagnosis of acute intoxication by pesticides. *Rev Toxicol* 2004;4:98-102.
 28. Campetelli G, Lombarte M, Biset H, Rigalli A, Basualdo MS. A rat-human scale-up procedure for the endocrine system. *Comp Chem Eng* 2014;71:512-20.

