

Prediction of Reproductive System Affection in Sprague Dawley Rats by Food Intake Exposed with Fenthion, Using Naïve Bayes Classifier and Genetic Algorithms

Juan David Sandino¹, Dario Amaya Hurtado² and Olga Lucia Ramos³

Virtual Applications Group-GAV, Universidad Militar Nueva Granada –UMNG.

<http://dx.doi.org/10.13005/bbra/2572>

(Received: 24 October 2017; accepted: 15 November 2017)

Improper application of pesticides in agricultural crops and indirect effects caused by exposure to them through consumption of contaminated crops, nowadays represent a serious risk to public health harmony. It is vital then, to know the degree of toxicity of each of these chemicals in order to properly regulate its application and sensitize the population at risk. Therefore, this paper shows the results of an algorithm with the ability to predict the effects on the reproductive system in Sprague Dawley rats, caused by the intake of food exposed with Fenthion. The original data were processed using the Naïve Bayes classifier, then optimized using genetic algorithms. It is concluded that the prediction algorithm does the job properly, processing qualitative information with relatively low computational cost, which allows its easy portability to different development platforms.

Keywords: Artificial intelligence, Machine learning, Organophosphate.

Currently, given the food demand in the global market, the use of pesticides in agriculture has been essential to achieve optimal crop yields. In order to show the impact of pesticides on health, it has been published results from studies that justify how these compounds are applied indiscriminately, showing the effects produced by direct or indirect exposure,^{1, 2, 3, 4}. Though, regulators for pesticide application, do not define good pesticide management practices, generating concern in local prevention agencies to adequately train the farmers in order to avoid health consequences^{5, 6, 7}. Additionally, it is common to find patients with clinical pictures of accidental poisonings caused by these products, so it is vital timely care of patients to minimize the risks and

consequences on human health in such accidents⁸.

The Fenthion (CAS 55-38-9) is an organophosphate applied to pest controls on agricultural crops, cases of public health and residential pest control. Despite being a compound of fast degradation in the environment, its effects on the reproductive system^{9,10}, do not stop being troubling to health agencies to properly regulate their application. Hence, the great importance to run practices for the prevention of poisoning, such as promoting education about good management practices and awareness of health risk for poisoning, such as to convey the importance of properly monitoring the implementation of a pesticide in food^{11, 12, 13}.

*Corresponding author E-mail:

This is an  Open Access article licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted Non Commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

Published by Oriental Scientific Publishing Company © 2017



The scope of medicine and toxicology as we know it today, would not have had the same impact if not for scientific experimentation on animals. So, it is essential to develop studies which can predict effects from eating contaminated food with pesticides, in each one of the systems of these living beings, since they are considered an essential approach to the study and improvement of the quality of human health and ensure the safety of public health^{14, 15}.

Some research focused on identifying and predicting effects on health, by the consumption of pesticides^{16,17,18}, demonstrate the usefulness of the application of machine learning techniques, such as the execution of Naïve Bayes classifiers and neural networks. Inversely, it is noted that the accuracy of implemented techniques descends slightly to treat a high number of predictor variables.

Based on the above, this study aims to develop an algorithm to predict the effects on the reproductive system, caused by the intake of Fenthion in Sprague Dawley rats by implementing machine learning techniques, optimized under the application of genetic algorithms. This algorithm will have the flexibility to analyze multiple databases, both for the study of health effects in animals like in humans, for several commonly used pesticides or interest in the region. Finally this technique will have the ability to be ported easily to other development platforms that allow easily distributing information to the population in general, thus transmitting the importance of adopting good management practices and to warn of the effects due to direct exposure of pesticides.

METHODS AND MATERIALS

First, data were collected from ToxRefDB database¹⁹, which presents results of toxicity studies in different *in vivo* animals, according to the chemical of interest. For this work, the available information of a laboratory test with Sprague Dawley rats was processed, given its recognized skills to be a model organism²⁰. In this study it was supplied Fenthion, with a purity of 96.9%, orally, by a period of 10 weeks, to those rodents of both sexes²¹.

The collected information was filtered to obtain the predictor variables and the variable

to predict. In the first case was analyzed sex, the applied dose (mg / kg / day) and the generation in which changes were observed in rats. In the opposite case the effect or final alteration in the rodent was filtered, if it existed, during the execution of the experiment (Figure 1).

The available information processing is developed by applying the Naïve Bayes classifier (NB)²², taking advantage of the ability to work with qualitative data, pretending that the predictor variables are independent from each other, and their great results in supervised learning applications^{23, 24}. The original data were classified randomly into two groups: the first devoted to the training phase of the classifier and the second to test it. Under these assumptions and applying Bayes' theorem, calculating probability of an event occurring E_i , given a condition H_i , it is set according to the Ec. 1.

$$P(E_i | H_i) = \frac{P(H_i | E_i)}{P(E_i)} \quad \dots(1)$$

For this situation, the NB classifier, assimilates the probability (prediction) for a variable E_i (Final alteration), given a set of predictor variables H_1 , H_2 y H_3 (sex, dose and generation) is defined in the Ec. 2 and 3.

$$P(E_i | H_1, H_2, H_3) = \frac{PP(E_i)}{PP(E_1) + PP(E_2) + PP(E_3)} \quad \dots(2)$$

$$PP(E_i) = P(E_i | H_1) * P(E_i | H_2) * P(E_i | H_3) \quad \dots(3)$$

Therefore, it appears that if one of the terms $P(E_i | H_k)$ in the product it is equal to zero (0), the entire calculation of the partial probability will be affected and consequently will distort the final calculation. These cases are presented when the probability of a variable that does not appear in the array of data available in the training stage is calculated. The solution to this problem was to apply Laplace smoothing²⁵, in which counters each one of the joints are started in one (1).

Evaluating the accuracy of the classifier NB it is also partitioned according to the training and testing sets. For that, the probability for each of the original data in each set was calculated and the count of true positives and false negatives were performed using Ec. 4. The analyzed variable was considered true positive as long as their likelihood is greater than 50%. Finally, the accuracy was calculated for the classifier through the Ec. 5.

$$Ac_P = \frac{VP}{VP + FN} \dots(4)$$

$$Ac = (Ac_{Train} * 0.5 + Ac_{Test} * 0.5) \dots(5)$$

In order to improve the accuracy of the NB classifier, it was applied genetic algorithms based techniques, taking several classifiers as individuals in a population. To optimize the accuracy, modified variables corresponded to the distribution of data for training and testing of each classifier, along with the location thereof in these two categories, altering the constant value of the random seed implemented for data distribution.

Among the most relevant parameters are included to handle an initial population randomized, real numbers are managed in the phenotypes of individuals, with genotype length of 14 bits, genotype – phenotype conversion through the gray code, roulette crossing technique, elitism technique and mutation technique, altering one

(1) random gene for each genotype of the selected individual. The algorithm was programmed and run on a PC, which most relevant technical specifications included Intel® Core™ i5-2500 processor (4 cores at 3.3GHz), 8GB RAM and Windows 10 x64 operative system. At Figure 2 it is depicted the pseudocode of the proposed algorithm.

RESULTS AND DISCUSION

The proposed algorithm was implemented in an application programmed in C#, which accesses and processes the database information contained in .xlsx files, with the ability to automatically add the sets of each of the predictor variables for predicting the desirable effect. Within the input parameters to optimize the classifier, it is possible to adjust 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). In Table 1 the results of each of the possible combinations for prediction of effects on the reproductive system are shown.

Firstly it is observed that alterations by exposure to Fenthion vary directly proportional by increasing the dose and vice versa, regardless whether alterations are present or not on the reproductive system. At low dosage levels, the odds of not suffering effects for future generations increase significantly. Secondly, it is perceived that male rats show a greater resistance to female to suffer effects on the reproductive system,

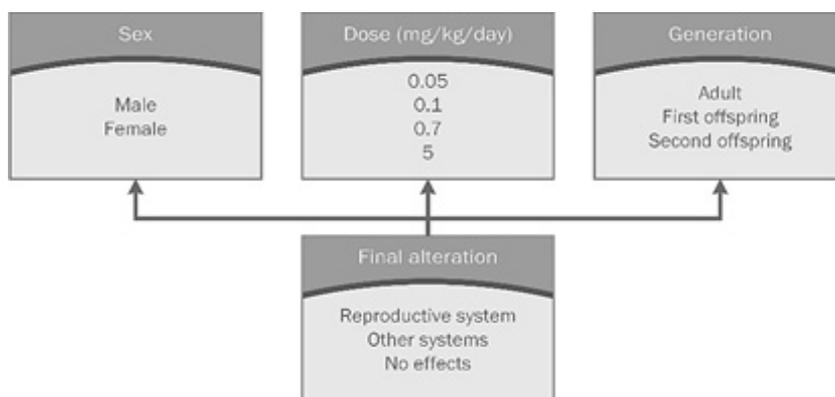


Fig. 1. Filtered data for effects prediction by ingestion of Fenthion

Input data:

P_E, P_M :	Percentage of elitism and mutation
n :	Number of Individuals
it, ϵ :	Stop conditions (iterations and tolerance)

1. Initialization

- 1.1. Load database to the application
- 1.2. Creating the initial population randomly
- 1.3. $i = 0$

2. Handling NB classifiers using GA

- 2.1. Sort individuals from high to low, depending on the accuracy Ac (phenotype)
- 2.2. Move the amount $(n * P_E)$ of ordered individuals to the next generation
- 2.3. Select randomly two individuals of the current generation
- 2.4. $Cc(0.0 < S_A < 5000.0)$ in order to obtain descendants
If ($0.0 < S_A < 5000.0$) and $(40\% < P_T < 80\%)$, add offspring to the next generation and count $n - (n * P_E)$ individuals, else, go to step 2.3
- 2.5. Select randomly one individual of the next generation
- 2.6. Genetically alter one gene of such individual
- 2.7. If $(0.0 < S_A < 5000.0)$ and $(40\% < P_T < 80\%)$, replace the original individual with the mutated one and count $(n * P_M)$ individuals, else, go to the step 2.5

3. Evaluation

- 3.1. Select the best individual (classifier) of the current generation
- 3.2. If $(Ac > \epsilon)$ or $(i < it)$; $i = i + 1$; go to step 2
- 3.3. Use the classifier for predicting effects

End**Fig. 2.** Proposed algorithm for effects prediction

regardless of the dose and studied generation. In addition, the probability of suffering effects on the reproductive system is considerably higher in the parents of the first generation, while the following generations are susceptible to effects on other systems, a factor which is also influenced by the dosage level. Finally it is reflected in the different dose levels, the likelihood of having abnormalities in the reproductive system is virtually unchanged, while probabilities of having effects in other systems shows a remarkable variation, increased over the future generations.

Moreover, NB classifier optimization was analyzed through the variation of the population and the proportion of evolution techniques. It were tested cases with low, normal and high quantity of individuals (10, 25 and 60 individuals

respectively). In the first case, the evolution of the population showed no tendency to optimize the classifier error, irrespective of the proportions of the evolution techniques. The ideal situation is noticed in the second case, shown in Figure 3, in which it was executed with proportions of elitism, crossover and mutation of 15, 85 and 30% respectively, with a general error of 2.25%, random seed 1916 and a proportion of "training - test" of 51.00% and 49.00% respectively. Finally, in the third case only small variations are observed in the error of the classifier, watching a population that hardly evolved over the generations.

Performance and behavior of the algorithm for classifier optimization was performed appropriately, switching the number of individuals in the population, as depicted in Table 2. For

example, it is contemplated that the ratio used in training and testing data suitably ranges; if the number of individuals is low, the proportion of data is better balanced and vice versa. In addition,

it is perceived that the accuracy of the classifier converges to the same value range using different seeds for the random distribution of the original data.

Table 1. Results of the proposed algorithm for predicting effects

Dose (mg/kg/day)	Generation	Sex	Reproductive system	Other systems	No effects
0.05	Adult	M	34.81%	18.77%	46.42%
		F	44.70%	26.14%	39.39%
	First offspring	M	22.54%	17.36%	60.10%
		F	22.89%	24.80%	52.31%
	Second offspring	M	2.72%	10.08%	87.20%
		F	2.97%	15.47%	81.56%
0.10	Adult	M	39.38%	21.24%	39.38%
		F	38.23%	29.00%	32.77%
	First offspring	M	26.52%	20.43%	53.05%
		F	26.33%	28.53%	45.14%
	Second offspring	M	3.48%	12.89%	83.63%
		F	3.74%	19.43%	76.84%
0.70	Adult	M	39.23%	56.41%	4.36%
		F	32.07%	64.87%	3.05%
	First offspring	M	30.52%	62.70%	6.78%
		F	24.51%	70.82%	4.67%
	Second offspring	M	7.39%	72.89%	19.71%
		F	5.83%	80.85%	13.32%
5.00	Adult	M	50.72%	47.87%	1.41%
		F	42.53%	56.46%	1.01%
	First offspring	M	41.60%	56.09%	2.31%
		F	33.97%	64.41%	1.62%
	Second offspring	M	12.29%	79.52%	8.19%
		F	9.37%	85.28%	5.35%

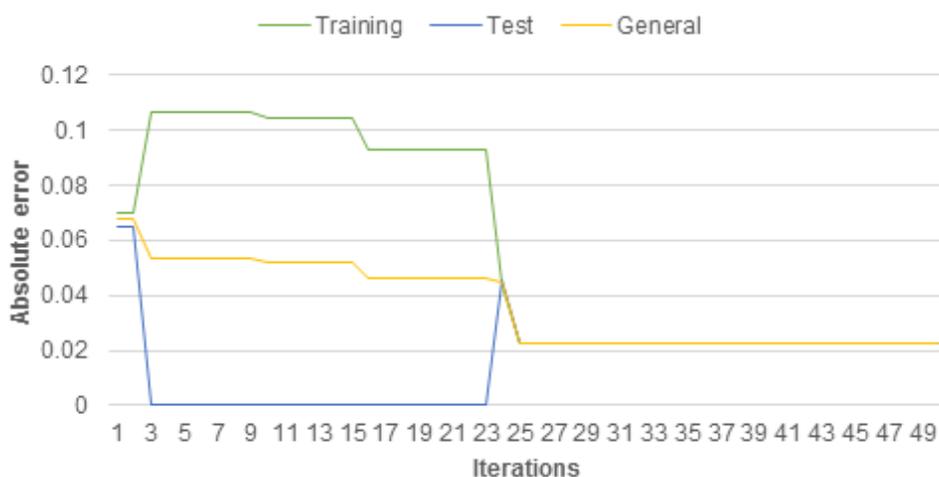


Fig. 3. Evolution of classifier accuracy over generations

Table 2. Algorithm performance for 50 generations

n	Elapsed time (ms)	Mean (ms)	Standard deviation (ms)	Standard error of the mean (ms ^{-1/2})	Seed of random value	Training proportion (%)	Classifier error (%)
15	213	4.26	0.49	71.65	1865	42.70	3.28
20	206	4.12	0.59	64.88	1508	78.90	5.71
25	204	4.08	0.27	95.51	3197	40.30	2.35
30	207	4.14	0.61	64.21	2512	48.00	2.13
35	204	4.08	0.34	85.69	3976	50.00	2.25
40	207	4.14	0.35	84.45	3726	54.90	4.17

The NB classifier error was tempered considerably by the execution of genetic algorithms, unlike if the sorter is running under normal parameters, where the magnitude of this variable may worsen from manually choose a random seed and proportion of “training - testing” data.

CONCLUSIONS

The probabilities of suffering effects on the reproductive system, are considerably higher for the first generation rats, for doses between 0.05 to 0.1 mg/kg/day, despite the risk of inheriting these effects. With the passage of generations decrease significantly, while by increasing the dosage levels ranging from 0.70 to 5.00 mg/kg/day, the risk of disease in other systems gradually increase. Regardless of the dosage and the generation studied male rats, reflect to be more resistant to suffer effects on the reproductive system.

It was allowed propose and carry out an algorithm with the ability to predict effects on the reproductive system, by the ingestion of Fenthion, through processing of qualitative and quantitative data, adopting the Naïve Bayes classifier and optimized with the implementation of genetic algorithms.

The algorithm presented in this work, supports other prediction effect analyzes for other animal species or human studies according to data availability, because this technique is robust enough to study other databases, which to predict effects on other localized systems for several pesticides.

ACKNOWLEDGMENTS

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

REFERENCES

1. J. Dich, S. H. Zahm, A. Hanberg y H. O. Adami, «Pesticides and cancer.» *Cancer Causes Control*, vol. 8, n° 3, pp. 420-443, 1997.
2. Y. C. González-Álvarez, «Intoxicaciones por sustancias químicas reportados al sistema de vigilancia epidemiológica - Sivigila.» Bogotá, 2011.
3. S. Gomez-Arroyo, C. Martínez-Valenzuela, Y. Carbajal-López, A. Martínez-Arroyo, M. E. Calderón-Segura, R. Villalobos-Pietrini y S. M. Waliszewski, «Riesgo genotóxico por la exposición ocupacional a plaguicidas en américa latina.» *Revista internacional de contaminación ambiental*, vol. 29, pp. 159-180, 2013.
4. L. J. Domínguez-Majín, «Caracterización epidemiológica de las intoxicaciones por plaguicidas.» *Informe quincenal epidemiológico nacional (IQUEN)*, vol. 18, n° 20, pp. 243-255, Noviembre 2013.
5. G. González-Vides, «Intoxicación con plaguicidas: casuística del hospital universitario del caribe y de la clínica universitaria san juan de dios de cartagena.» Bogotá, 2010.
6. INS, «Intoxicaciones por sustancias químicas - Instituto Nacional de Salud.» Bogotá, 2014.
7. Y. Polanco, J. C. Salazar y B. Curbow, «Un análisis cuantitativo del uso de plaguicidas en los campesinos colombianos: percepción del control y la confianza en este uso.» *Revista Facultad Nacional de Salud Pública*, vol. 32, n° 3, pp.

- 373-382, 2014.
8. D. G. Fernández, L. C. Mancipe y D. C. Fernández, «Intoxicación por organofosforados,» *Revista Med*, vol. 18, n° 1, pp. 84-92, 2010.
 9. MedlinePlus, 2013. [En línea]. Available: <http://www.nlm.nih.gov/medlineplus/ency/article/002834.htm>. [Último acceso: 11 8 2015].
 10. USEPA, «Interim Reregistration Eligibility Decision for Fenthion,» 9 1 2015. [En línea]. Available: <http://www.epa.gov/pesticides/reregistration/REDs/0290ired.pdf>. [Último acceso: 19 8 2015].
 11. A. F. Hernández, A. Pla, M. A. Gómez, G. Pena, F. Gil, G. Pino y L. Rodrigo, «Susceptibilidad a los insecticidas organofosforados en trabajadores de invernadero: importancia de los marcadores bioquímicos,» de *III congreso de la sociedad española de agricultura ecológica SEAE*, 1998.
 12. A. J. Idrovo, «Vigilancia de las intoxicaciones con plaguicidas en Colombia,» *Revista de Salud Pública*, vol. 2, n° 1, pp. 36-46, 2000.
 13. R. E. Orozco-Cardona y C. Ceballos, «Intoxicación por sustancias químicas,» Medellín, 2014.
 14. Committee on the Use of Laboratory Animals in Biomedical and Behavioral Research, *Use of Laboratory Animals in Biomedical and Behavioral Research*, National Academies Press, 1988, p. 112.
 15. TRS, «The use of non-human animals in research: a guide for scientists,» The Royal Society, 2004.
 16. R. Rayo, G. Espinosa y F. Giralt, «Using an ensemble of neural based QSARs for the prediction of toxicological properties of chemical contaminants,» *Process Safety and Environmental Protection*, vol. 83, n° 4, pp. 387 - 392, 2005.
 17. M. Mishra, H. Fei y J. Huan, «Computational prediction of toxicity,» de *Bioinformatics and Biomedicine (BIBM)*, 2010 *IEEE International Conference on*, 2010.
 18. M. Mishra, B. Potetz y H. Jun, «Bayesian Classifiers for Chemical Toxicity Prediction,» de *Bioinformatics and Biomedicine (BIBM)*, 2011 *IEEE International Conference on*, Atlanta, 2011.
 19. USEPA, «ToxRefDB | Computational Toxicology Research Program (CompTox) | Research & Development | US EPA,» 2013. [En línea]. Available: <http://www.epa.gov/ncct/toxrefdb/>. [Último acceso: 23 9 2015].
 20. P. M. Lannaccone y H. J. Jaboc, «Rats!,» *Disease Models & Mechanisms*, vol. 2, n° 5-6, pp. 206-210, 2009.
 21. R. Kowalski, G. Clemens y V. Jasty, «A Two-generation Reproduction Study with Fenthion (Baytex) in the Rat: Lab Proj- ect Nos. 99811; 1166; 8765,» Mobay Chemical Co., 1989.
 22. H. Zhang, «Exploring conditions for the optimality of Naive Bayes,» *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, n° 2, pp. 183-198, 2005.
 23. R. Caruana y A. Niculescu-Mizil, «An Empirical Comparison of Supervised Learning Algorithms,» de *In Proc. 23 rd Intl. Conf. Machine learning (ICML'06)*, 2006.
 24. I. Witten, E. Frank y M. Hall, «4.2 Statistical Modeling,» de *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011, p. 93.
 25. C. D. Manning, P. Raghavan y H. Schütze, «Text classification and Naive Bayes,» de *An Introduction to Information Retrieval*, Cambridge, England, Cambridge University Press, 2009, p. 260.