Prediction of Reproductive System Affectation 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 to care of patients to minimize the risks and consequences on human health in such accidents.⁸

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, 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.⁹ ¹⁰ ¹¹ ¹² ¹³
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.  

Some research focused on identifying and predicting effects on health, by the consumption of pesticides, 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 database, 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 organism. 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 sexes. 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), 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. 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.

\[
P(E_i \mid H_i) = \frac{P(H_i \mid E_i)}{P(E_i)}
\]

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

\[
P(E_i \mid H_1, H_2, H_3) = \frac{PP(E_i)}{PP(E_1) + PP(E_2) + PP(E_3)}
\]

\[
PP(E_i) = P(E_i \mid H_1) \times P(E_i \mid H_2) \times P(E_i \mid H_3)
\]

Therefore, it appears that if one of the terms $P(E_i \mid H_k)$ 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 smoothing, 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.

$$Ac = \frac{TP}{TP + FN}$$  ... (4)

$$Ac = (Ac_{\text{Train}} \times 0.5 + Ac_{\text{Test}} \times 0.5)$$  ... (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 DISCUSSION

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, 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 DISCUSSION

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, 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 DISCUSSION

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, 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 DISCUSSION

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,
Input data:
\[ P_E, P_M, n, it, e \]
- Percentage of elitism and mutation
- Number of Individuals
- 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 \( A_c \) (phenotype)
   2.2. Move the amount \( (n \times P_E) \) of ordered individuals to the next generation
   2.3. Select randomly two individuals of the current generation
   2.4. Combine genotypes in order to obtain descendants
       
       \[ \text{If } (s_0 < 5000.0) \text{ and } (40\% < P_M < 80\%) \text{, add offspring to the next generation and count } n - (n \times P_E) \text{ individuals, else, go to step 2.3} \]
   2.5. Select randomly one individual of the next generation
   2.6. Genetically alter one gene of such individual
   2.7. If \( A_c > \theta \) or \( i < it \), \( i = i + 1 \), go to step 2

3. **Evaluation**
   3.1. Select the best individual (classifier) of the current generation
   3.2. If \( A_c > \theta \) 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

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

**Fig. 3.** Evolution of classifier accuracy over generations
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.

**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**

373-382, 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.


