AMERICAN-EURASIAN JOURNAL OF SUSTAINABLE AGRICULTURE
ISSN: 1995-0748, EISSN: 1998-1074
2016, volume(10), issue(6): pages (83-91)
Published Online 31 December 2016 in http://www.aensiweb.com/AEJSA/

Identification of pesticides in *Annona muricata* Using artificial intelligence (Pengenalan racun perosak dalam Annona muricata Menggunakan kecerdasan buatan)

[1]Maira Sánchez, [2]Olga Ramos, [3]Dario Amaya

[1]Research Assistant, Mechatronic Engineering, Nueva Granada Military University, Bogotá, Cundinamarca, Colombia.
[2,3]Professor and Research, Mechatronic Engineering, Nueva Granada Military University, Bogotá, Cundinamarca, Colombia

Received 18 September 2016; Accepted 28 December 2016

Address For Correspondence:
Maira Sanchez, Nueva Granada Military University, Mechatronic Engineering, Faculty of Engineering, 57+16500000 – Ext 1280, Bogotá Colombia.

Copyright © 2016 by authors and American-Eurasian Network for Scientific Information.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Open Access

BACKGROUND
The benefits of using pesticides in the agricultural production are multiple. However, the excessive and uncontrolled use of these has generated many problems, starting with environmental affection, damage to the human metabolism, among others.

OBJECTIVE
Classify and identify of ^{1}H and ^{13}C NMR spectra of Malathion, Chlorpyrifos, Diazinon and Dimethoate pesticides used in cultivation of *Annona muricata* through pattern recognition systems, using artificial intelligence techniques, specifically neural networks.

RESULTS
The identification rate was in average of 97.89%, maintaining a high value of 97% for all pesticides, 14 neurons were needed to correctly recognize pesticides, which implies a 46% reduction in the number of neurons in relation to the network trained with the ^{1}H NMR spectra data.

CONCLUSION
The implemented recognition algorithms are able to recognize the differences between shifts (ppm) and signals of ^{1}H and ^{13}C NMR spectra, managing effectively identify the corresponding pesticide.
For most pesticides, recognition was above 97%, except Diazinon where a value of recognition of 61.86% was obtained with the data from the ^{1}H NMR spectra. This value is low compared with other compounds.

KEY WORDS
Plaguicides, Annona muricata, NMR ^{13}C, NMR ^{1}H, neural networks.

INTRODUCTION

Pesticides are substances used to prevent, eliminate and control pests, whether plants or animals, including diseases of unwanted species that interfere with the production [1]. In Colombia, one of the pillars of its economy is the agriculture; therefore, the use of pesticides is a necessity in crops of fruits and vegetables mainly.

Unfortunately, the use of these products is not controlled or supervised by any state entity. For such motive, nowadays, studies or investigations that contribute knowledge about their interaction with the environment, humans, and existing ecosystems are of great interest.

In food, pesticides are found in small amounts or traces, where the analysis and detection is performed in the laboratory.

One of the used methods is nuclear magnetic spectroscopy (NMR), which facilitates the identification of these compounds found in the sample by analyzing the various signals and displacements present in the ^{1}H and ^{13}C NMR spectra, specific of every compound.

Nowadays, there are studies and programs, which predict the peaks and probable displacements that a molecule or compound in ^{1}H and ^{13}C NMR spectrum can have, but these elucidate the spectra from the drawn molecular structure. Until now, on the market there are few software or programs that perform the reverse process, that is to say, that from spectra they elucidate efficiently the compound from which they come, thus being expensive. This is due to the high computational cost required for the simulation of NMR spectra, because this value grows exponentially with the number of elements making up the molecule [2].

Since the NMR spectra is characteristic of a molecule, it can be employed along with systems pattern recognition, as is the case of Artificial Neural Networks (ANNs). These are mathematical models with geometry similar to the connection of the neurons in the brain [3] , which are defined as systems of interconnections, highly distributed with nonlinear adaptation (EPs), that work between themselves to produce a stimulus of exit [4]. These are trained by trial and error using different combinations for the connections, in order to recognize the relationships bet ween data items input and output.

There are studies, which by means of RNAs, achieved good rates of identification and rejection of mass spectra and FT-IR, reducing false negatives and positives. Among these, a work for identifying gasoline using near infrared (NIR) can be found, in which the effective classification of three types of fuel is obtained, using the method of probabilistic neural network (PNN) [5]. Another study, demonstrated the feasibility of analyzing the modes of action of four herbicides from the ^{1}H NMR spectrum analysis and subsequent classification using neural networks, opening up possibilities for the discovery of new herbicides, drugs or metabolic profile applications [6].

Other researches based on neural networks, show the ability of this method to process data obtained from FT-IR spectra, and identify and discriminate between genotypes of *Campylobacter* bacteria [7].

This paper presents the classification and identification of ^{1}H and ^{13}C NMR spectra of Malathion, Chlorpyrifos, Diazinon and Dimethoate pesticides used in cultivation of *Annona muricata* through pattern recognition systems, using artificial intelligence techniques, specifically neural networks. These networks take into account a set of all possible individual patterns to be recognized, all properly labeled. Based on the recognition of these patterns, the data and signals acquired from NMR spectra corresponding to each of the compounds are described and classified. For training and learning processes of the neural networks, the backpropagation algorithm was used, changing the way to calculate the error.

MATERIALS AND METHODS

Pesticides:

The pesticides were chosen according to the toxicological classification reported by the ICA [8]. According to this, two types of pesticides were evaluated: first are the organophosphorus, which are fat-soluble compounds whose general formula is derived from phosphoric acid and can be found in the environment as phosphates, phosphonates, phosphoramidothioate, phosphoramidates [9]. Second, are the organochlorine pesticides, characterized for being hydrocarbons high in chlorine atoms, and having a variable toxicity according to the degree of chlorination. Most of these products persist in the environment and affect wildlife, reason why they are prohibited or restricted in certain countries [10].

According to the above four pesticides were chosen given their toxicity. The first compound was malathion (2-[(dimethoxy phosphorothioyl) sulfanyl] butanedioate, Diethyl), is a moderately toxic insecticide [8], it penetrates rapidly into the body when making contact with the surface of the skin and eyes. It is a toxic product for birds, aquatic invertebrates, amphibians and highly toxic for bees. It is a cholinesterase inhibitor, a possible carcinogen and a suspected endocrine switch [11].

The second chosen pesticide was Chlorpyrifos (O,O-diethyl O-3,5,6-trichloropyridin-2-yl phosphorothioate), which is a slightly dangerous insecticide. Prolonged exposure to chlorpyrifos may cause effects on the adrenal glands, also show the same effects as those associated with cholinesterase inhibition. By inhalation, it may aggravate existing chronic problems such as asthma, emphysema or bronchitis [12].

The third compound was Diazinon (O,O-Diethyl O-[4-methyl-6-(propan-2-yl)pyrimidin-2-yl] phosphorothioate), which is extremely toxic to warm-blooded land animals (birds), highly toxic to livestock, bees, crustaceans, fish and aquatic plants (algae) [13].

Finally, the pesticide Dimethoate was chosen (O,O-dimethyl S-[2-(methylamino)-2-oxoethyl] dithiophosphate), classified as a moderately toxic pesticide, primarily affects birds, bees and fish. It is a cholinesterase inhibitor insecticide, possible carcinogen and reproductive toxin promoter [14].

NMR Spectra:

From the databases nmrdb.org Tool for spectroscopists [15] and Spectral Database for Organic Compounds SDBS [16] and MestReNova® (v. 10.0) computational tool, ^{1}H and ^{13}C NMR spectra of the compounds were obtained.

Based on the computational tool developed jointly between the Ecole Polytechnique Fédéral of Lausanne in Switzerland, National University and Valle University in Colombia, together with MestReNova® 10.0 software, the ^{1}H and ^{13}C NMR spectra are elucidated, based on the molecular structure of each of the analyzed compounds, which were taken from the SDBS database. In Figure 1, the molecular structure of the compounds is shown.

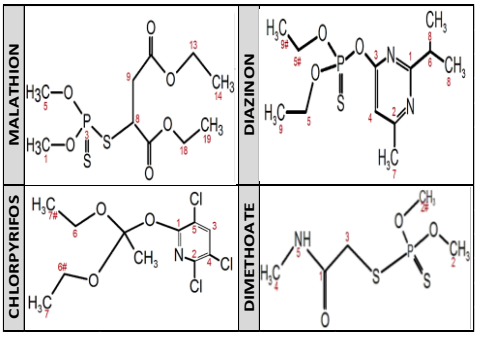
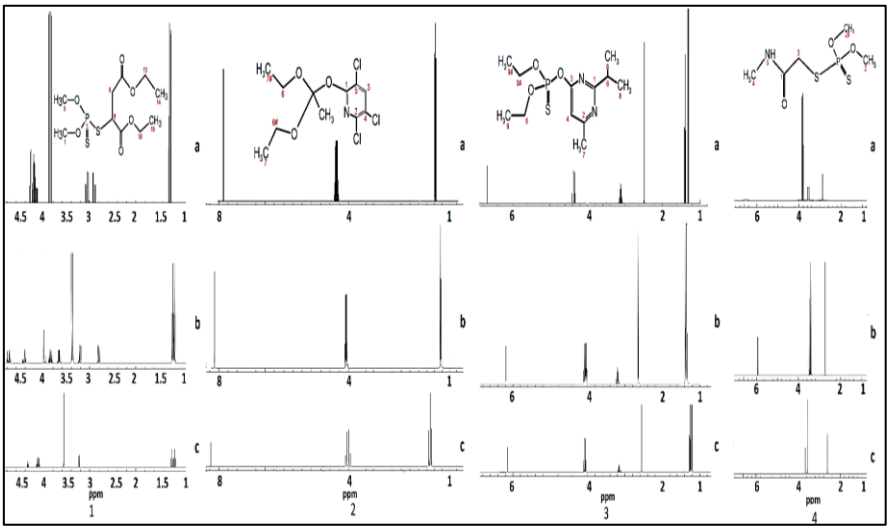


Fig. 1: Structures of pesticides used in the cultivation of *Annona muricata.*

At the same time, parameters such as frequency and solvent according to those reported in the SDBS database were adjusted. Therefore, for both types of analysis, deuterated chloroform ($CDCl_3$) was used as solvent, and frequencies of 399.65 MHz for ^{1}H NMR and 100.4 MHz for ^{13}C NMR where used.

Figure 2 shows the NMR spectra obtained for the pesticides in study, with which the identification algorithm was developed.

A



B

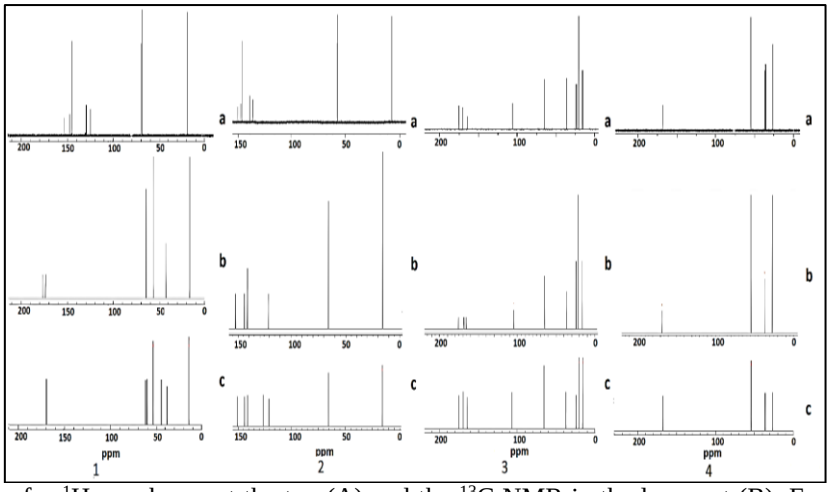


Fig. 1: NMR spectra for ^{1}H are shown at the top (A) and the ^{13}C NMR in the low part (B). For both cases, we have 1-Malathion, 2-Chlorpyrifos, 3-Diazinon, and 4-Dimethoate. The databases correspond to: a) SDBS, b) MestReNova® c) nmrdb.

In relation to the Figure 2, the ^{1}H NMR spectra for the pesticide Malathion (Figure 2-A, Spectrum 1) presents different displacements for Hydrogens 13, 13*, 18* and 18, from R-CH_2-X (X= N, O, Halogen). For chlorpyrifos and diazinon (Figure 2-A, Spectrum 2, 3), no difference occurs in the spectra, while for Dimethoate (Figure 2-A, Spectrum 4), the MestReNova® database shows a signal at 5,901 ppm corresponding to Hydrogen 5 (NH) labile proton of the molecule. The reason that signal does not appear in the other databases is because they excluded labile protons.

On the other hand, for the ^{13}C NMR spectra of every molecule (Figure 2-B), no displacement difference can be distinguish between the three databases.

Pattern recognition:

Pattern recognition is used for describing and classifying data from objects, people, performances or as in this case signals. Systems for recognizing patterns implementing neural networks work based on a set with all possible individual patterns to recognize, properly labeled. There are two models of RNAs suitable for the classification and identification of patterns, Multilayer Perceptron (MLP) and probabilistic neural network (PNN). In this paper, the classification algorithm proposed is described below:

a) Architecture:

The chosen network is a *Feed-Forward* type Multilayer Perceptron network model, which consists of a hidden layer and an output layer. The activation function for the hidden layer is a sigmoid function, represented by Equation 1:

$$\boldsymbol{f(x) = \frac{1}{1+e^{-x}}} \tag{1}$$

The activation function for the output layer is called softmax, represented by Equation 2.

$$\boldsymbol{g(x) = \frac{e^{x}}{\sum_{i=1}^{n} e^{x}}} \tag{2}$$

The network topology is shown in Figure 3, were the functions previously mentioned were selected due to the advantages they give for training and for classification.

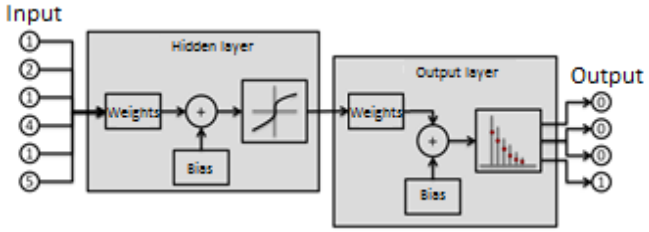


Fig. 2: Implemented Network Architecture

b) Training:

For the training and learning processes, a *backpropagation* algorithm is implemented, which updates the values of the weights of neurons, according to the present output error [17]. Since the required neural network is used for classification and identification processes, and not to perform regressions, the error calculation must be performed using the method of crossed entropy error.

The calculation of cross entropy uses the related values with the probability of correctly identifying patterns values, which adjusts the training process.

$$\boldsymbol{H(p,q) = -\sum_{n=1}^{N} p_n \log(q_n)} \quad (3)$$

Where:

p_n= Expected probability of recognition
q_n= Obtained probability of recognition
N= number of samples

The equation can be rewritten as follows in Equation 4.

$$\boldsymbol{L(w) = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log(\hat{y}_n) + (1 - y_n)\log(1 - \hat{y}_n)]} \quad (4)$$

Where:
y_n= Desired Output
$\hat{y}_n$= Obtained Output

The training data for the neural networks were the ^{1}H and ^{13}C NMR spectra obtained from the *MestReNova*® software and the computational tool *nmrdb.org*, through image processing techniques, which in addition to extracting the characteristics, can be used to solve problems related to scale and noise present in each of them. Finally, spectra from the *SDBS* database was extracted and used for the evaluation of the two networks.

RESULTS AND DISCUSSION

According to the results presented, it was found that classifiers recognize the differences between the shifts (ppm) or signals from ^{1}H and ^{13}C NMR spectra, therefore, identifies the pesticide from which these signals came. In relation to previous, the figures below show the behavior of the identification algorithms based on artificial intelligence in terms of the number of neurons in the hidden layer.

For the neural network in charge of identifying compounds through the ^{1}H NMR spectrum, a high value of classification was obtained from four neurons, as shown in Figure 4. However, it does not guarantee that the identification of the four compounds is being performed properly.

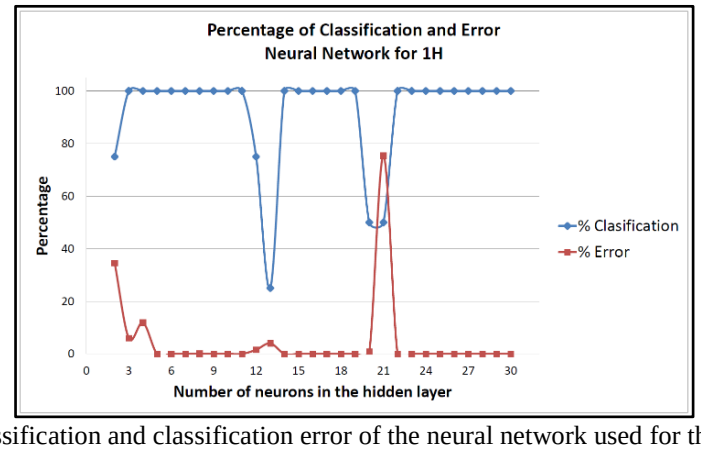


Fig. 4: Percentage of classification and classification error of the neural network used for the ^{1}H NMR.

In Figure 5, the percentages of individual identification for each compound according to the number of neurons used in the hidden layer of the network are presented. Showing that there is no link between the correct identification of the four spectra and the classification error. It can be seen that for some amounts of neurons with minimal error according to the data of Figure 3, i.e. four neurons, the identification of each compound is not the best.

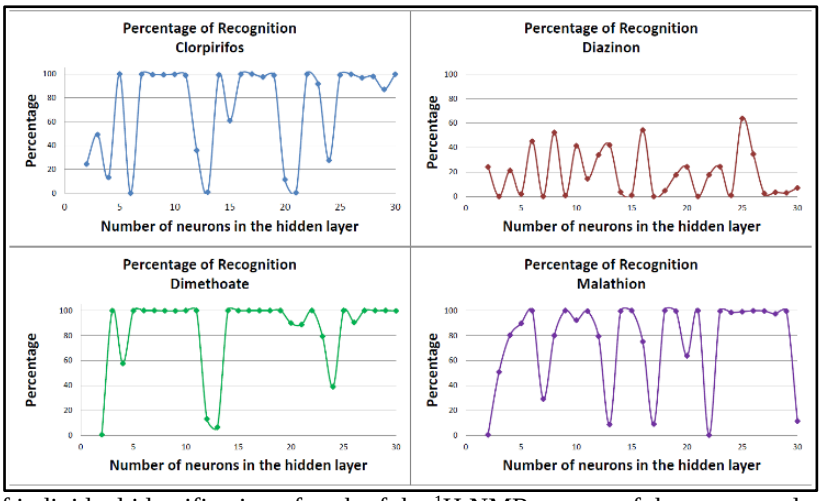


Fig. 5: Percentage of individual identification of each of the ^{1}H NMR spectra of the compounds studied in terms of the number of neurons in the hidden layer.

Considering the data in Figure 5, the appropriate number of neurons in the hidden layer for the trained network with data obtained from ^{1}H NMR spectra is 25. With this quantity of neurons, a percentage of average identification of 88.44 % was achieved, being the Diazinon the compound with the lowest index of recognition.

In Figure 6, the data related to the neural network training used to identify pesticides using ^{13}C NMR spectra are presented, an like the neural network for ^{1}H NMR had minimal errors of classification with few neurons in the hidden layer, specifically for three neurons.

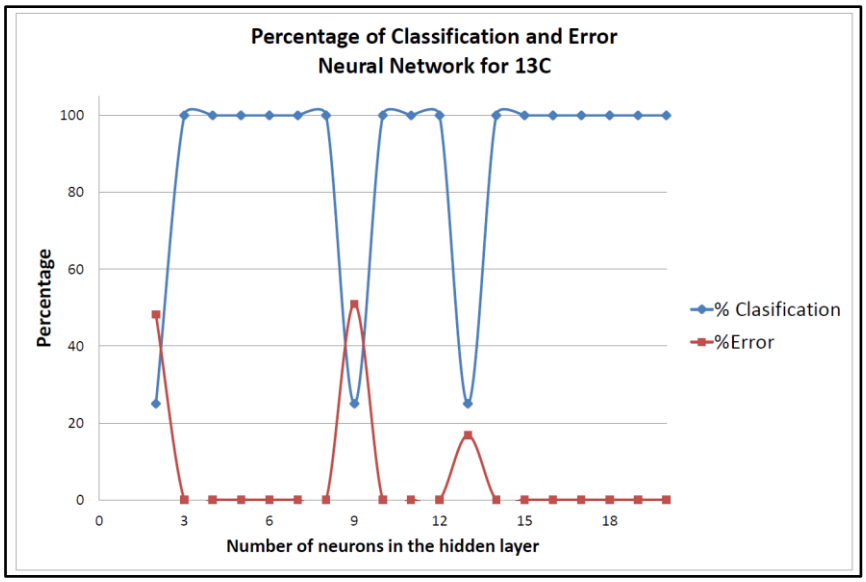


Fig. 6: Percentage of classification and classification error for the ^{13}C NMR neural network.

Even though the values of classification are initially high, the same does not happen with the identification of all compounds. In Figure 7, the recognition of pesticides through the ^{13}C NMR spectrum presented a better performance by having 14 to 18 neurons in the hidden layer. By adding more, like 19 and 20 neurons, Chlorpyrifos and Dimethoate compounds were not correctly identified.

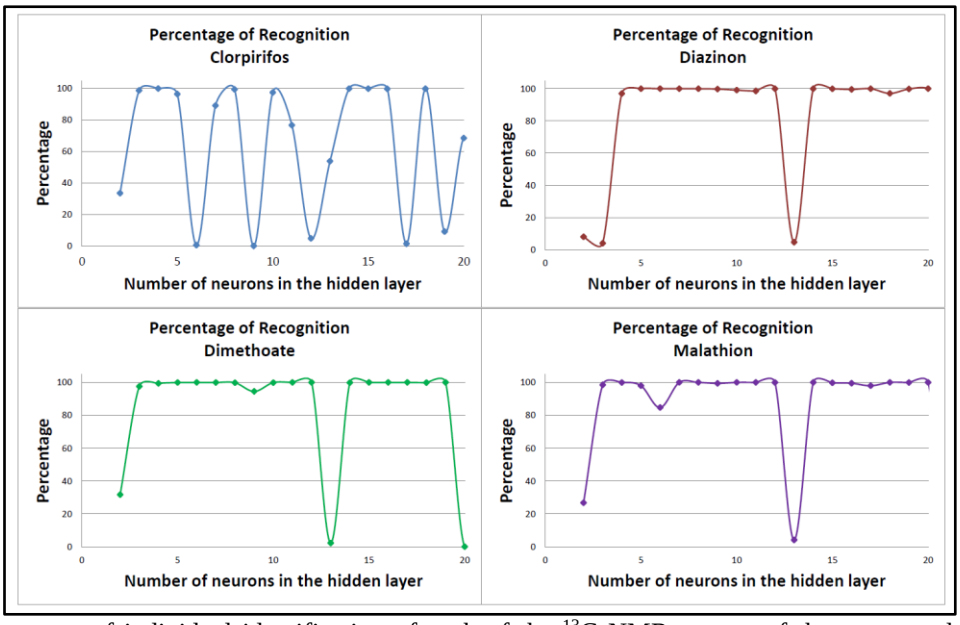


Fig. 7: Percentage of individual identification of each of the ^{13}C NMR spectra of the compounds studied in function of the number of neurons in the hidden layer.

According to the results shown in Figure 7, 14 neurons were needed to correctly recognize pesticides, which implies a 46% reduction in the number of neurons in relation to the network trained with the ^{1}H NMR spectra data. The identification rate was in average of 97.89%, maintaining a high value of 97% for all pesticides.

In Figure 8, the results of the programmed algorithm for identifying pesticides by means of the ^{1}H NMR spectra are shown. The algorithm receives the spectra as an image and the neural network discriminates from which compound it is, returning the result with the respective index of identification. In this case, the identification of the Malathion compound.

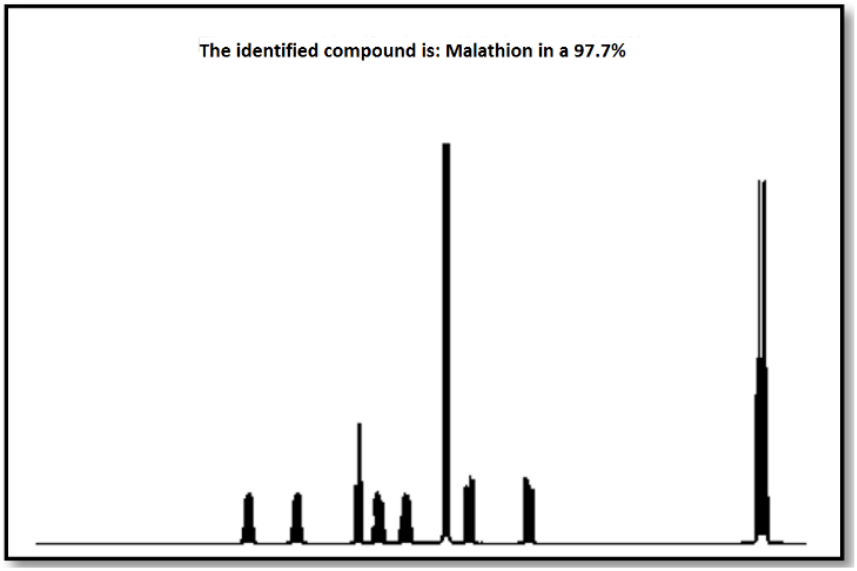


Fig. 8: Identification of one of the 1H NMR spectra corresponding to Malathion.

Similarly, in Figure 9, the result for Malathion is shown, using the neural network trained with data obtained from the ^{13}C NMR spectra.

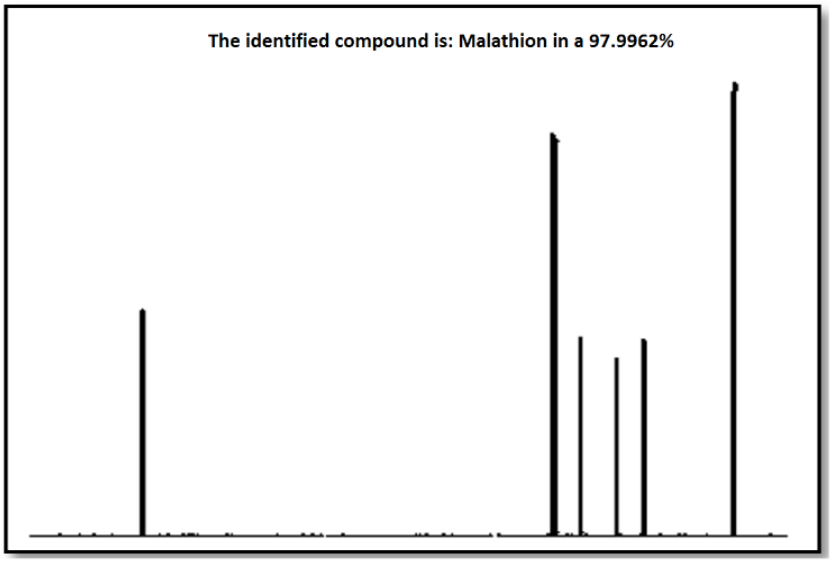


Fig. 9: Identification of one of the 13C NMR spectra corresponding to Malathion.

Finally, in Table 1 the values of recognition for the four pesticides according to the type of spectrum used in the identification process are shown.

Table 1: Percentage of success in the identification of each of the compounds through their respective 1H and ^{13}C NMR spectra.

Identification	Chlorpyrifos	Diazinon	Dimethoate	Malathion
RMN ^{1}H	97.24	61.86	97.60	97.7
RMN ^{13}C	97.84	97.98	97.77	97.99

Conclusions:

The implemented recognition algorithms are able to recognize the differences between shifts (ppm) and signals of ^{1}H and ^{13}C NMR spectra, managing effectively identify the corresponding pesticide.

Not always a correct classification of data by the neural network ensures good performance in identifying patterns, i.e. while training with 7 neurons for ^{1}H NMR and 12 neurons for ^{13}C NMR, low value misclassification with failed recognition for Malathion and Diazinon in the first case and Chlorpyrifos in the second were obtained.

For most pesticides, recognition was above 97%, except Diazinon where a value of recognition of 61.86% was obtained with the data from the ^{1}H NMR spectra. This value is low compared with other compounds.

There is no direct relationship between the number of neurons needed to process information from the ^{1}H and ^{13}C NMR spectra, because in order to identify compounds for ^{1}H NMR 25 neurons were required, while for ^{13}C only 14 neurons were needed.

ACKNOWLEDGMENTS

The authors express their gratitude to the Research Vice-chancellorship of the Military University Nueva Granada, for financing the IMP-ING-1777 project, 2015.

REFERENCES

[1] Organización de las Naciones Unidas Para la Agricultura (FAO). 2002. Código Internacional de Conducta para la Distribución y Utilización de Plaguicidas, [On line]. Available: http://www.fao.org/3/a-a0220s.pdf.

[2] Castillo, A.M., L. Patiny, J. Wist, 2011. Fast and accurate algorithm for the simulation of NMR spectra of large spin systems, *Journal of Magnetic Resonance*, 209(nº 2): 123-130.

[3] Gómez, J.P., 2010. Prevalencia, Caracterización Quimio taxonómica Mediante Espectroscopía de Infrarrojos e Identificación Con Redes Neuronales Artificiales y Análisis Multivariante de Especies Termofílicas de Campylobacter Aisladas de Aves y Productos Avícolas, *Tesis Doctoral*.

[4] Cardozo Vásque, A., 2012. Desarrollo de un modelo de red neuronal artificial para la reducción de escala (downscaling) de datos de temperatura del modelo Climático Global Canadiense 3.1 a estaciones meteorológicas colombianas. *Tesis de Maestria.*

[5] Balabin, R.M., R.Z. Safieva, E.I. Lomakina, 2010. Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques, *Analytica Chimica Acta*, 671(nº 1-2): 27-35.
[6] Ott, K.-H., N. Arani, B. Singh, G.W., 2003. Metabonómica clasifica vías afectadas por compuestos bioactivos. Artificial clasificación de la red neuronal de los espectros de RMN de extractos de plantas, *Fitoquímica*, 62(nº 6): 971-985.
[7] Mouwen, D., R. Capita, C. Alonso-Calleja, J. Prieto-Gómez, M. Prieto, 2006. Artificial neural network based identification of Campylobacter species by Fourier transform infrared spectroscopy. *Journal of Microbiological Methods*, 67(nº 1): 131-140.
[8] www.huila.gov.co. 2007. Informe de Evaluación de Plaguicidas y Plan de Acción para su Uso Más Seguro (PERSUAP),» Julio 2007. [On line]. Available: http://www.huila.gov.co/documentos/I/Informe%20de%20Evaluacion%20de%20Plaguicidas.pdf. [Last login: 28 04 2015].
[9] Ferrer, A., 2003. Intoxicación por plaguicidas, *Anales Sis San Navarra*, 26.
[10] Menéndez Díez, F., 2009. Higiene industrial. Manual Para la Formación del Especialista, 9 ed., Lex Nova.
[11] Proficol Andina B.V Sucursal Colombia, 2010. Hoja de Datos de Seguridad de los Productos-Malathion 57 EC,» Julio 2010. [On line]. Available: http://www.adama.com/colombia/es/Images/HS%20Malathion%2057%20EC_tcm104-57103.pdf.
[12] Nufarm, S.A., 2011. «Clorpyrifos Agrogen 480 EC Hoja de Datos de Seguridad de Materiales HDSM,» 15 04 2011. [On line]. Available: http://www.nufarm.com/assets/17093/1/CLORPYRIFOSAGROGEN480EC.pdf.
[13] ANALJASA Agroquímicos, «Ficha técnica,» [On line]. Available: http://www.anajalsa.com.mx/pdf/ficha_21.pdf.
[14] Nufarm, S.A., 2015. «Hoja de Datos de Seguridad- Dimetoato 40% CE,» [En línea]. Available: http://www.nufarm.com/assets/22303/1/DIMETOATO40NufarmHojadeSeguridad_MSDS_.pdf.
[15] Ecole Polytechnique Fédérale de Lausanne (EPFL), 2015. Universidad del Valle, Universidad Nacional de Colombia - Bogotá, Universidad Industrial de Santander , «nmrdb.org Tools for NMR spectroscopists,» [On line]. Available: http://www.nmrdb.org/predictor_all/index.shtml?v=v2.17.5.
[16] National Institute of Advanced Industrial Science and Technology (AIST), 2015. «Spectral Database for Organic Compounds, SDBS,» [On line]. Available: http://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_frame_disp.cgi?sdbsno=45534.
[17] Rodríguez Fernández, A.M., 2010. Astrofísica Computacional: Aplicación De Técnicas De Inteligencia Artificial En La Clasificación Y Parametrización De Espectros Estelares, *Tesis de doctorado*, Coruña.