

ORIGINAL

An artificial intelligence-based approach for an urgent detection of the pesticide responsible of intoxication

Un enfoque basado en la inteligencia artificial para la detección urgente del plaguicida responsable de la intoxicación

Rajae Ghanimi¹ , Fadoua Ghanimi¹ , Ilyas Ghanimi¹ , Abdelmajid Soulaymani¹ 

¹Ibn Tofail University, Av. de L'Université, Kénitra-Morocco.

Cite as: Ghanimi R, Ghanimi F, Ghanimi I, Soulaymani A. An artificial intelligence-based approach for an urgent detection of the pesticide responsible of intoxication. Data and Metadata. 2023;2:114. <https://doi.org/10.56294/dm2023114>


Submitted: 22-09-2023

Revised: 03-11-2023

Accepted: 28-12-2023

Published: 29-12-2023

Editor: Prof. Dr. Javier González Argote 

Guest Editor: Yousef Farhaoui 

Note: Paper presented at the International Conference on Artificial Intelligence and Smart Environments (ICAISE'2023).

ABSTRACT

Acute poisoning by pesticides in Morocco is an important public health issue, because the use of pesticides has become both massive and anarchic. This is the cause of deaths whose incidence is unfortunately increasing. Unfortunately, these deaths are not always accidental. Pesticides are also used as a means of suicide; according to the WHO, these are means suicide chemicals most used in the world, since, out of the 800 000 suicides recorded per year, more than a third are caused by this type of product. Even more serious, these suicides are currently being observed among children and teenagers. Faced with this alarming figure, and in order to prevent deaths and improve emergency treatment of cases of pesticide poisoning, it becomes important to use the potential of artificial intelligence in the treatment of these admissions. Our approach is essentially based on machine learning algorithms, including decision support software capable of predicting, based on major clinical signs, the most likely pesticide responsible of the intoxication in the triage room. This, before moving on to the confirmation stage based on biological and toxicological investigations, which are often costly and time-consuming.

Keywords: Pesticide; Machine Learning; Artificial Intelligence; Diagnostics; Emergencies; Triage Room.

RESUMEN

La intoxicación aguda por plaguicidas en Marruecos es un importante problema de salud pública, ya que el uso de plaguicidas se ha vuelto masivo y anárquico. Esta es la causa de muertes cuya incidencia desgraciadamente va en aumento. Por desgracia, estas muertes no siempre son accidentales. Los pesticidas también se utilizan como medio de suicidio; según la OMS, son los productos químicos suicidas más utilizados en el mundo, ya que, de los 800 000 suicidios registrados al año, más de un tercio están causados por este tipo de productos. Y lo que es más grave, estos suicidios se observan actualmente entre niños y adolescentes. Ante esta alarmante cifra, y con el fin de prevenir muertes y mejorar el tratamiento de urgencia de los casos de intoxicación por plaguicidas, se hace importante utilizar el potencial de la inteligencia artificial en el tratamiento de estos ingresos. Nuestro enfoque se basa esencialmente en algoritmos de aprendizaje automático, incluyendo un software de apoyo a la toma de decisiones capaz de predecir, en función de los principales signos clínicos, el plaguicida más probable responsable de la intoxicación en la sala de triaje. Todo ello, antes de pasar a la fase de confirmación basada en investigaciones biológicas y toxicológicas, que a menudo son costosas y requieren mucho tiempo.

Palabras clave: Pesticida; Aprendizaje Automático; Inteligencia Artificial; Diagnóstico; Emergencias; Sala de Triage.

INTRODUCTION

Poisoning is a public health problem worldwide and is one of the most common reasons for attending hospital emergency departments.⁽¹⁾ Although the incidence of poisoning is difficult to estimate precisely, the wide availability and accessibility of chemicals and their widespread use in various applications, including medicine, agriculture and industry, have increased the risk of poisoning. Acute toxicity of pesticides results from misuse, accidental use of pesticides (domestic accidents) or intentional poisoning, often extremely serious. Organophosphate pesticides and Carbamates cause the most common cases of pesticide poisoning.⁽²⁾ Exposure is mainly via the skin-mucosal and respiratory routes (inhalation), the oral route of exposure would concern the general population more through accidental or intentional ingestion of pesticides. According to the World Health Organization (WHO), there are one million serious pesticide poisonings worldwide each year, causing around 220 000 deaths per year. Early diagnosis of these cases of poisoning is decisive in terms of prognosis.⁽³⁾ Our work consists of a Machine Learning (ML) algorithm capable of predicting the type of pesticide responsible for the poisoning based on the clinical signs of the patient in the emergency room. to develop our algorithm. We worked on the database of the National Anti-Poison Center of MOROCCO, bringing together 8,107 cases of pesticide poisoning from March 1980 to December 2014, in order to extract the data necessary for training an algorithm. Our patients come from both urban and rural areas of our data are composed of different ages and sexes. These etiologies were detected through family history and interview, clinical symptoms, and laboratory examination of gastric cleaning products.

Chavan et al.⁽⁴⁾ presented a classification model using the k-nearest neighbor (k-NN) algorithm. Their model focused on the analysis of 118 chemicals from the NEDO (New Energy and Industrial Technology Development Organization) RTD database, now known as the Hazard Evaluation Support System (HESS). This model relied on the two categories of acute toxicity (LD50) as variables to predict, and on eight fingerprints from the PaDEL as predictive variables. The results demonstrated the effectiveness of the model in predicting LD50 categories for 70 of the 94 chemicals in the training set, and for 19 of the 24 chemicals in the testing set.⁽⁴⁾

Other studies, such as those of JC Carvaillo et al.⁽⁵⁾ introduced the AOP-help Finder program. This program was developed to predict the toxicity of products by analyzing a multitude of data from the scientific literature. To do this, two distinct approaches were used: the first involved textual analysis targeting relevant terms, including chemical names (e.g., bisphenol S, pesticides) and descriptions of disease biological processes. The second approach focused on identifying critical keywords. Researchers have developed directories containing the names of known substances (for example, bisphenol S from PubChem) and numerous Adverse Outcome Pathways (AOP).

In the context of paraquat poisoning, Huiling Chen et al.⁽⁶⁾ explored the use of common blood indices to diagnose PQ toxicity and assess its severity using a machine learning approach. They employed a support vector machine-based method along with feature selection to predict the risk of PQ poisoning. This approach was validated on a sample of 79 individuals, making it possible to differentiate living from deceased patients, and was rigorously evaluated for its accuracy, sensitivity and specificity.

The team of Huiling Chen et al.⁽⁷⁾ focused on early identification of paraquat poisoning (PQ) in patients to ensure rapid and accurate prognosis. Their study demonstrated that routine blood tests can be used to preliminarily establish the toxicity or prognosis of PQ poisoning, without requiring PQ concentration data. This approach represents an additional and innovative tool to assess the prognosis of PQ poisoning.

Despite the abundance of studies on technology and artificial intelligence in the fields of healthcare, diagnostics and pharmacology, researchers have observed a lack of publications specifically focused on the use of artificial intelligence to diagnose cases of poisoning. This identified gap arouses their interest for future research in this innovative field.

METHODS

The approach is essentially based on Machine Learning (ML) algorithms, capable of predicting, based on major clinical signs, the most probable pesticide in the triage room. Hopefully, this prediction by the software has to occur before moving on to the confirmation stage based on biological and toxicological investigations, which are oftencostly and time-consuming. Based on the patient's clinical signs reported by the doctor in charge in emergency room, the work consisted of predicting the type of pesticide responsible for the poisoning.

Database description

The approach is essentially based on Machine Learning (ML) algorithms, capable of predicting, based on major clinical signs, the most probable pesticide in the triage room. Hopefully, this prediction by the software has to occur before moving on to the confirmation stage based on biological and toxicological investigations, which are often costly and time-consuming. Based on the patient's clinical signs reported by the doctor in charge in emergency room, the work consist to predicting the type of pesticide responsible for the poisoning.

The data used is collected from the national poison control center of Morocco. The Database included

8107 cases of pesticide poisoning from March 1980 to December 2014. Our patients come from both urban and rural areas of our data are composed of different ages: 1 newborn (maternal-fetal poisoning); 12 trotters, 417 children, 2473 adolescents; 5180 adults and 23 elderly people.

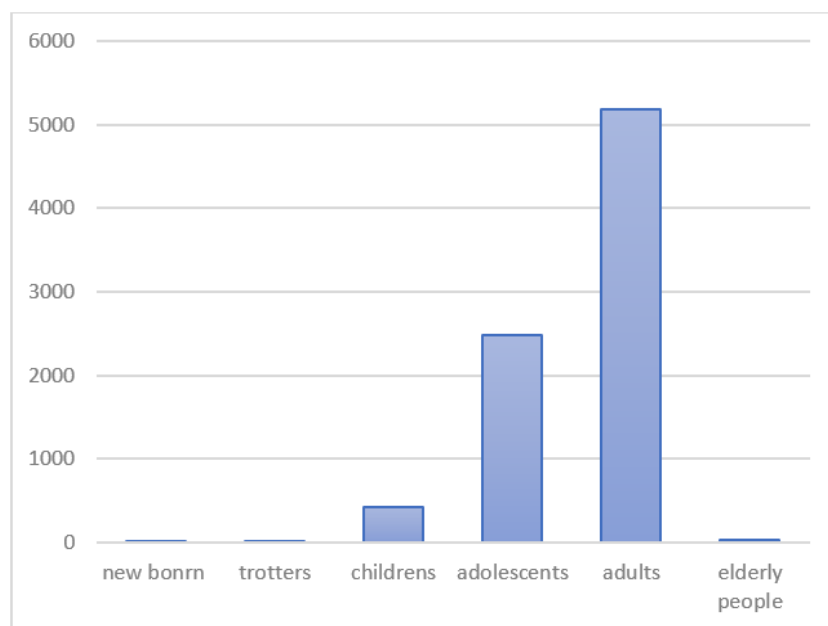


Figure 1. Distribution of data by age range

The distribution by sex shows a female predominance, 5150 women (i.e. 64 %) compared to 2957 men (i.e. 36 %). And the predominance of the suicidal cause (98 %), other causes (drug addiction, professional, criminal and abortion) represent less than 2 %. Pesticides represent the second cause of poisoning in our database after drugs.

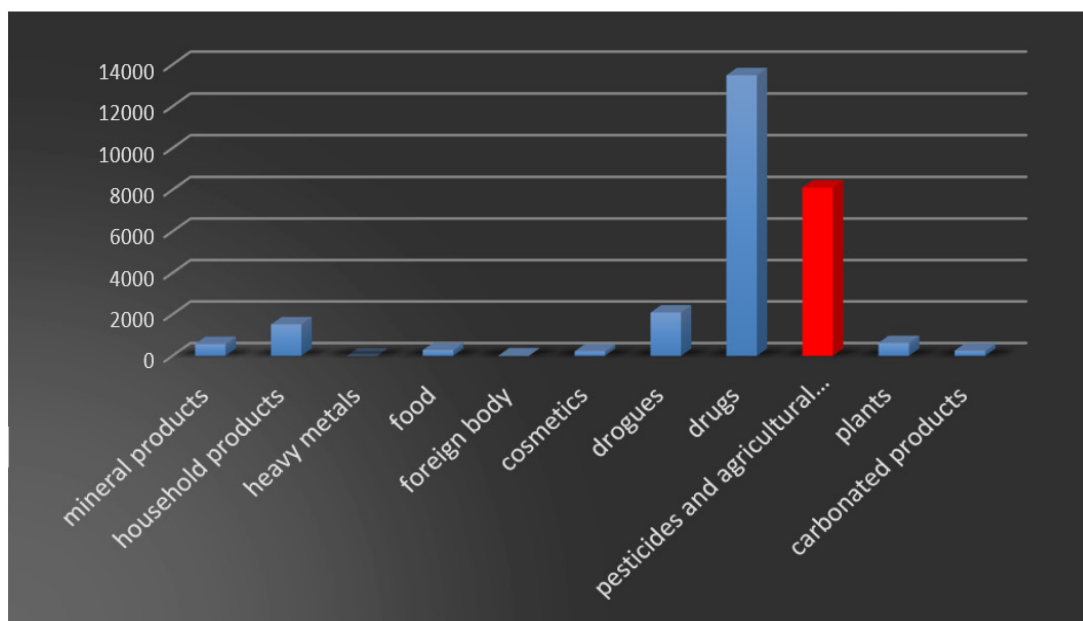


Figure 2. Database by category of toxic substance

Four different ML models were tested in this study for prediction: SVM, Decision Tree, Random Forest and XGBoost. GridSearchCV function from scikit-learn library was used to perform the search for model's parameters. The evaluation metrics used to evaluate the models were precision, recall and accuracy.

Support Vector Machine

Support Vector Machine (known as SVM) is a supervised learning method designed for binary classification "either positive or negative class". It is applied for the purpose of finding patterns from the collection of data.

⁽⁸⁾ Generally, the pattern classification applies an activity to be involved in two main steps; the first is mapping the input to higher dimension feature space, this is done due to the fact of the SVM that usually depends on geometrics characteristics of inputted data, and, the second is finding the most suitable hyper plane that classifies the mapped features in the higher dimensional space. It's beneficial for classification and regression approaches. The margin is the space between the hyperplane and the nearby data point. The core target is discovering the hyperplane with the greatest database split; into two classes for gaining novel vectors with proper classification.

Classification with the SVM is based on the determination of a hyperplane which separates two classes. The hyperplane is defined based on the maximum margin. The maximum margin is distance between the hyperplane and the closest samples of each class. However, in order not to be limited to linearly separable data, it is possible to use a kernel in order to separate non-linearly separable data. This involves transforming the representation of the data in a larger space where they will be linearly separable. ^(9,10,11,12,13)

To do this, we use the scalar product of our data with the space of larger dimensions. Nonlinear kernels make it possible to match this space of higher dimensions. Among the types of kernels commonly used, we find the radial basis function, the polynomial function, and the Gaussian function.

Decision Tree

The decision tree is based on decision rules in each of the nodes. These rules of decisions are based on feature attributes. The leaves of the tree match to the results of successive decisions of the model. ⁽¹⁴⁾ The depth of the tree determines its capacity: the greater the depth, the greater the capacity, but the greater the risk of overfitting is big. Hyperparameters, such as the minimum number of data points to split a node and the minimum number of data per leaf, control the depth of the tree.

When constructing the tree, the decision rule for each node is chosen based on of the minimization of the entropy or the impurity of the branches which will result. Thus, the characteristics which make it possible to better distinguish the classes are chosen in the first decision rules of the tree. ^(15,16,17,18)

Random forest

Random Forest is a supervised learning model used for classification and regression. The Random Forest consists of several decisional trees to classify a new object from an input vector while the random forest's input vector is the input of every tree. ⁽⁹⁾ The final decision is taken by considering the average of all trees. However, in the decision tree, the decision rule at each node is based on the best features that give the most homogeneous splitting in each of the divisions. In the decision forest, the division is chosen on the best features of a subset of features chosen randomly. So an individual tree in the forest of trees may be less efficient than the decision tree, but the average of all the trees in the forest makes it possible to reduce the overall error and generally have better results. ⁽⁹⁾ The scientist proves that this method achieves accurate and stable prediction with high performance. ^(19,20,21,22)

XGBoost

XGBoost is a decision tree ensemble based on gradient boosting designed to be highly scalable. ⁽¹⁰⁾ XGboost uses randomization techniques to reduce overfitting and to increment training speed. In addition, XGBoost implements several methods to increment the training speed of decision trees not directly related to ensemble accuracy. Specifically, XGBoost focuses on reducing the computational complexity for finding the best split, which is the most time-consuming part of decision tree construction algorithms. ⁽²³⁾

The proposed method

We evaluated the performance of four different supervised machine learning algorithms. The different machine learning techniques will only be summarized here but have been described in more detail above. In short, machine learning corresponds to the use of algorithms which learn non-linear associations in a way retrospectively using data to estimate the risk of a predefined outcome.

To this is added the bagging technique, corresponding to the division of this data by "groups" or "bags".

The algorithms keep part of the data aside in order to then test its reliability. There are many internal validation techniques of the algorithm. Here we used a validation technique crossed, with a validation sample of 1 to 5. This consists of dividing the data into 5 samples, then selects one as validation while the other 4 samples participate in learning, then the operation is repeated by selecting another sample of validation. This allows for a more robust estimate.

We selected the variables to integrate into the algorithms among those which had a result significant after the statistical analysis, while trying to integrate a limited number of variables into each algorithm. This was intended to avoid the creation of an overly complex model, which could be subject to the phenomenon of overfitting and cannot subsequently be generalized. Overfitting corresponds to the development of an algorithm that is too precise, too close to the source population, and which is not reliable when tested on a new series of

data and is therefore not generalizable to the target population.^(24,25,26)

RESULTS

To analysis of the result in this experiment, we considered different analyses to examine the four machine learning models for the prediction of the nature of poisoning pesticides applied to the dataset. In terms of accuracy, XGBoost achieved the highest accuracy of 75 % and Decision tree achieved the worst performance 52 %. With respect to F1-score, XGBoost achieved the highest score 84 %, Random Forest achieved 83 %, SVM performs 74 %. and Decision tree 52 % (see table 1)

According to these results we conclude that XGBoost Algorithm is more effective than the other models for predicting the nature of the pesticide responsible of poisoning case. Especially since XGBoost has demonstrated in several studies its performance in the field of medical diagnostics.^(11,12,13)

Table 1. Results for the three machine learning algorithms with the test dataset				
	Accuracy score	Precision score	Recall score	F1-score
XGBoost	0,75	0,91	0,78	0,84
SVM	0,62	0,72	0,76	0,74
Decision tree	0,52	0,50	0,57	0,52
Random forest	0,74	0,85	0,81	0,83

CONCLUSION

It is recognized that it is difficult to predict the nature of the pesticide responsible of poisoning in time due to several risk factors. Due to these factors, scientists have turned to modern approaches such machine learning to predict disease.

Three machine learning models, namely respectively: XGBoost, Random Forest and SVM were developed in this work. XGBoost predicted the best with a performance of 84 % in the F1-score used in this study. As a result, the objectives set at the introduction were achieved and this research responded to the expectations of emergency clinicians.

REFERENCES

1. Mégarbane B. Présentation clinique des principales , intoxications et approche par les toxi dromes. Réanimation 2012 ; 21 : S482-93.
2. El-Sarnagawy, G.N., Abdelnoor, A.A., Abuelfadl, A.A. et al. Comparison between variousscoring systems in predicting the need for intensive care unit admission of acute pesticide-poisoned patients. Environ SciPollut Res29,3399934009(2022).
3. Thabet, H., Brahmi, N., Elghord, H., Kouraichi, N., Amamou, M. (2013). Intoxications par les insecticides organophosphorés et carbamates. In: Intoxications aiguës. Références en ré animation. Collection de la SRLF. Springer, Paris.
4. Chavan, S.; Friedman, R.; Nicholls, I.A. Acute Toxicity-Supported Chronic Toxicity Prediction: A k-Nearest Neighbor CoupledRead-Across Strategy. Int. J. Mol. Sci. 2015, 16, 11659-11677.
5. Jean-Charles Carvaillo, Robert Barouki, Xavier Coumoul, and Karine Audouze2019 Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach ;Environmental Health Perspectives 127:4 CID: 047005.
6. Huiling Chen, Lufeng Hu , Huaizhong Li , Guangliang Hong , Tao Zhang, Jianshe Ma, Zhongqiu Lu : An Effective Machine LearningApproach for Prognosis of Paraquat Poisoning Patients Using Blood Routine Indexes; Basic &Clinical Pharmacology &ToxicologyVolume120, Issue1January 2017-Pages 86-96
7. Lufeng Hu, Guangliang Hong, Jianshe Ma, Xianqin Wang, Huiling Chen,An efficient machine learning approach for diagnosis of paraquat-poisoned patients,Computers in Biology and Medicine,Volume 59,2015,Pages 116-124,
8. V.N.Vapnik,The Nature of Statistical Learning Theory,Springer-Verlag, New York. (1995).

9. L. Breiman. Machine Learning 45 (1): 5-32 (2001).
10. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pages 785-794, New York, NY, USA, (2016).
11. Sami, S.M., Bhuiyan, M.I.H. (2022). Power Transformer Fault Diagnosis with Intrinsic Time-Scale Decomposition and XGBoost Classifier. In: Arefin, M.S., Kaiser, M.S., Bandyopadhyay, A., Ahad, M.A.R., Ray, K. (eds) Proceedings of the International Conference on Big Data, IoT, and Machine Learning. Lecture Notes on Data Engineering and Communications Technologies, vol 95. Springer, Singapore.
12. Jiang, YQ., Cao, SE., Cao, S. et al. Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. J Cancer Res Clin Oncol 147, 821-833 (2021).
13. Li, Q., Yang, H., Wang, P. et al. XGBoost-based and tumor-immune characterized gene signature for the prediction of metastatic status in breast cancer. J Transl Med 20, 177 (2022).
14. S.R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics, 21(3) :660-674, 1991.
15. Auza-Santiv   ez JC, D   az JAC, Cruz OAV, Robles-Nina SM, Escalante CS, Huanca BA. mHealth in health systems: barriers to implementation. Health Leadership and Quality of Life 2022;1:7-7. <https://doi.org/10.56294/hl20227>
16. Murillo-Ticona TA, Berneso-Soto ML. Los Entornos Virtuales de Aprendizaje al rescate del servicio educativo. Sincretismo 2020;1
17. Uman JMM, Arias LVC, Romero-Carazas R. Factores que dificultan la graduaci  n: El caso de la carrera profesional de contabilidad en las universidades peruanas. Revista Cient  fica Empresarial Debe-Haber 2023;1:58-74.
18. Gonzalez-Argote J. Analyzing the Trends and Impact of Health Policy Research: A Bibliometric Study. Health Leadership and Quality of Life 2023;2:28-28. <https://doi.org/10.56294/hl202328>
19. Coa YMF, Crisostomo NWF, D   az-Barriga GE. Desarrollo econ  mico sostenible bajo un r  gimen social sin preceptos   ticos y morales: auditor  a forense en contraposici  n de la corrupci  n. Revista Cient  fica Empresarial Debe-Haber 2023;1:48-62
20. Alaoui, S.S., and all. "Hate Speech Detection Using Text Mining and Machine Learning", International Journal of Decision Support System Technology, 2022, 14(1), 80. DOI: 10.4018/IJDSST.286680
21. Alaoui, S.S., and all. , "Data openness for efficient e-governance in the age of big data", International Journal of Cloud Computing, 2021, 10(5-6), pp. 522-532, <https://doi.org/10.1504/IJCC.2021.120391>
22. El Mouatasim, A., and all. "Nesterov Step Reduced Gradient Algorithm for Convex Programming Problems", Lecture Notes in Networks and Systems, 2020, 81, pp. 140-148. https://doi.org/10.1007/978-3-030-23672-4_11
23. Tarik, A., and all. "Recommender System for Orientation Student" Lecture Notes in Networks and Systems, 2020, 81, pp. 367-370. https://doi.org/10.1007/978-3-030-23672-4_27
24. Sossi Alaoui, S., and all. "A comparative study of the four well-known classification algorithms in data mining", Lecture Notes in Networks and Systems, 2018, 25, pp. 362-373. https://doi.org/10.1007/978-3-319-69137-4_32
25. Gonzalez-Argote J. Patterns in Leadership and Management Research: A Bibliometric Review. Health Leadership and Quality of Life 2022;1:10-10. <https://doi.org/10.56294/hl202210>
26. Castillo-Gonzalez W. Charting the Field of Human Factors and Ergonomics: A Bibliometric Exploration. Health Leadership and Quality of Life 2022;1:6-6. <https://doi.org/10.56294/hl20226>

FINANCING

No financing.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Rajae Ghanimi, Fadoua Ghanimi, Ilyas Ghanimi, Abdelmajid Soulaymani.

Research: Rajae Ghanimi, Fadoua Ghanimi, Ilyas Ghanimi, Abdelmajid Soulaymani.

Drafting - original draft: Rajae Ghanimi, Fadoua Ghanimi, Ilyas Ghanimi, Abdelmajid Soulaymani.

Writing - proofreading and editing: Rajae Ghanimi, Fadoua Ghanimi, Ilyas Ghanimi, Abdelmajid Soulaymani.