A machine learning based approach for urgent poisoning diagnostic in the emergency

Un enfoque basado en el aprendizaje automático para el diagnóstico urgente de envenenamiento en situaciones de emergencia

Rajae Ghanimi, Khalil Chouikri, Ilyas Ghanimi, Fadoua Ghanimi, Abdelmajid Soulaymani

ABSTRACT

Faced with the scale of cases of acute poisoning, whether accidental or voluntary and requiring admission to the emergency services, the integration of the in silico approach in the process of diagnosis, prognosis and treatment is of paramount importance. This approach, centered on artificial intelligence (AI), is based on prediction based on significant clinical data, thus supporting practitioners and helping them to target the most likely toxic substances. The objective is to make a prediction upstream of the confirmation stage, which would require biological and toxicological investigations that are often costly and time-consuming. With this in mind, our work focuses on the development of a Machine Learning (ML) algorithm capable of predicting the causative toxic agent, providing essential information on the predominant clinical signs. Although many studies in the literature have addressed the use of technology and artificial intelligence in diagnosis, monitoring and pharmacology, we did not find any publications concerning the use of artificial intelligence for the diagnosis or the aid in the diagnosis of poisoning cases. This innovation will therefore constitute the strong point of our research work. Our machine learning algorithm is based on a prediction process based on the in-depth analysis of clinical data provided by the clinical examination of the patient as soon as he arrives at the emergency room. By taking into account a set of parameters such as the symptoms present, the medical history and the circumstances surrounding the exposure, the model can establish relevant links between the clinical signs and the potential toxic agents. By emphasizing the speed and accuracy of prediction, while recognizing the crucial importance of biological and toxicological analysis to confirm diagnoses, our approach has the potential to optimize clinical management by directing the physician to appropriate measures more quickly. As a decision support tool, it offers a relevant first predictive assessment from the patient’s admission.

Keywords: Intoxication; Diagnostic; Machine Learning; Artificial Intelligence; Poisoning; Emergency Department; Emergency.

RESUMEN

Ante la magnitud de casos de intoxicación aguda, ya sea accidental o intencionada, que requieren admisión en los servicios de emergencia, la integración del enfoque in silico en el proceso de diagnóstico, pronóstico y tratamiento es de suma importancia. Este enfoque, centrado en la inteligencia artificial (IA), se basa en la predicción a partir de datos clínicos significativos, apoyando así a los profesionales y ayudándoles a identificar las sustancias tóxicas más probables. El objetivo es realizar una predicción antes de la etapa de confirmación, que a menudo requiere investigaciones biológicas y toxicológicas costosas y que llevan tiempo.
INTRODUCTION

In the field of toxicological emergencies, identifying the substance(s) causing the poisoning is an essential and decisive step. However, there may be times when a comatose patient presents with an unlabeled container or is unable to provide a consistent history. In these cases, a focused physical examination and routine clinical laboratory testing can often provide a tentative diagnosis, allowing empiric interventions or specific toxicological testing to be considered; in this case, management of the intoxication remains essentially symptomatic and is based largely on this clinical approach. The usefulness of a systematic emergency search for the toxin responsible during the management of poisoned patients still arouses debate and the relevance of such analyzes due to economic (high cost) and temporal (vital emergency) constraints, the choice to prescribe such toxicological analysis must be considered and justified. Situations where the indication of a specific treatment and the assessment of severity or prognosis are linked to the results of toxicological investigations are still rare in the literature.

Today, with the advent of artificial intelligence (AI) and the use of machine learning (ML) algorithms, these technologies have become indispensable in all fields, especially in the field of health. They offer remarkable opportunities to improve the different stages of the care process. The use cases for ML are manifold, ranging from making a near-perfect diagnosis to predicting the most accurate readmissions and identifying high-risk patients. These predictions are based on aggregate records of symptoms from anonymized patients.

The main objective of our work is to develop a new algorithm based on a machine learning model, to help doctors to diagnose the different cases of poisoning which could be serious and dangerous threatening the life prognosis. The idea comes from the fact that in the event of an emergency, each minute of delay in the diagnosis risks darkening the prognosis. Especially that the cognitive capacities of a doctor do not allow him to restore the whole of the toxic substances at the origin of the clinical signs of the patient, in particular, in the absence of objective indicators at the time of the interrogation. Our machine learning model based on different algorithms would be able to guide the doctor in his diagnostic process, directing him towards the most probable leads.

Various scientific studies have demonstrated that acute poisoning is a common condition in patients presenting to emergency departments around the world, requiring costly medical care. The variety of toxic substances involved is constantly changing. Some of the most commonly implicated toxic agents include chemicals, insecticides, drugs, and herbs. The clinical approach of poisoned patients is based on the analysis of the history of intoxication, the search for clinical signs and characteristic biological markers constituting a toxidrome.

According to estimates by the World Health Organization (WHO), the number of cases of unintentional acute poisoning varies between 3,5 and 5 million each year worldwide. Of these cases, 3 million are serious, resulting in 20 000 annual deaths. In industrialized countries, the most frequent cases of voluntary poisoning are linked to overdoses of analgesics, anxiolytics and antidepressants. In addition, mortality in patients with alcohol use disorder (AUD) increased by more than 20% in 2020 and 2021, during the COVID-19 pandemic. Several factors contribute to the increase in the number of poisoning cases in developing countries, including the lack of regulation, facilitating access to various pharmaceutical and chemical products without individual and regulatory oversight.

Palabras clave: Intoxicación; Diagnóstico; Machine Learning; Inteligencia Artificial; Envenenamiento; Departamento de Emergencias; Emergencia.

Data and Metadata. 2023; 2:110
collective protective measures.\textsuperscript{(10,11,12,13,14,15)} In 2012, the WHO revealed that 193,460 people die each year from unintentional poisonings, and 84 \% of these cases occur in middle and low-income countries.\textsuperscript{(16)} As a result, the health effects of poisoning in these countries can be considerably more severe than in high-income countries. The situation is alarming: nearly one million people dying each year from suicide by chemicals. In Africa, toxic exposures are often linked to the economic and professional activities of populations.\textsuperscript{(16)} The practices of traditional pharmacopoeia represent another source of intoxication for disadvantaged populations, particularly through suicide and drug consumption among young people.\textsuperscript{(16)} Other causes of poisoning include the use of abortifacient plants among young women, drug poisoning, drug use among young people, as well as the use of counterfeit drugs in Africa, characterized by imprecise composition and inadequate labeling. The continued use of pesticides withdrawn in the West is also a major problem in Africa.\textsuperscript{(17)}

Chavan et al.\textsuperscript{(18)} presented a classification model based on the k-nearest neighbor (k-NN) algorithm. This model was built for 118 chemicals from the NEDO (Repeated Dose Toxicity New Energy and Industrial Technology Development Organization) RTD database, currently known as the Hazard Evaluation Support System (HESS). Both acute toxicity classes (LD50) were used as the response, and a set of eight PaDEL-derived fingerprints served as predictors. The model developed from the Estate fingerprints correctly predicted the LD50 classes for 70 of the 94 chemicals in the training set and 19 of the 24 chemicals in the test set. For each chemical, an individual category was formed by extracting the corresponding k analogues that were identified by the k-NN classification. These categories were used to perform the read-across study for the prediction of chronic toxicity, in particular the lowest observed effect levels (LOEL).\textsuperscript{(18,19,20)}

Other works, in particular those of JC Carvaillo, R Baroukli, X Coumoul and K Audouze\textsuperscript{(19,20,21)} have developed the AOP-helpFinder program, which makes it possible to predict the toxic effect of products by carefully analyzing a vast amount of data published in the scientific literature. The program is based on two methods: the first consists of a text analysis based on the search for words of interest, in particular terms designating chemical substances such as bisphenol S or pesticides, as well as terms describing pathological biological processes. The second method is based on the search for keywords of interest. To identify the relevant terms to search for, the team of researchers compiled "dictionaries" grouping together all the known names of the substance to be studied, such as bisphenol S, found in the PubChem database, as well as thousands of pathologies and adverse effects pathways (Adverse Outcome Pathways, AOPs).

Other works, such as those by Huiling Chen et al.\textsuperscript{(20)} have studied paraquat poisoning and explored the use of common blood indices to identify the degree of PQ toxicity and/or diagnose paraquat poisoning in patients using a machine learning-based approach. Specifically, using a support vector machine-based method combined with a feature selection technique, the research team was able to accurately predict PQ poisoning risk status. This method was tested on 79 people (42 men and 37 women; 41 living patients and 38 deceased patients). The detection method has been rigorously evaluated on a real data set to determine its accuracy, sensitivity and specificity. Feature selection was also applied to identify factors correlated with risk status, and results showed that there are significant differences in common blood indices between PQ-infected individuals who are deceased and those who are alive (p-value = 0.01). Feature selection also revealed that the most important correlated indices are white blood cells and neutrophils. The results of this work demonstrated that the toxicity or prognosis of paraquat poisoning can be preliminarily determined by routine blood tests, without PQ concentration data, thus providing an additional tool and novel approach to assess the prognosis of PQ poisoning.

Although many studies in the literature have addressed the use of technology and artificial intelligence in diagnosis, monitoring and pharmacology, we did not find any publications concerning the use of artificial intelligence for the diagnosis or the aid in the diagnosis of poisoning cases. This innovation will therefore be the subject of our research work.

Our project presents an innovative methodology aimed at building a machine learning model using various algorithms in order to assist emergency doctors (or doctors working in emergency departments) in their poisoning diagnosis process. This model aims to guide them towards the most likely leads, thus facilitating their clinical decision-making. To better understand our approach, we will present in detail the methodology adopted. The flowchart below (figure 1) summarizes the different stages of the research process carried out.

\section*{METHOD}

The first step of our methodology is to collect and prepare the relevant poisoning data from different sources, such as the toxicological databases of the National Poison Control Center (of MOROCCO) and scientific publications. This step is crucial to guarantee the quality and reliability of the data used in our model. Next, we perform a data exploration and analysis phase, during which we identify key features that can be used as predictors in our machine learning model. We use advanced statistical techniques and visualization tools to understand the relationships between different variables and their impact on the diagnosis of intoxication.
Once the relevant characteristics have been identified, we move on to the model selection step. We test four machine learning algorithms, such as support vector machines, Random forest, decision trees, etc., to see which performs best in terms of accuracy. The fourth step of our methodology consists in training the selected model on the training data. We use cross-validation techniques to assess model performance and avoid overfitting. Next, we proceed to evaluate the model on an independent test dataset to estimate its ability to generalize to new data. This step is essential to ensure that the model is reliable and robust in real situations.

Finally, we present the results obtained and interpret the predictions of the model. We also highlight the advantages and limitations of our approach, highlighting the prospects for improving and extending this work.

Before starting the development and design phase of our algorithm, the establishment of a reference database is of crucial importance to list the major toxic effects of substances, in particular those presenting life-threatening risks, i.e. which damage essential organs (heart, lung, brain, kidneys, liver). In this study, we exploited the database of the National Poison Control Center of Morocco, including 29 268 poisoning cases recorded between March 1980 and December 2014 (anonymized cases: no identifying information). The goal was to extract the data needed to train our algorithm. This database is a valuable resource for developing a successful machine learning model. By using this real data, our algorithm will be able to identify relevant patterns, establish links between certain symptoms and potential toxic agents, and thus provide informed predictions for the diagnosis of intoxication.

With the aim of a rigorous analysis of our database and to ensure consistency with the clinical semiology of poisoning, we undertook an in-depth study of the literature on the clinical signs of poisoning related to substances listed in the International Classification of Diseases ICD-11. This exhaustive review allowed us to collect information on the most common toxic substances as well as their main clinical symptoms in case of intoxication, in particular those classified in the Common Classification of Diseases - CIM-11, taking into account the Moroccan national context.

To do this, we carefully analyzed more than 50 bibliographical references, including articles from the National Poison Control Center. (21,22) Thanks to this review of the literature, we were able to establish a reference database grouping the pathophysiological profile of each chemical substance, thus guaranteeing a perfect match with the data from scientific publications. (23,24,25,26,27,28,29) Table 1 presents part of this synthesis work.

To ensure a thorough and precise analysis of our dataset, while remaining in perfect harmony with the knowledge established in the scientific literature. It is a fundamental pillar of our approach, allowing us to develop an efficient and reliable machine learning algorithm, capable of identifying the characteristic clinical signs associated with each toxic substance, and of providing valuable assistance to healthcare professionals in the diagnosis and management of poisoning cases.
To include a case of poisoning in our training dataset, we have established certain essential criteria to be met, thus ensuring the relevance and reliability of the information collected. These criteria are as follows:

1. Elimination of non-toxic causes: Before including a case of poisoning in our database, any non-toxic cause must be retitled. This assures us that the cases retained are indeed linked to toxic agents.

2. Identification of the toxic agent: For each case, it is crucial that the responsible toxic agent is correctly identified. The diagnostic approach can be divided into two distinct nosological frameworks:
   a) Poisoning by a misidentified poison: data from the interrogation preselected the nature (insecticide, medicine, drug without specifying the subtype): refer to the clinical signs and the data from the interrogation.
   b) When there is suspicion of poisoning by an unknown toxin, our approach is to gradually downgrade to poisoning by an identified toxic agent. The objective is to confirm this intoxication based on an in-depth clinical examination followed by additional examinations, ranging from the simplest to the most sophisticated.

3. Presence of significant symptoms: The cases included in our database must present major clinical symptoms, justifying urgent treatment.

This allows us to focus on the most critical poisoning situations and ensure appropriate patient care. The application of these rigorous criteria in the selection of poisoning cases guarantees the quality and relevance of the training data for our algorithm. By selecting only cases with clear identification of the toxic agent and significant symptoms, we ensure a powerful and accurate machine-learning model capable of providing informed predictions and helping healthcare professionals make informed decisions when managing ED poisoning cases.

1. Among the criteria for exclusion:
   2. Asymptomatic cases were eliminated;
   3. Cases with incomplete information: such as the absence of the name of the toxic substance;
   4. Patients admitted to the emergency room in agony;
   5. Patients who have presented minimal symptoms: such as asthenia, dry mouth;
   6. Cases where the cause of poisoning is clinically obvious (insect bites);
   7. Data that does not comply with clinical picture of literature data;
   9. Duplicates, inconsistent or superfluous information.

In summary, after a careful selection and preparation process, our database contains well-structured and suitable poisoning cases to start learning our artificial intelligence algorithm. This data provides a comprehensive view of symptoms, involved toxicants, and demographic characteristics, paving the way for in-depth analysis and accurate predictions to aid healthcare professionals in the management of poisoning cases.

Our learning database contains 29,268 cases of poisoning recorded in Morocco between 1981 and 2014 in the registers of the National Antipoison Center, covering the entire Moroccan territory, all regions as well as all age groups and sexes. However, we encountered challenges when handling this data, as it was misclassified and presented in a non-standardized manner.

https://doi.org/10.56294/dm2023110
Some symptoms were scattered, information was missing, and data was not properly labeled, requiring extensive manual preparation. To remedy this situation, we undertook a process of reclassification of symptoms according to the principles of medical semiology, with the aim of grouping them appropriately. Frequent and relevant symptoms such as headaches, vomiting, abdominal pain and hyperthermia remained individualized.

Following this selection, we retained a set of 27,701 cases of poisoning. The data retained includes 11 variables and 143 toxic substances (see figure 2).

Of these variables, 10 are categorical in nature and represent target responses, while one variable is numeric, representing patient age. Categorical variables include patient sex and presenting symptoms, such as headache, vomiting, coma, abdominal pain, neurological disorders, hyperthermia, cardiac disorders, ophthalmologic disorders, and respiratory disorders (see table 2 for detailed meanings of each variable).

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Descriptions and (under symptoms)</th>
<th>Variable type</th>
<th>Range of value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age of the patient</td>
<td>Numerical</td>
<td>2-94</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex of the patient $F=0$; $M=1$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Respiratory disorders</td>
<td>Respiratory distress, dyspnea, chest pain, congestion: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Cardiac disorders</td>
<td>Palpitations, faintness, syncope, bradycardia, Arrhythmia: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Neurological Disorders</td>
<td>Agitation, convulsion, epilepsy, tremor, hypotonia, hallucinations: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Ophthalmologic disorders</td>
<td>Mydriasis, Miosis: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Headache</td>
<td>Headaches, migraines: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Vomiting</td>
<td>Vomiting, nausea: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Hyperthermia</td>
<td>Fever: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Coma</td>
<td>Coma, obnubilation, dizziness: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>Abdominal pain: $1=yes$; $0 =no$</td>
<td>Categorical</td>
<td>0,1</td>
</tr>
</tbody>
</table>

The patients come from both urban and rural areas and comprise different ages: newborns (maternal-fetal intoxication); baby walkers, childrens, teenagers; adults and elderly peoples (see figure 3).

From figure 3 we can see that in the data set, most patients with intoxication are aging between 13 and 50, while only few people have intoxication aged under 13 or above 50.
Figure 2. Graphical representation of toxic substances
Machine Learning Algorithms

Machine learning is a field of artificial intelligence. It includes the methods aiming to improve or construct models from data. Machine learning is the science of automatically learning a predictive function from a labeled or unlabeled data. Its applications are many and varied, ranging from search systems and character recognition to web search.

Machine learning is a fairly broad field, but we may classify the problems it addresses on four main classes.

- **Supervised learning**: The goal of supervised learning is to learn to make predictions, given a list of labeled examples, i.e. accompanied by the value to be predicted.
- **Unsupervised learning**: consists of identifying groups in an unlabeled data. This allows us to understand their general characteristics, and possibly infer the properties of an observation based on the group to which it belongs.
- **Semi-supervised learning**: involves learning labels from a partially labeled dataset. The first advantage of this approach is that it avoids having to label the entire training examples.
- **Reinforcement learning**: In reinforcement learning, the learning system can interact with its environment and perform actions. In return for these actions, he obtains a reward, which can be positive if the action was a good choice, or negative otherwise. The reward can sometimes come after a long series of actions.

**XGboost**

In boosting techniques, weak classifiers are built sequentially on the entire database; at each step a classifier is built from the previous classifiers by weighting predicted observations more heavily incorrectly so as to focus efforts on these difficult-to-classify observations.

The best-known example of a boosting type method is undoubtedly XGBoost (eXtreme Gradient Boosting). This is a popular and performing technique for building a strong classifier from weak classifiers. It is an example of gradient boosting where the weak classifiers are decision or regression trees, which consists of using an approximation of order 2 rather than order 1 in the minimization, as well as a regularization term in order to avoid overfitting.

**Support Vector Machine**

Support vector machines (also called support vector machines), or SVMs, are powerful machine learning algorithms. Support Vector Machine was chosen as one of the models because it is an algorithm for classification and regression. The Support Vector Machines are a class of learning algorithms initially defined for discrimination, that is to say the prediction of a variable binary qualitative. They were then generalized to the prediction of a variable quantitative. In the case of the discrimination of a dichotomous variable, they are based on the search for the optimal margin hyperplane which, when it is possible, correctly classifies or separates the data while being as far as possible from all observations. The principle is therefore to find a classifier, or a discrimination function, with the greatest possible generalization capacity (prediction quality).

This approach stems directly from Vapnik's work in the theory of learning from 1995. It focused on the generalization (or prediction) properties of a model by controlling its complexity. The founding principle of SVMs is precisely to integrate the control of the complexity, i.e. the number of parameters associated in this case with the number of support vectors. The other guiding idea of Vapnik in this development is to avoid substituting for the initial objective.

The basic principle of SVMs is to reduce the problem of discrimination to the linear one of finding an optimal hyperplane. Two ideas or tricks to achieve this goal:

a) The first consists in defining the hyperplane as the solution of a problem constrained optimization whose objective function is not expressed than using scalar products between vectors and in which the number “active” constraints or support vectors control the complexity of the model.

b) The transition to the search for nonlinear separating surfaces is obtained by the introduction of a kernel function in the scalar product implicitly inducing a nonlinear transformation of the data to an intermediate space (feature space) of greater dimension. This tool is becoming widely used in many types of application and proves to be a serious competitor of the most efficient algorithms (aggregation of models). It should be noted that, on the algorithmic level, these algorithms are more penalized by the number of observations, i.e. the number of potential support vectors, than by the number of variables. Nevertheless, efficient versions of the algorithms allow taking into account large databases in times of calculation acceptable.

**Decision Tree**

A decision tree is a structure widely used in Machine Learning. Its operation is based on heuristics constructed using the techniques of supervised learning. Decision trees have a hierarchical structure and are composed of
nodes and leaves (also called terminal nodes) connected by branches. In their graphical representation, the root is placed at the top and the leaves at the bottom. The internal nodes are called decision nodes. They can contain one or more rules (also called tests or conditions).

The values that a variable can take in a decision tree are called instances or attributes. The terminal nodes contain the class also called target. After its construction, a decision tree can be translated in the form of a set of decision rules.\(^{15}\)

**Random Forest**

Grouping methods are very powerful methods in practice, which are based on the idea that combining many weak predictors makes it possible to obtain performance that is far superior to the individual performances of these weak predictor, because their errors compensate for each other.

The power of ensemble methods is revealed when weak learners are independent conditional on the data, in other words as different from each other as possible, so that their errors can compensate for each other. To achieve this objective, the idea of random forests, also proposed by Leo Breiman, is to build individual trees not only on different samples (as for bagging), but also using different variables.\(^{16,17}\)

In practice, random forests are one of the most efficient and simplest algorithms to set up. They also have the advantage of little dependence on their hyperparameters, namely the number of variables considered at each node, the number of observations used for each tree, the maximum number of observations in the leaves of the tree and the number of trees.

**RESULTS**

To evaluate the performance of the Random Forest Classifier, XGBoost, SVM and Decision Tree Classifier models for causal toxic agent detection, we used the accuracy as the metric commonly used by researcher. A confusion matrix is a specific table allowing visualization of the performance of a classification algorithm. While each row contains information associated with actual classifications, each column indicates the number of predicted classification done by a classification model.

The number TP denotes the number of true positives (positive samples that have been classified as positive), FN denotes the number of false negatives (positive samples classified as negative), FP presents the number of false positives (negative samples classified as positive), and TN represents the number of true negatives (negative samples classified as negative). The confusion matrix for a binary classifier is shown in table 3.\(^{38}\)

As can be seen, such a representation allows visual inspection of prediction errors with little effort as they are clearly located outside the diagonal of the table.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Negative</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

A number of widely-used statistical measures are often derived from a confusion matrix including:

- **Accuracy** \((Ac)\), which is the fraction of correct classifications by the number of samples studied according to the following equation 1.

\[
Ac=\frac{TP+TN}{TP+FP+TN+FN} \quad (1)
\]

- **Precision** \((Pr)\), which is defined as the ratio of true positives to all predicted positives and can be derived using equation 2.

\[
Pr=\frac{TP}{TP+FP} \quad (2)
\]

- **Recall** which is also called True Positive Rate and Sensitivity. It is defined as the proportion of actual positives that are correctly identified as positives which can be computed using equation 2. Examples include the percentage of patients that are correctly identified as having the condition.

\[
R=\frac{TP}{TP+FN} \quad (3)
\]

The Accuracy score showed that all models achieved good results for our task. Random Forest Classifier has the highest Accuracy score of 99.86, followed by One Vs Rest Classifier with an Accuracy score of 99.46, SVM with an Accuracy score of 99.16 and Decision Tree Classifier with an Accuracy score of 99.04.

The results suggest that all four models are effective in detecting the agent responsible for intoxication. However, the Random Forest Classifier model outperformed the other models in terms of F1-score and Accuracy score. (table 4)

https://doi.org/10.56294/dm2023110
Table 4. Results for our new approach applied to the balanced data

<table>
<thead>
<tr>
<th>Models</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy_score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest Classifier</td>
<td>89.66</td>
<td>98.73</td>
<td>99.86</td>
</tr>
<tr>
<td>OneVsRest Classifier</td>
<td>98.67</td>
<td>99.71</td>
<td>99.46</td>
</tr>
<tr>
<td>SVM</td>
<td>98.33</td>
<td>99.12</td>
<td>99.16</td>
</tr>
<tr>
<td>Decision Tree Classifier</td>
<td>99.63</td>
<td>97.51</td>
<td>99.04</td>
</tr>
</tbody>
</table>

DISCUSSION

With the primary goal of finding a way to predict and treat the disease, researchers continue to take advantage of constant advances in technology. In this context, after using different methods such as early stage screening, data augmentation, they turned to the use of ML which is becoming a popular tool in the research community. The work presented in the context of this article has been the subject of a contribution to the design of intelligent systems for the prediction of the toxic substance (etiological diagnosis) based on Machine Learning algorithms. Thus, we have presented in this article the overall architecture of our model based on machine learning for the prediction of intoxication. We have detailed our contribution starting with data collection, data pre-processing, model creation and finally evaluation. We finalize the paper with the results obtained and the discussion. This work has therefore enabled the validation and evaluation of the performance of each of the techniques studied. The comparative study undertaken between the Random Forest Classifier, One Vs Rest Classifier, SVM and Decision Tree Classifier with the aim of a decisive choice of the method best suited to the prediction was therefore made.

The results obtained first showed that these four models were similar in terms of accuracy, which gives them the advantage of being integrated into an online prediction system. Models are accurately evaluated so that they can be compared and the best model is determined based on need. The results of this study should be considered in light of certain limitations. It should be noted that the dataset used here is a subset of the original database, which contained 76 attributes instead of 11, which is used in this study to improve accuracy and reduce the false negative rate of machine learning algorithms. The limitations of this study indicated the following areas as recommendations for future work. First, include other attributes from the original dataset to discover the machine-learning algorithm with the highest accuracy and lowest false negative rate. Second, since each patient has different health conditions, it is recommended to group patients with similar health conditions and ages to study the accuracy and false negative rate of each machine-learning algorithm. Third, the poisoning dataset we have compiled provides valuable information about poisoning cases; however, it does not detail specific patient medical histories. However, it is essential to consider the patient’s medical history, as these can radically influence the profile of the intoxication and even modulate its severity. Unfortunately, we do not have information about the profile of women, and although we can assume that the dataset took into account the parameter of pregnancy, since abortion is among the causes of intoxication. Another limitation of our database is that it consists mainly of patients of Moroccan origin, and it is unclear to what extent genetic origins might influence the intoxication profile. To address these limitations, we plan to expand the scope of our study beyond the Moroccan datasets by collecting additional data from diverse populations. We will seek to study how to generalize the machine-learning framework that we propose by comparing it to other algorithms on these new data collected.

Our goal is to develop a personalized health solution, taking into account all parameters and other toxicants. This innovative approach will provide a better understanding of the interactions between patients’ medical history, their poisoning profile and genetic factors, leading to more efficient and personalized management of poisoning cases.

REFERENCES


41. Gutiérrez VF. La estructura organizacional del Gobierno Regional de Moquegua y su eficiencia funcional.

https://doi.org/10.56294/dm2023110
Sincretismo 2021;2.


49. Murillo-Ticona TA, Berneso-Soto ML. Los Entornos Virtuales de Aprendizaje al rescate del servicio educativo. Sincretismo 2020;1


FINANCING
No financing.

CONFLICT OF INTEREST
None.

AUTHORSHIP CONTRIBUTION
Conceptualization: Rajae Ghanimi, Khalil Chouikri, Ilyas Ghanimi, Fadoua Ghanimi, Abdelmajid Soulaymani.
Research: Rajae Ghanimi, Khalil Chouikri, Ilyas Ghanimi, Fadoua Ghanimi, Abdelmajid Soulaymani.
Drafting - original draft: Rajae Ghanimi, Khalil Chouikri, Ilyas Ghanimi, Fadoua Ghanimi, Abdelmajid Soulaymani.
Writing - proofreading and editing: Rajae Ghanimi, Khalil Chouikri, Ilyas Ghanimi, Fadoua Ghanimi, Abdelmajid Soulaymani.