Artificial Intelligence to analyze and limit drug-related problems in hospitals

Sarah BEN OTHMAN [a], Bertrand DECAUDIN [b], Pascal ODOU [b], Chloé ROUSSELIÈRE [c], Etienne COUSEIN [c] and Slim HAMMADI [a]
[a] *Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France,*
[b] *Univ. Lille, CHU Lille, ULR 7365 - GRITA, F-59000 Lille, France*
[c] *CHU Lille, Institut de Pharmacie, 59000 Lille, France*

Abstract. The health product circuit corresponds to the chain of steps that a medicine goes through in hospital, from prescription to administration. The safety and regulation of all the stages of this circuit are major issues to ensure the safety and protect the well-being of hospitalized patients. In this paper we present an automatic system for analyzing prescriptions using Artificial Intelligence (AI) and Machine Learning (ML), with the aim of ensuring patient safety by limiting the risk of prescription errors or drug iatrogeny. Our study is made in collaboration with Lille University Hospital (LUH). We exploited the MIMIC-III (Medical Information Mart for Intensive Care) a large, single-center database containing information corresponding to patients admitted to critical care units at a large tertiary care hospital.

Keywords. Artificial Intelligence, Decision support system, Drug interactions, Medical prescription review.

1. Introduction

Iatrogeny refers to all disorders, undesirable effects or illnesses caused by a medical intervention, treatment or medication and represents a major public health issue, both in terms of health and economics. Medication-related iatrogeny, induced by the administration of one or more drugs, concerns 6 to 10% of hospitalized patients and is responsible for 3.5% of deaths [1]. In a French study, it was shown that 33% of serious adverse drug reactions were drug-related and that 51.2% could have been avoided [2]. In this context, making therapeutic management safe is a priority for health care institutions in order to reduce drug-related problems by acting simultaneously on the various stages of the health product circuit: prescription, transcription, delivery, administration and follow-up. There are many possible errors in a prescription such as unusual dosage for a specific active ingredient, inappropriate route of administration, toxic interaction between two drugs, etc. The analysis of prescriptions is therefore complex and must absolutely consider the following three criteria: the prescription itself, compatibility with the patient's biological characteristics and analyses and possible drug interactions. Automated and computerized systems were quickly identified as tools to prevent these medication errors and therefore as means to increase health safety. Several studies have already been published on the development of alert rules and computerized Clinical Decision Support (CDS) alert systems [3] [4]. However, these systems tend to send many unnecessary alerts, which leads to alert fatigue in healthcare [5]. Other studies are interested in patient and drug profiling to refine the analysis of prescriptions and target the patients and drugs most at risk [6]. In this paper, we propose a novel AI module to analyze the compliance of pharmaceutical prescriptions, in collaboration with LUH. Our system exploits a database containing the information of each prescription, as well as the characteristics and biological analyses of patients. The validity of the prescription, as well as its compatibility with the patient's condition and analyses are studied using clustering and classification methods. Potential drug interactions between two drugs prescribed to the same patient during the same hospital stay are also analyzed.

2. Material and methods

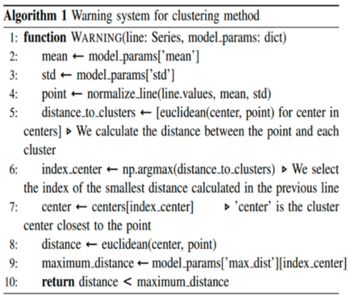
We use the MIMIC-III data ('Medical Information Mart for Intensive Care') which is a database accessible to all researchers worldwide after completing an appropriate online training and signing a data use agreement. The open nature of this database makes it perfect for making studies reproducible and for conducting academic or industrial research projects [7]. The MIMIC-III data include deidentified clinical data on 38,597 adult patients admitted between 2001 and 2012 to the intensive care unit at Beth Israel Deaconess Medical Center in Boston, Massachusetts. The data also include 7870 neonates admitted between 2001 and 2008. It stores data in a structured format and displays characteristics like mass or multi-ownership.

2.1. Data pre-processing

MIMIC-III data are grouped into 26 different tables, grouping different types of medical data. Our work focuses on the PRESCRIPTIONS table. First, we merge the PATIENTS table with the PRESCRIPTIONS table. Then, for each prescription row in this new table, we add the laboratory measurements from the LABEVENTS table that share the same hospital stay identifier as the prescription. Our database presents missing data which is an unavoidable problem in statistical practice. For randomly missing data, the probability that a variable is not measured depends only on the state of some other observed variables. When this probability no longer depends on the observed variables, the data are said to be missing completely at random [8]. We therefore deleted all prescription lines with missing drug or dose. With the help of pharmacists, we have identified the relevant biological assays kept in our dataframe such as Anion Gap, Bicarbonate, Calcium (Total), Chloride, Creatinine, Glucose, Hematocrit, Hemoglobin, Mean Corpuscular Hemoglobin, etc. According to the drug, administration route and prescription unit, we divided the resulting dataframe into several dataframes. The prescriptions in the MIMIC-III database can be considered as "valid" prescriptions (already validated by a pharmacist) or "positive" according to good practices in prescribing drugs. We therefore only have validated prescriptions, the "positives" and must create "negatives", i.e., problematic prescriptions that would be refused by a pharmacist. In fact, imbalanced datasets cause a difficulty for most standard ML classifiers which assume a nearly balanced class distribution and are biased to the majority class [9]. Thus, we use reprocessing methods that modify the training dataset to balance distributions by creating these "negative" or problematic prescription rows. We start with prescription rows from the MIMIC database and then randomly change the numerical values (doses, posologies, etc.) by considering a confidence interval calculated as the mean of the column ± its standard deviation. The negative prescriptions created are individually problematic prescriptions, without considering potential drug interactions and should therefore only be used to test methods that analyze each prescription individually.

2.2. Analysis of each prescription individually

In order to study each prescription individually we started by testing clustering methods. To create the clusters, only MIMIC data is used, i.e., valid and compliant prescriptions. For each training dataframe corresponding to a Drug (D), a Prescription Unit (PU) and a specific Route of Administration (RA), we cluster the data, with a k-means clustering model and with a hierarchical clustering model. The k-means clustering algorithm takes as input the data to be clustered and a number of clusters to be formed. However, the hierarchical clustering model takes as input only a distance from which two clusters should not be merged and starts by considering each point as a cluster. Then, the algorithm finds the two closest clusters to each other, and merges them into one cluster. Each point of the cluster corresponds to a prescription corresponding to this D, PU and RA. Algorithm 1 presents the alert algorithm that analyzes a new prescription to be studied by also taking as input the parameters of the clustering model used and trained and indicates if the prescription is normal.



```
Algorithm 1 Warning system for clustering method
function WARNING(line: Series, model_params: dict)
    mean ← model_params['mean']
    std ← model_params['std']
    point ← normalize_line(line.values, mean, std)
    distance_to_clusters ← [euclidean(center, point) for center in centers] ▷ We calculate the distance between the point and each cluster
    index_center ← np.argmax(distance_to_clusters) ▷ We select the index of the smallest distance calculated in the previous line
    center ← centers[index_center]    ▷ 'center' is the cluster center closest to the point
    distance ← euclidean(center, point)
    maximum_distance ← model_params['max_dist'][index_center]
    return distance < maximum_distance
```

Figure 1 : Alert algorithm analyzing a new prescription

In order to classify positive and negative prescription, XGBoost and Random Forest (RF) methods have been tested to analyze validity of a prescription individually, without taking into account possible drug interactions with other prescriptions of the patient for the same hospital stay

2.3. Methods of drug interaction analysis

In addition to studying each prescription individually, it is interesting to study possible drug interactions with other drugs prescribed to the same patient during the same hospital stay. To analyze potentially clinically relevant drug-drug interactions, we start by listing in a list of lists (LL) the drugs that do not interact, i.e., the drugs that have been prescribed to the same patient, in the same hospital stay, during the same period. To test whether a new prescription could conflict with other drugs that the patient is already taking, we start by listing all the drugs prescribed to this patient during the same hospital stay as the prescription to be tested. Then, we check if the drug to be tested has already been prescribed with these drugs in the LL. If the answer is no, then an alert is returned to warn of the potential risk of drug interaction.

3. Results

To study the results of the clustering methods, we focused on the clustering of the drug ACET325, with the prescription unit mg and with the administration route 'PO'. We chose this drug because the MIMIC database contains more than 30000 prescriptions of this drug. We first tested our clustering method with the K-means algorithm. We varied the number of clusters parameter and obtained the best results for 5 clusters (Figure 2).

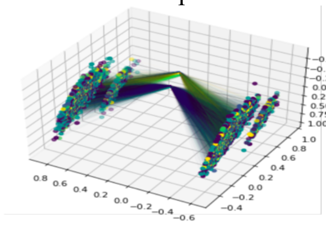


Figure 2: 3D visualization of clusters of ACET325 prescriptions (5 clusters)

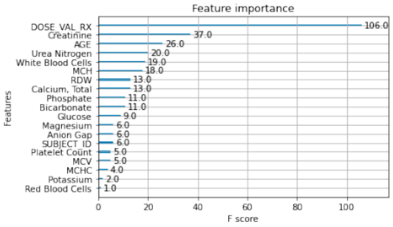


Figure 3: Importance of features in the classification of prescriptions using XGBoost

The 3D visualization shows that the centers of the various clusters are well separated and not almost confused/overlapped. The use of a metric such as the accuracy rate, when, as here, most observations belong to the same category (negative prescriptions), will lead to get a model that is not enough "smart" and almost systematically predicts the dominant class. The performance evaluation criteria chosen are therefore a combination of Precision, Recall, and F1-score. This combination integrates the expected concordance by chance, as well as the sensitivity, which focuses on the TPR (True Positive Rate), while the specificity is given as an indication but is not used to classify the models.

Table 1. Evaluation metrics of the ML methods used

	Clustering		**Classification**	
Score	***Kmeans***	***Hierarchical Clustering***	***XGBoost***	***Random Forest***
Accuracy	0,4373	0,4391	0,7960	0,9942
Precision	0,3043	0,3078	0,7845	0,9906
Recall	0,4146	0,4174	0,8162	0,9978
F1score	0,3510	0,3544	0,8000	0,9942

The results returned when the models were run on the test set are shown in Table 1. For instance, the overall accuracy on the test set is around 43% and the F1score=35% for both methods. Then we tested our clustering method with the hierarchical algorithm. This algorithm finds out more than 2800 clusters among these prescriptions, with a TPR=2916 and a TNR (True Negative Rate) = 5216. Regarding XGBoost and RF methods, the two confusion matrices given were used to calculate the Accuracy, Precision, Recall and F1score methods (Table 1). However, the negative data we generated are not suitable for testing drug interactions between prescriptions. To test drug interactions, new negative data could be created by creating several prescriptions with drugs that have clinically relevant interactions for the same patient and same hospital stay. To do this, we can rely on the results of published drug interaction studies. We would also need the help of the pharmacists to define and validate a complete list of drug-drug interactions.

4. DISCUSSION

Clustering methods used tend to over-categorize prescriptions as negative. This can be explained by the fact that using the Euclidean distance, we give equal importance to each column of the data. However, some features could be considered more than others. A future track would be to determine the weight of each feature with the help of the medical profession and LUH, for a more adapted calculation of the distances between the prescriptions. The graph showing the importance of each feature in the classification process of XGBoost clearly shows that there are 3 most important parameters: the dose of drug prescribed, the age of the patient, and the creatinine level (Figure 3). The same result was given by RF but in a different order. These three features were validated by pharmacists. Indeed, the dose prescribed to the patient is intuitively the central parameter. Moreover, the recommended dose varies greatly depending on the age of the patient. In addition, many drugs cannot be prescribed if the patient has renal failure. Creatinine levels are an indicator of kidney function: high creatinine levels indicate kidney failure. It is valuable to know which features are most important in the prescription classification process. This could lead in the future to give more weight to the already defined most important features in the calculation of the distances between prescriptions.

5. Conclusion

Our study in collaboration with LUH allowed us to explore different algorithmic and statistical methods to analyze pharmaceutical prescriptions, with the aim of limiting drug iatrogeny. For this work, we used the MIMIC-III database. Classification and clustering methods have been proposed and compared. Concerning the future perspectives, it would be interesting to explore the tracks explained before to improve the clustering methods and explore more techniques for minority oversampling.

References

[1] Kohn, L.T., Corrigan, J.M., Donaldson, M.S., McKay, T. Pike, K.C. (2000). To err is human: Building a safer system. Washington, DC: National Academy Press.

[2] L'état de santé de la population en France, Rapport 2017.

[3] A. Barbier, C. Rousselière, L. Robert et al., Élaboration d'un guide méthodologique sur l'implantation d'un système d'aide à la décision pharmaceutique : retour d'expérience d'un centre hospitalier uni- versitaire fran ̧cais, Annales Pharmaceutiques Françaises, https://doi.org/10.1016/j.pharma.2022.06.009

[4] E. Cuvelier, L. Robert, E. Musy, C. Rousselière, R. Marcilly, S. Gautier, P. Odou, J.-B. Beuscart, B. Décaudin, The clinical pharmacist's role in enhancing the relevance of a clinical decision support system, International Journal of Medical Informatics, Vol 155, 2021, 104568, ISSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2021.104568.

[5] H. Bouami et P. Millot, Healthcare Delivery System Security: the orchestration of automated and organizational solutions , IFAC-Pap., vol. 52, no 19, p. 127-132, janv. 2019, doi: 10.1016/j.ifacol.2019.12.123.

[6] E. Cousein et al., Effect of automated drug distribution systems on medication error rates in a short-stay geriatric unit , J. Eval. Clin. Pract., vol. 20, no 5, p. 678-684, 2014, doi: 10.1111/jep.12202.

[7] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035

[8] Rubin, Donald B. "Inference and Missing Data." *Biometrika* 63, no. 3 (1976): 581–92. https://doi.org/10.2307/2335739.

[9] J. Corny et al., A machine learning–based clinical decision support system to identify prescriptions with a high risk of medication error , J. Am. Med. Inform. Assoc., vol. 27, no 11, p. 1688-1694, nov. 2020, doi: 10.1093/jamia/ocaa154.