Using Association Rules in Antimicrobial Resistance in Stone Disease Patients

Ioannis MANOLITSIS [a,1,†], Georgios FERETZAKIS [b,c,†], Lazaros TZELVES [a],
Dimitris KALLES [b], Evangelos LOUPELIS [d], Stamatis KATSIMPERIS [a],
Thanos KOSMIDIS [f], Athanasios ANASTASIOU [e], Dimitrios KOUTSOURIS [e],
Stavroula KOFOPOULOU [e], Vassilios S. VERYKIOS [b], Andreas SKOLARIKOS [a]
and Ioannis VARKARAKIS [a]

[a] *Second Department of Urology, National and Kapodistrian University of Athens,
Sismanogleio General Hospital, Athens, Greece*
[b] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[c] *Sismanogleio General Hospital, Department of Quality Control, Research and
Continuing Education, Marousi, Greece*
[d] *Sismanogleio General Hospital, IT department, Marousi, Greece*
[e] *National Technical University of Athens (NTUA)*
[f] *CareAcross Ltd, London, United Kingdom*
†*Equal contribution (1st authors)*

Abstract. Association rule mining is a very popular unsupervised machine learning
technique for discovering patterns in large datasets. Patients with stone disease
commonly suffer from urinary tract infections (UTI), complicated by the
emergence of antimicrobial resistance (AMR), due to the excessive use of
antibiotics. In this study, we explore the use of association rule mining in the AMR
profile of patients suffering from stone disease.

Keywords. Association rules, unsupervised ML, Antimicrobial Resistance, AMR

1. Introduction and background

Urolithiasis is a disease commonly associated with urinary tract infections (UTI), as
about one-third of all cases are complicated by the presence of UTIs [1]. The selection
of proper antibiotics in patients with stone disease is a challenge, taking into account
their poly-microbial profile, widespread use of antibiotics, and the emergence of
antimicrobial resistance (AMR) [2]. The presence of various resistance determinants
among pathogenic microorganisms poses the need for a systematic assessment of
potential association factors [3]. The unequivocal emergence and growth of artificial
intelligence (AI) and machine learning (ML) techniques in the field of microbiology
can assist in providing better healthcare in the era of AMR [4-6]. Association rules
are one of the unsupervised data mining techniques used to identify relationships
between items present in a database [7]. In this study, we explore association rule

---

[1] Corresponding Author, Ioannis Manolitsis, MD, PhD(c), Second Department of Urology, National
and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece; E-mail address:
giannismanolit@gmail.com.

mining in the AMR profile of patients suffering from stone disease.For discovering correlations between exposures and disease outcomes in clinical data, association rule mining has been suggested to be more sensitive than regression techniques or chi-square testing [8] and has also beenused in studies of nosocomial infection trends [9,10].

2. Methods and Materials

This observational retrospective study has been approved by the Institutional Review Board of Sismanogleio Hospital. The raw data from urolithiasis patients attending the Urology Department during 2019 were extracted. Data from 239 patients were retrieved, including gender (binary), Gram stain (categorical- positive, negative), sample type (categorical-blood, urine, pus), antibiotics (categorical-39 different antibiotics), and a binary class attribute (antimicrobial susceptibility-resistant, sensitive). The dataset consisted of 5.156 instances in total.

In our case, the data set is organized into rows and columns, with one 'transaction' per row and one 'item' per column. Each 'transaction' represents a susceptibility testing of an isolate, and the 'items' correspond tothe categorical variables: gender, microorganism species, sample type, antibiotics, and antimicrobial susceptibility.The mathematical formulation of the association rules can be presented as follows [11]. Let $I = \{i_1, i_2, \dots i_m)$ a set of items and $D = \{T_1, T_2, \dots T_m)$ a set of transactions. Each transaction T is a set of items such that $T \subseteq I$. If a transaction T contains X, this is denoted by $X \subseteq T$. An association rule is an "if-then" rule stating that X associates with Y, denoted by $X \Rightarrow Y$, where $X, Y \subseteq I$, and $X \cap Y = \emptyset$. The left-hand side, LHS, of a rule is called antecedent, and the right-hand side, RHS, is called the consequent. Given the rule $X \Rightarrow Y$, the support ($supp$), the confidence ($conf$) and the $lift$ are defined as follows:

$$supp(X \cup Y) = \frac{number\ of\ transactions\ containing\ X\ and\ Y}{total\ number\ of\ transactions}$$

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} = \frac{number\ of\ transactions\ containing\ X\ and\ Y}{number\ of\ transactions\ containinig\ X}$$

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \times supp(Y)}$$

Support is the percentage of transactions containing both parts of the rule, while confidence is the number of cases in which the rule is correct relative to the number of cases to which it is applicable. An interesting rule must at least have support and confidence values greater than the user-specified minimum thresholds, *minsup* and *minconf*, respectively.The lift measures how many times more often X and Y occur together than expected if they are statistically independent of each other. Lift is a measure of how X and Y are really related rather than coincidentally happening together. Only if the lift is greater than 1we have an indication that there is some usefulness to the rule. A larger value of lift suggests a greater strength of the association rule $X \Rightarrow Y$.Note that the implication symbol means co-occurrence and not

causality.We performedassociation rule mining using the Rprogramming language. Specifically, we used the '*arulesViz*' and '*arules*' packages and the Apriori algorithm [11], one of the most popular algorithms for mining association rules.

3. Results and Discussion

The association rules which can be deduced based on the given antimicrobial susceptibility data set can be parametrized according to the knowledge that clinicians want to extract from the database. For example, by setting *minsup* and *minconf* to 0.01 and 0.35, respectively, the four rules sorted by decreasing order of confidence are presented in Table 1.

Table 1. Sample output

LHS			RHS		Measures		
Gender	Type	Gram	Antibiotic	Susceptibility	Support	Confidence	Lift
Male	Urine	Negative	Ampicillin	R	0.01144	0.81944	2.69455
Male	-	Negative	Ampicillin	R	0.01299	0.81707	2.68675
-	-	Negative	Ampicillin	R	0.03200	0.75688	2.48883
-	Urine	Negative	Ampicillin	R	0.02890	0.75252	2.47450

For example, the first row of the above table can be formulated as follows:IF the patient is male AND the sample type is urine AND the Gram-stain of the isolate is negative AND the antibiotic used is ampicillin, THEN the susceptibility testing will show that this microorganism is resistant to ampicillin.Thefull set of the rules from the above example alongside the two-key plot, the network graph of the ten most important rules sorted by confidence, and the corresponding R code used are available in the online appendix [12].

Patients with stone disease can suffer from various infectious complications, some of which are potentially life-threatening [13]. The methods that are traditionally employed for the surveillance of AMR are antibiotic sensitivity reports and resistance rates of various antibiotics. However, these methods cannot identify underlying associations and do not consider aggregated data of patients that are essential in monitoring and tackling AMR [14]. Rule mining could discover various associations between binary variables in large sets of data. So, in our case, when AMR data are divided into sensitive and resistant interpretations, association rules appear to be a suitable technique for analyzing AMR patterns. In addition, association rule mining can identify patterns in a dataset, thus providing information about the relationships between antimicrobial susceptibilities [15].

4. Conclusion

There are two key take-away messages from our work. The first is that by drawing tools from the pool of machine learning techniques, one can generate rules that can either explain existing data or predict outcomes of situations. A mature collection of techniques (and association rules are among them) can be adopted and empower professionals to explore data relationships in new ways. The second message is that the new techniques also require a fresh look at the intersection of these techniques with the

medical practice since any substantially new tool requires updating the training curricula of the people who will use it.

Acknowledgements

References

[1]   Chen D, Zhang Y, Huang J, et al. The analysis of microbial spectrum and antibiotic resistance of uropathogens isolated from patients with urinary stones. 2018;72(6):e13205.
[2]   De Lorenzis E, Alba AB, Cepeda M, et al. Bacterial spectrum and antibiotic resistance of urinary tract infections in patients treated for upper urinary tract calculi: a multicenter analysis. European Journal of Clinical Microbiology & Infectious Diseases. 2020;39(10):1971-81.doi: 10.1007/s10096-020-03947-z.
[3]   Feretzakis G, Loupelis E, Sakagianni A, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. Antibiotics (Basel). 2020 Jan 31;9(2):50. https://doi: 10.3390/antibiotics9020050
[4]   Feretzakis G, Sakagianni A, Loupelis E, et al. Machine Learning for Antibiotic Resistance Prediction: A Prototype Using Off-the-Shelf Techniques and Entry-Level Data to Guide Empiric Antimicrobial Therapy. Healthc Inform Res. 2021;27(3):214-21.
[5]   Feretzakis G, Loupelis E, Sakagianni A,et al. Using Machine Learning Algorithms to Predict Antimicrobial Resistance and Assist Empirical Treatment. Stud Health Technol Inform. 2020 Jun 26;272:75-78. https://doi: 10.3233/SHTI200497.
[6]   Yu KH, Beam AL,Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018 Oct; 2(10), 719–731. https://doi.org/10.1038/s41551-018-0305-z
[7]   R. Agrawal, T. Imielinski, A. Swami, Mining association rules betweensets of items in large databases, in: Proceedings of 1993 ACM-SIGMOD International Conference on Management of Data, Washington, DC, May 1993, pp. 207-216.
[8]   Cleophas, T. J., Zwinderman, A. H. (2015). "Association Rules Between Exposure and Outcome (50 and 60 Patients)," in Machine Learning in Medicine - a Complete Overview, Springer, p. 475–479.
[9]   Cazer CL, Al-Mamun MA, Kaniyamattam K, et al. Shared Multidrug Resistance Patterns in Chicken-Associated Escherichia coli Identified by Association Rule Mining. Front Microbiol. 2019 Apr 12;10:687. doi: 10.3389/fmicb.2019.00687. Erratum in: Front Microbiol. 2020 Jun 30;11:1359.
[10] Giannopoulou, E. G., Kemerlis, V. P., Polemis, M., et al.(2007). "A large scale data mining approach to antibiotic resistance surveillance," in Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07) (Maribor: IEEE).
[11] R. Agrawal, R. Srikant. "Fast algorithms for mining association rules", Proceedings of 20th International Conference on Very Large Data Bases, Santiago de Chile, Chile, September 12-15, 1994.
[12] Feretzakis G Using Association rules in Antimicrobial Resistance in stone disease patients. Available online: www.learningalgorithm.eu/AssociationRules_ICIMTH2022.html (accessed on 18 March 2022).
[13] Tzelves L, Skolarikos A. Suction Use During Endourological Procedures. Current Urology Reports. 2020;21(11):46.doi: 10.1007/s11934-020-00998-9.
[14] Feretzakis G, Loupelis E,Petropoulou S, et al. Using Microbiological Data Analysis to Tackle Antibiotic Resistance of Klebsiella Pneumoniae. Stud Health Technol Inform. 2019 Jul 4;262:180-183.
[15] R. Mureşan, C. Zaharia,D. Zaharie. "Using Additive Bayesian Networks and Association Rules in Antimicrobial Resistance Analysis," *2020 International Conference on e-Health and Bioengineering (EHB)*, 2020, pp. 1-4.