Evaluation of drug and disease influence on occurrence of hyperkalemia: a data mining approach

Petra VOLK MARKOVICa, Tomislav LAPTOSb, Marko ROBNIK-SIKONJAb, Andreja CUFARa,1

a University Medical Centre Ljubljana
b University of Ljubljana, Faculty of computer and information science

Keywords. Data mining, hyperkalemia, interactions, association rules

Introduction

Health records offer a plethora of data that can be transformed into useful information using data mining tools and clinical knowledge of interdisciplinary healthcare teams [1]. Clinical data and prescribed medications were analysed to uncover rules and connections with the hyperkalemia occurrence.

1. Methods

The retrospective study was conducted on 150 patients with hyperkalemia during either the admission or hospitalization or their visit to outpatient clinics. Medical records were examined by clinical pharmacists.

The demographic characteristics, diagnosis, serum potassium level and medication therapy administered to the patient one day before the detection of hyperkalemia were collected. Medical data and additional background knowledge were analysed using two different machine learning algorithms for association rule mining to extract knowledge in the form of association rules [2, 3]. Detected rules were reviewed and evaluated by clinical pharmacists.

2. Results

Both algorithms were tested on artificial data and on the clinical data set from University Medical Centre Ljubljana. Two given hierarchies (ATC for drugs and ICD for diseases) were used as a background knowledge. Generated rules were classified by clinical pharmacists as either (1) interesting (unexpected but plausible), (2) potentially

1 Corresponding Author.
interesting (requiring further analysis), (3) expected (confirming available background),
(4) trivial (connecting diagnosis and therapy of the same disease), or (5) undefined
(irrelevant due to high level of hierarchy).

Generalized association rule algorithm [3] used with both given hierarchies produced the biggest body of 8283 association rules that were sorted according to the rule confidence criterion [3]. 630 best rules were reviewed and 2.8% were classified as interesting, which is not surprising considering other studies involving association rules e.g., [2, 3]. The same algorithm was employed also without ATC and ICD hierarchies and generated 68 and 15 rules, respectively. It produced some interesting rules that were consistent with rules generated using given hierarchies.

The second algorithm used high-utility pattern mining [3]. It produced 8 sets of 6 to 27 rules using both given hierarchies that turned out to be consistent with generalized association rules produced using both hierarchies.

3. Discussion

The results are promising as several interesting interactions leading to serum potassium level disturbance were detected. The most interesting discovered rules describe an association between the occurrences of hyperkalemia and hypothyroidism. This association was is rarely found in medical textbooks and can therefore be assumed not obvious. It may be concluded that data mining methods employed are useful in detecting hidden information from health data to improve patients’ safety and quality of health care.

References