Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

© 2022 European Federation for Medical Informatics (EFMI) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220452

Electronic Phenotyping to Identify Patients with Arrythmia Disease from a Hospital Information System

Masaharu NAKAYAMA[a,1] and Ryusuke INOUE[b]
[a] *Medical Informatics, Tohoku University Graduate School of Medicine, Miyagi, Japan*
[b] *Medical IT center, Tohoku University Hospital, Miyagi, Japan*

Abstract. Electronic phenotyping is an important method to identify a disease group by collecting clinical data from hospital information systems. This study aimed to extract accurate cases of supraventricular arrythmia, ventricular arrythmia, and bradycardia from clinical data of a hospital information system. The electronic phenotyping algorithm was improved using the machine learning method. Subsequently, it showed a higher area under the curve for prediction and higher specificity. However, the algorithm needs further improvement to classify each arrythmia disease accurately. In conclusion, phenotyping using clinical data from hospital information systems has some affinities and issues depending on the disease.

Keywords. Electronic phenotyping, machine learning, arrythmia disease.

1. Introduction

The use of medical big data is expected to provide real-world evidence research. Hospital information systems are very useful in this regard as they contain massive amounts of high-quality clinical data. Electric phenotyping has attracted attention in recent years as a tool to accurately identify patients diagnosed with a particular disease based on clinical data from hospital information systems [1,2]. This study attempted to extract true cases of several arrythmia diseases that can cause fatal conditions, such as strokes and sudden death because the algorithm to identify arrythmia diseases has not been established well.

2. Methods

The target diseases are arrythmias, including supraventricular arrythmia, ventricular arrythmia, and bradycardia. In this study, supraventricular arrythmia was defined as atrial extrasystole, atrial tachycardia, paroxysmal supraventricular tachycardia, atrial flutter, or atrial fibrillation. Ventricular arrythmia was defined as ventricular tachycardia, ventricular fibrillation, or ventricular extrasystoles. Bradycardia was defined as sinus failure syndrome and atrioventricular blocks. We used data comprising health insurance claims, diagnosis procedure combination data, and clinical information extracted from the Standardized Structured Medical Information eXchange version 2 standard storage

[1] Corresponding Author, Masaharu Nakayama, Japan; E-mail: nakayama@cardio.med.tohoku.ac.jp.

of Tohoku University Hospital (TUH) from April 1 to December 31, 2013 [3]. The initial rule was based on the International Classifications of Diseases 10th revision codes, examinations, therapeutic drugs, and procedures related to arrythmia diseases. Two physicians, including a cardiologist, checked the medical records to determine the true cases out of each 60 cases randomly selected; then, they calculated the positive predictive value (PPV). Using the evaluations, we trained the XGBoost model using five-fold cross-validation in R to improve algorithms, and then we evaluated the area under the curve (AUC) [4]. Finally, we retrospectively tested the algorithm on a cohort from an inpatient database of the TUH Department of Cardiovascular Medicine. This study was approved by the ethics review committee of Tohoku University (No. 2020-1-459).

3. Results

Regarding supraventricular arrythmia, although the initial rule had low PPV (27.8%), our model with XGBoost showed high correlation with radiofrequency catheter ablation, verapamil, and mexiletine (0.847 of AUC). The specificity was 96.0%. About ventricular arrythmia, the PPV was only 8.3% based on the initial rule. The applied method showed high correlation between implantable cardioverter-defibrillator (ICD), exchange of ICD, and amiodarone with 0.801 of AUC. The specificity was 97.4%, although the PPV increased to only 27.3%. For bradycardia, the PPV was initially 8.3%. The method showed high correlation with implantation and replacement of the pacemaker with 0.861 of AUC, although the PPV increased to only 23.3%. The specificity was 95.2%.

4. Discussion

Our study's results showed high specificity in the electronic phenotyping of arrythmia disease, resulting in usefulness of the related clinical data, such as disease names, examinations, medication, devices, and procedures, to exclude arrythmia cases among patients. However, there was little power to classify each arrythmia disease accurately because the use of related medication, examinations, and procedures overlapped among each arrythmia disease. This suggests that other important information, such as the Minnesota Code for electrocardiograms, would be critical to identify specific arrythmia cases. In conclusion, the findings suggest that phenotyping using clinical data from hospital information systems has some affinities and issues depending on the disease. To address these issues, it may be necessary to narrow down the target group more clearly based on the disease and collect disease-specific data.

References

[1] Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc. 2014 Mar-Apr;21(2):221-230.

[2] Kashyap M, et al. Development and validation of phenotype classifiers across multiple sites in the observational health data sciences and informatics network. J Am Med Inform Assoc. 2020;27:877-83.

[3] Yamaguchi M, et al. Establishment of the MID-NET ® medical information database network as a reliable and valuable database for drug safety assessments in Japan. Pharma Drug Saf. 2019;28:1395-1404.

[4] Izukura R, et al. The development of an electronic phenotyping algorithm for identifying rhabdomyolysis patients in the MID-NET database. Stud Health Technol Inform. 2019 Aug 21;264:1498-1499.