Linnæus University · Universidad de Murcia · Nantes Université · Semmelweis University 1769 · euniwell European University for Well-Being

ASTUTENESS PROJECT

AI-Driven Tools in Healthcare: A Visual Trustworthy Treatment Decision Support Systems

www.um.es/astuteness

WP1: Supervised Deep Learning for Drug Related Problems

Responsible: Alisa Lincke (LNU)
Partners involved: UMU, LNU
Researchers: Francisco Jose Mora Caselles (UMU), Alisa Lincke (LNU), Tora Hammar (LNU), Olof Björneld (LNU)

October, 2023

European Universities · Erasmus+

With the support of the Erasmus+ Programme of the European Union

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101035821.

Contents

Background

There is a general increase of medicine use and more patients are prescribed several different medications at the same time, which in turn leads to an increased risk for ***drug-related problems (DRPs)***

DRPs occur frequently and are a common cause of suffering, hospitalizations, and death.

Identifying DRPs is a challenging task due to a number of factors to be considered (multiple diseases and medications, metabolism, usage time, dosage, pharmacological properties of a drug, etc.).

Methods to identify DRPs:

- expert judgment (standardized and subjective)
- medication review with the involvement of patients (non-standardized and subjective)
- knowledge-based algorithms such as ***clinical decision support systems (CDSS)*** (standardized and objective)
- ***Machine Learning (ML) algorithms*** to detect, and predict the risk of DRPs

Previous and Ongoing Research at LNU

- The prevalence of potential DRPs identified using ***Janusmed knowledge database*** developed in Sweden (Region Stockholm)
- The combination of ***machine learning algorithms*** with ***Janusmed knowledge database*** (as additional feature) to improve the prediction of DRPs

DISA Seed Projects (2021): “Data Intensive analysis for Identification and Prediction of Risk Medications and side effects in Kalmar region (DIIPRIM)”.

FORTE Project (2022-2025): “Prediction of medication risks and drug-related problems – a novel pharmacoepidemiological approach using real-world data, decision support algorithms and machine learning”.

Janusinfo REGION STOCKHOLM

WP1: Motivation

One of the nine categories of side effects included in Janusmed Riskprofile is QT-prolongation. Several different types of medications can cause **QT-prolongation** which in turn may lead to serious consequences such as arrhythmia, cardiac arrest, and death.

In the pilot of our research project we focus only on this risk category as many doctors want support related to medications with a risk of QT-prolongation

In this study, we have focused on predicting one of the QT-prolongation adverse outcomes (**ventricular arrhythmia,**) on patients from the **Kalmar County Region** using the medication data.

There are various causes for **ventricular arrhythmias**, such as long QT syndrome, heart failure, coronary artery disease, *medications (drug side effects)*, and others.

WP1: Aim and Objectives

The primary **aim** of this study is to predict the negative outcome (ventricular arrhythmias) in patients using two approaches supervised and unsupervised machine learning.

Objective 1 *(Supervised Machine Learning).* Compare two machine learning methods, one with capturing temporal dependencies (such as recurrent-neural networks (RNN)) and another without capturing temporal dependencies (such as traditional machine learning models, e.g., Random Forest (RF)), in predicting the negative outcome (Ventricular arrhythmia) using the medication data.

Objective 2 *(Unsupervised, Anomaly Detection, Pattern Mining).* Explore the medication use in patients with high and low Janusmed risk for patients with ventricular arrhythmias outcome.

WP 1: Methodology

Objective 1: the ***experimental research*** (supervised machine learning) with controlled experiments to evaluate the performance of machine learning models . The two models were selected which are ***Gated Recurrent Unit (GRU)*** and ***Random Forest (RF)*** after empirical benchmarking different models in the first iteration. Experimental research was selected for its reliability and validity to compare different machine learning models.

Objective 2: the ***exploratory research*** (data mining, unsupervised machine learning) to gain a better understanding of the dataset itself. The anomaly detection (**DBSCAN clustering algorithm**) was selected to find group of patients with anomalies. To further explore the clusters, **SDMap*** algorithm is applied to compare medications use. The exploratory research method is particularly useful for studying a large datasets (with more than 300 dimensions) and time series data.

WP 1: Dataset

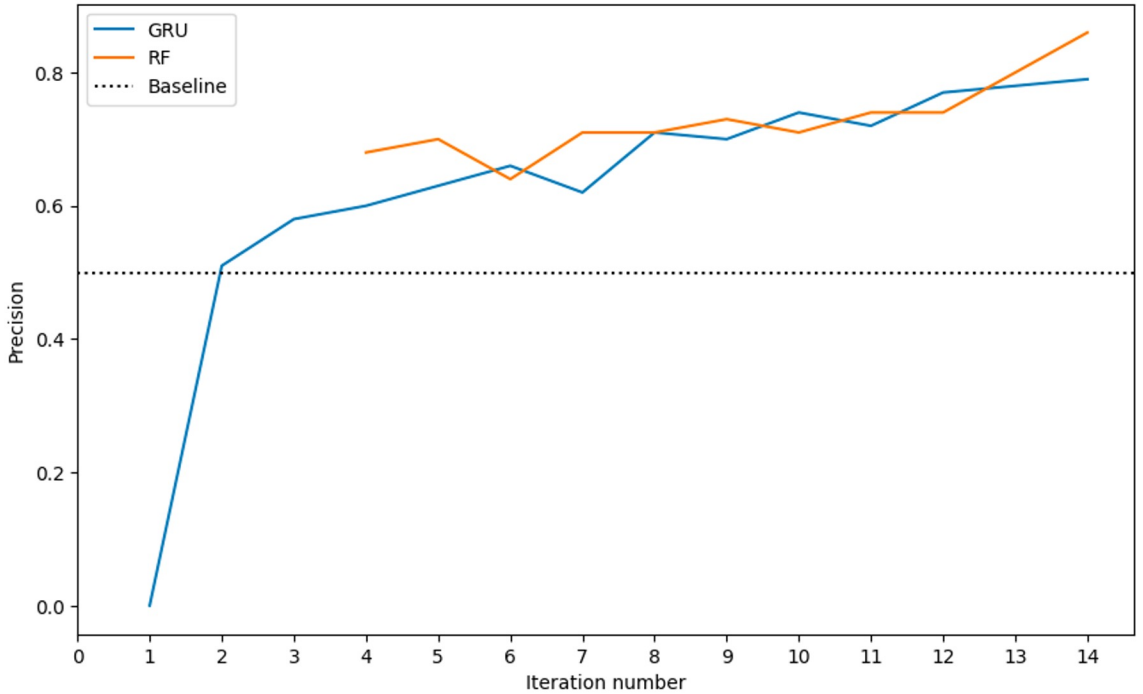
This study employs health data on drug-related events retrieved for the period of 10 years (01.01.2011-01.01.2021) from the Cambio COSMIC health care information system in Kalmar, Sweden. The dataset included a total of 281 767 patients (137 441 males and 144 326 females).

The use of an anonymized dataset for research was approved by the Swedish Ethical Review Authority under ethical permits (No. 2021-038800).

Data containing personal data (age and gender), medication events (administered in the hospital or picked up at the pharmacy), and negative effects of QT-interval events (such as 6 different clinical outcomes collected based on ICD-10 diagnosis codes).

For this pilot study, we have used only ***704*** **patients**, where 352 are without an outcome date (*positive class*) and another 352 patients with an outcome date (*negative class*).

Results I (Machine Learning Iterations Overview)



Results I (Machine Learning Final Results)

Table 1. Average metrics obtained when adding Janusmed risk values as an additional column/feature.

Model	**Accuracy, %**	**Precision, %**	**Recall, %**	**F1-SCORE, %**
GRU	69	71	63	67
Random Forest	**73**	74	73	73

Table 2. Average metrics obtained in models trained using only outcome patients (negative class) with a Janusmed risk level of more than 0.

Model	**Accuracy, %**	**Precision, %**	**Recall,%**	**F1-SCORE**
GRU	81	79	84	81
Random Forest	**86**	86	86	86

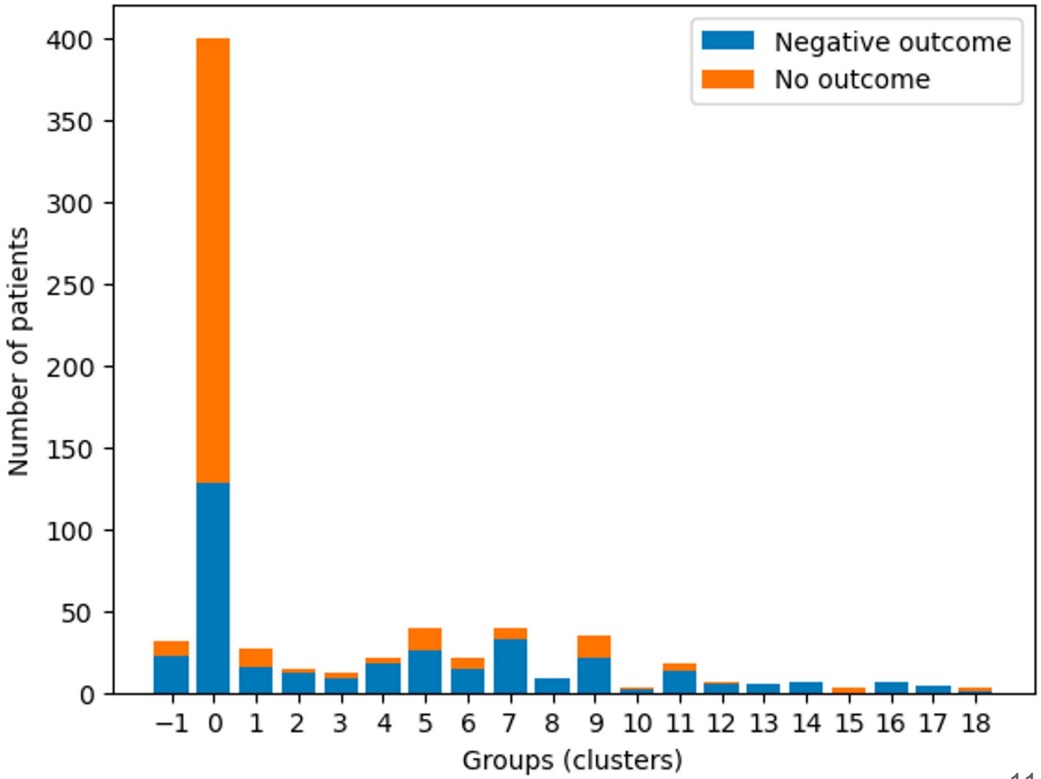
Result II (Data Mining, DBSCAN)

Patients have been grouped in 19 groups (clusters), where one group is called ‘anomaly’ (group number is -1).

Orange are patients without outcome.

Blue are patients with outcome.

Data used for clustering are: max Janusmed risk level for three periods (40 days each, 120 days before the outcome day).



Result II (Data Mining,SDMap*)

Table 1. Top 20 SDMap* results. Sorted by Qc. All patients with negative outcomes were selected. Target: Outcome = Yes.

Description	tp	fp	Confidence	Qc (c = 3)	Qg (g = 50)
C01B = '1'	74	0	1,0	74	1,48
C01B = '1', C07A = '1'	66	0	1,0	66	1,32
B01A = '1', C01B = '1'	60	0	1,0	60	1,2
C03C = '1', C07A = '1'	58	1	0,983	55	1,137
B01A = '1', C01B = '1', C07A = '1'	53	0	1,0	53	1,06
B01A = '1', C03C = '1'	53	1	0,981	50	1,039
C07A = '1', N02A = '1'	61	5	0,924	46	1,109
B01A = '1', C07A = '1', N02A = '1'	47	4	0,922	35	0,87
C07A = '1', C09A = '1'	74	13	0,851	35	1,175
C07A = '1', N02A = '1', N02B = '1'	49	5	0,907	34	0,891
C07A = '1', C09A = '1', C10A = '1'	52	6	0,897	34	0,929
C01D = '1'	47	5	0,904	32	0,855
A06A = '1', C07A = '1'	47	5	0,904	32	0,855
C07A = '1', C09C = '1'	56	8	0,875	32	0,966
C07A = '1', N02B = '1'	96	22	0,814	30	1,333
B01A = '1', C07A = '1', C09A = '1'	65	12	0,844	29	1,048
B01A = '1', C07A = '1', C09A = '1', C10A = '1'	47	6	0,887	29	0,839
B01A = '1', C07A = '1', C10A = '1'	97	24	0,802	25	1,311
A06A = '1', B01A = '1'	54	10	0,844	24	0,9

Result II (Data Mining,SDMap*)

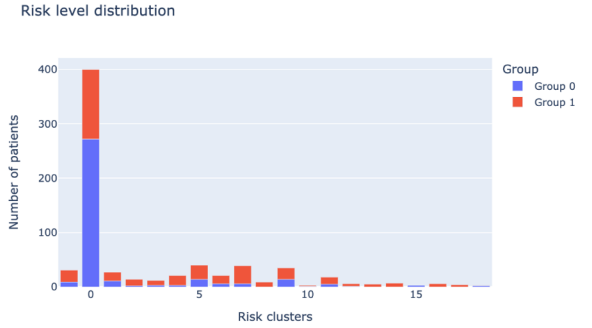
Table 2. Top 20 SDMap* results. Sorted by Qc. Only patients with negative outcome selected. Target: Risk = High.

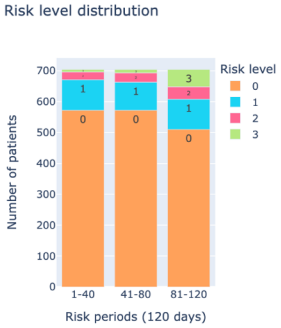
Description	tp	fp	Confidence	Qc (c = 3)	Qg (g = 50)
C03C = '1', C07A = '1'	43	1	0,977	40	0,843
B01A = '1', C03C = '1'	37	1	0,974	34	0,725
B01A = '1', C03C = '1', C07A = '1'	35	1	0,972	32	0,686
C01D = '1', C07A = '1'	30	0	1,0	30	0,6
J01C = '1', N02B = '1'	28	0	1,0	28	0,56
A06A = '1', B01A = '1'	34	2	0,944	28	0,654
B01A = '1', J01C = '1'	28	0	1,0	28	0,56
C07A = '1', J01C = '1'	28	0	1,0	28	0,56
A06A = '1', N02B = '1'	34	2	0,944	28	0,654
B01A = '1', C01D = '1'	28	0	1,0	28	0,56
B01A = '1', C01B = '1'	49	7	0,875	28	0,86
A06A = '1', C07A = '1'	33	2	0,943	27	0,635
B01A = '1', C01D = '1', C07A = '1'	27	0	1,0	27	0,54
C03C = '1', C10A = '1'	30	1	0,968	27	0,588
C03C = '1', N02B = '1'	29	1	0,967	26	0,569
C03C = '1', C07A = '1', N02B = '1'	29	1	0,967	26	0,569
C01B = '1', C07A = '1'	50	8	0,862	26	0,862
C03C = '1', C07A = '1', C10A = '1'	28	1	0,966	25	0,549
B01A = '1', C01B = '1', C07A = '1'	43	6	0,878	25	0,768

Result III (Dashboard)

Janusmed risk level

Selected populations (CLUSTER, GROUP): All selected





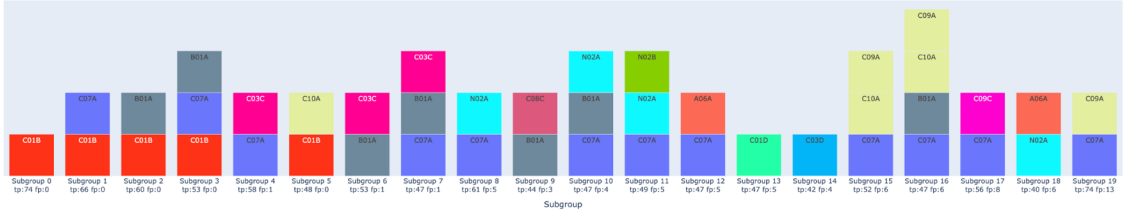
	Group 0	Group 1
%	50	50
Number of patients	352	352
Average medications	14.45	24.21

Sort results by:

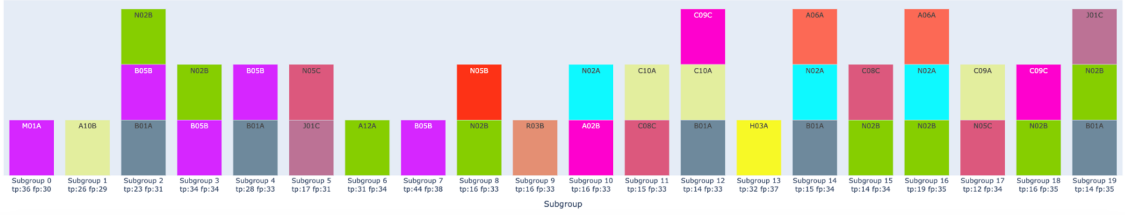
Qc

Select value of c:

SDMap* results for Group 1 (Medications)



SDMap* results for Group 0 (Medications)



WP1: Challenges and Limitations

- Small number of patients with negative outcome (413)
- The ground truth is not a 100% truth. We do not know how many of the outcomes that were caused by medications and how many were caused by other reasons, and there could be patients with the outcome that is not visible in the data (for example if they died without getting the diagnosis)
- ATC CODE is categorical variable which generates ~1000 attributes using one-hot encoding.
- ATC CODE is a basis but more descriptive information about the medication type needs to be included as additional features (such as form, dosage, etc.)
- SDMap* algorithm performance decreases a lot with having many attributes/features.

WP1: Conclusion

- Janusmed Risk Profile is a valuable attribute/feature
- More information about medication needs to be added as additional features, attributes
- DBSCAN algorithm helps to identify different groups of patients which might be useful for further exploration.
- RF outperformed GRU, hence the dataset is too small to have scientific evidence.
- SDMap* algorithm looks promising technique for finding medication combination
- The interactive visualization dashboard can be used to explore medication use and compare different patient groups.
- The proposed approach is general and can be applied to predict other drug-related negative outcomes
- **The work done in WP1 will be presented in Big Data Conference 2023 at LNU**.