Discovering Drug-Drug Interactions Using Association Rule Mining from Electronic Health Records

Nazyrova, N., Chahed, S., Dwek, M., Getting, S. and Chaussalet, T.

https://doi.org/10.1109/inista59065.2023.10310637

Discovering Drug-Drug Interactions Using Association Rule Mining from Electronic Health Records

Nodira Nazyrova[1], Salma Chahed[1], Miriam Dwek[2], Stephen J. Getting[2], Thierry Chaussalet[1]
[1]*School of Computer Science and Engineering,* [2]*School of Life Sciences*
University of Westminster, London, United Kingdom
N.Nazyrova@westminster.ac.uk, S.Chahed@westminster.ac.uk, M.V.Dwek@westminster.ac.uk,
S.J.Getting@westminster.ac.uk, chausst@westminster.ac.uk

***Abstract*—In this paper, we propose utilising Electronic Health Records (EHR) to discover previously unknown drug-drug interactions (DDI) that may result in high rates of hospital readmissions. We used association rule mining and categorised drug combinations as high or low risk based on the adverse events they caused. We demonstrate that the drug combinations in the high-risk group contain significantly more drug-drug interactions than those in the low-risk group. This approach is efficient for discovering potential drug interactions that lead to negative outcomes, thus should be given priority and evaluated in clinical trials. In fact, severe drug interactions can have life-threatening consequences and result in adverse clinical outcomes. Our findings were achieved using a new association rule metric, which better accounts for the adverse drug events caused by DDI.**

***Index Terms*—drug-drug interactions, association rule mining, adverse drug events, polypharmacy, hospital readmission**

I. Introduction

Before a drug is approved for the market, pharmaceutical companies run clinical trials to ensure the safety of the end users. The first phase clinical trials investigate the safety of the drug on a small group of people, while the subsequent phase clinical trials may investigate the best dose of treatments and side effects on a larger group [1]. Following drug approval, further post-marketing clinical trials are conducted to investigate its long-term safety in a real-world setting. Those clinical trials are limited due to the substantial resources and efforts required. To address this drawback, this paper aims to investigate the use of Electronic Health Records (EHR) for detecting potential new drug interactions that lead to adverse clinical outcomes in elderly patients. 30-day hospital readmission is used as a measure of adverse drug events (ADE) caused by drug-drug interactions (DDI); however, other adverse clinical events, such as in-hospital and out-of-hospital mortality, excessive length of stay in hospital or different variations of hospital readmission (7-day, 14-day) can also be used. Our proposal involves utilising Association Rule Mining (ARM) to pinpoint drug combinations that pose a high risk of adverse drug-drug interactions. Additionally, we have introduced a new evaluation metric that aids in identifying previously unknown drug interactions.

II. Problem Background

Polypharmacy is defined as taking five or more medications concurrently, often more than medically necessary [2]. It is more common in older adults, as many have multiple chronic conditions, which increases the complexity of therapeutic management. However, nearly 50% of older adults take one or more medications that are not medically necessary [2]. Research has clearly established a strong relationship between polypharmacy and negative clinical consequences, including drug-drug interactions, adverse drug reactions and multiple geriatric syndromes, such as functional decline, cognitive impairment and falls [2] [3]. These clinical outcomes may lead to rehospitalisation, extended stay in the hospital or patient mortality. According to Glans *et al.* [4], 40% of patients over 65 years were readmitted within 30 days of discharge due to medication-related reasons. Moreover, 54% of older people have at least one clinically significant DDI in their prescriptions [5]. Therefore, identifying drug-drug interactions in clinical prescriptions is a critical area of research. Pharmaceutical companies study the pharmacokinetics and pharmacodynamics of the drugs, trying to predict possible drug interactions during clinical trials. After the drug is released to the market, adverse event reports are collected from health care professionals, consumers, and manufacturers in pharmacovigilance systems, such as the Food and Drug Administration (FDA) Adverse Reporting System (FAERS) and Spontaneous Reporting System (SRS). Detection of DDIs in the drug developmental stage and post-marketing surveillance is challenging as evaluating all possible drug combinations is not feasible. Data mining and machine learning methods can provide a valuable benchmark in detecting new potential DDIs. Such methods may rely on adverse event reports and drug properties, such as chemical structure, targets, anatomical therapeutic chemical classification codes, side effects, medication and clinical observations.

III. Literature Review

Nowadays, many data-driven methods for detecting drug-drug interactions have been proposed. These methods can be broadly categorised into literature-based extraction methods, machine learning-based prediction methods and

pharmacovigilance-based data mining methods. Literature-based extraction methods extract drug-drug interactions from biomedical research papers using natural language processing techniques. With this method, DDI detection is considered a relation extraction task aiming to identify specific relations between the entity pair through documents containing mentions of the drug combinations. Among many, Support Vector Machine (SVM) based models [6] [7], deep learning-based models, including CNN-Based methods [8] [9], RNN-based models [10], and LSTM-based models [11] are some of the most accurate. Such an approach enables the identification of DDIs from existing scientific journals, published articles and technical reports. Therefore, only known DDI can be extracted from the literature. Another approach to finding potential DDIs is machine learning-based models. Traditional classifier algorithms and regression-based methods usually use similarity and dissimilarity metrics to construct features and then apply classifiers or regression models to predict potential DDIs. Numerous studies [12] [13] [14] used feature similarities of drug pairs to predict DDIs using various conventional classification algorithms. Unlike literature-based extraction methods, machine learning-based prediction methods use multiple data sources containing various drug information. Diverse data sources provide heterogeneous and multimodality data for building prediction models. An advance in clinical databases and pharmacovigilance systems enabled the identification of DDIs using specialised databases, such as DrugBank, Micromedex, and Medline [15] [16], using machine learning algorithms. Data mining algorithms, such as reporting odds ratio or Multiitem Gamma Poisson Shrinker (MGPS) [15], have been used to generate and rank the different $drug \rightarrow ADE$ associations in earlier years. Further, more advanced methods using deep neural networks for extracting DDI [16] [17] were used and demonstrated promising results. Association rule mining (ARM) is a technique for exploring interesting relationships among items in a dataset. Association rules were expressed in the form of $X \rightarrow Y$, providing the information in the form of an "$if \rightarrow then$" statement. Association rules are commonly used to detect DDIs using spontaneous reporting systems [15] [18] [19] [20] [21]. Thakrar *et al.* [19] developed one of the early association rule-based methodologies to detect ADEs associated with DDIs using spontaneous report systems. They used two different statistical assumptions for detecting signals of DDI: the additive interaction model and the multiplicative interaction model. The additive interaction model assumes no interaction between two drugs when the excess risk associated with Drug *X* in the absence of Drug *Y* is the same as the excess risk associated with Drug *X* in the presence of Drug *Y*. The multiplicative interaction model assumes no interaction on the multiplicative scales; the relative risk associated with Drug *X* is the same as in both the absence and presence of exposure to Drug *Y*. Later study [20] used ARM for identifying rules in the form of $drug \rightarrow ADE$ from the SRS dataset. The results of this study show that association rules detected both familiar and unfamiliar combinations of drugs and their ADEs in the SRS dataset. However, the study did not consider the rules involving a combination of drugs, such as X and $Y \rightarrow ADE$. This problem was addressed in a study by Noguchi *et al.* [15], where ADEs caused by drug interactions were explored using ARM on the SRS dataset. They proposed an ARM model based on 'lift' and 'conviction', demonstrating high detection power. The results showed a good discriminative performance of the association rule in finding DDIs (sensitivity = 99.05%, specificity = 93.60%). However, since the true data about drug-drug interactions were not available, authors treated drug-drug combinations detected in the previous study [21] as the true data. Most methods currently used for detecting DDIs rely on the existing drug databases or scientific literature. Such work in organising knowledge about drug interactions is essential and has great value in the clinical domain. However, this approach is appropriate only for rediscovering existing clinically proven DDIs. A large corpus of clinical records, such as hospital records with prescription data, is not being studied to analyse ADEs. The aim of this study is to analyse prescription data and use hospital readmission as an indirect signal of ADE for detecting DDIs. This work presents a significant chance to analyse EHR to uncover both current and potential DDIs. It also helps identify priority candidates for further clinical investigations. However, this approach is linked to challenges in evaluating potential drug interactions that are unknown yet. Only indirect evaluation methods can be utilised for new drug interactions that have not yet been clinically proven, as will be described in subsequent sections.

IV. Methodology

To identify DDIs using EHR, the association rule mining approach was used. This approach was further enhanced using a new evaluation score. The full methodology for discovering new potential drug interactions using EHR is provided in Fig. 1.

A. Data Source

The MIMIC IV (Medical Information Mart for Intensive Care) dataset was used for detecting drug-drug interaction from the hospital patients' prescriptions. The MIMIC-IV dataset includes deidentified demographical data, diagnosis, prescriptions, critical health meters and other vital signs [22]. From this dataset, the cohort of elderly people aged 65 and more was selected for the analysis as this group of patients is hospitalised more often and most likely to receive multiple treatments. Only the prescription table with information about the admissions to the hospital was used. A total of 185,157 hospital stays of 73,888 patients were included in the analysis. The prescribed medications were deduplicated and standardised, replacing brand names with generic names where necessary. In total, 2,697,759 prescriptions with 801 unique drugs were analysed.

B. Definitions of Adverse Drug Events

An adverse drug event (ADE) is defined as harm experienced by a patient due to exposure to a medication [23]. ADEs

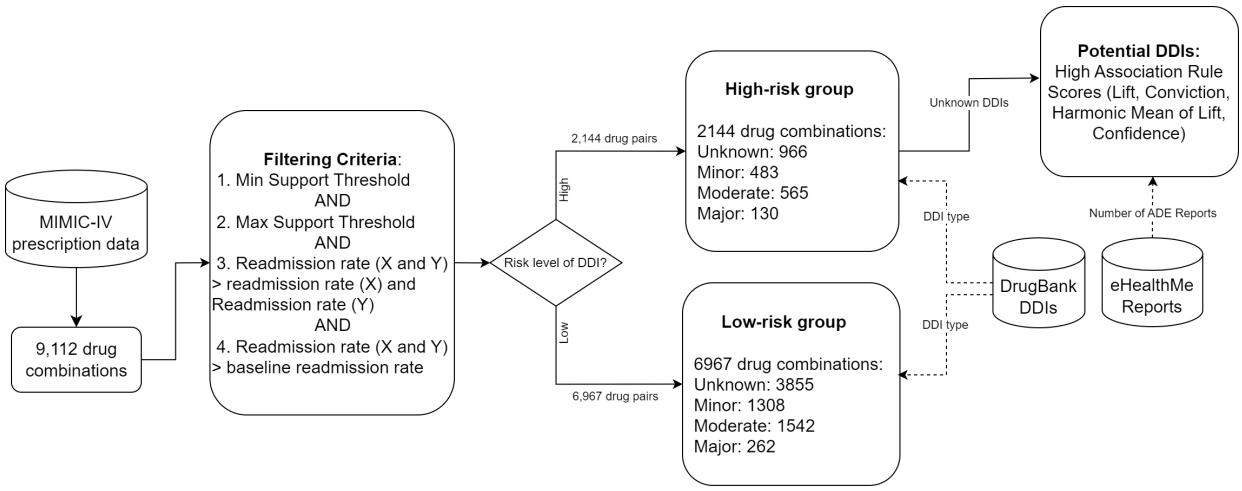

Fig. 1. Methodology for discovering new DDIs using Association Rule Mining approach.

are tightly connected with rehospitalisation, with 40% of such cases accounting for medication-related reasons. Therefore, in this study, 30-day hospital readmission is used as an ADE for DDI signal detection. Thus, drug combination is said to have ADE when the hospital readmission of patients taking both drugs is higher than that of patients taking any of the drugs. Additionally, as ADE should capture the cases of abnormally excessive hospital readmissions, the hospital readmission rate associated with both drugs should be higher than the average baseline readmission rate for the entire cohort of patients.

C. Association Rule Mining Model

Association rule mining is an unsupervised learning technique for detecting interesting relationships between variables in large datasets. An association rule is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint itemsets. In the case of drug-drug interactions, the association rule takes form $X \cup Y \rightarrow ADE$. For this study, an Apriori algorithm is used to implement ARM, as it mitigates the challenge of hard computation of association rules. The Apriori algorithm only considers frequent itemsets which exceed the minimum support threshold. Setting a minimum support threshold can help to eliminate rules that lack generalizability and have limited applicability to diverse cases. The support of an itemset $S(X)$ is the number of records containing X, and the support for the association rules is usually written using a probability notation:

$$Support(X \Rightarrow Y) = P(X \cup Y) \tag{1}$$

Once the frequent itemset is identified, the confidence metric is used to measure the strength of the association rules. Confidence for a 2-itemset is calculated as a percentage of transactions having both items divided by the percentage of transactions having only one item. It can also be written using the following probability notation:

$$Confidence(X \Rightarrow Y) = P(Y|X) \tag{2}$$

Confidence determines how often items in Y appear in records that contain X. In the presence of high-frequency drug in the drug combination, the confidence metric might misinterpret the interestingness of the rule. Moreover, the confidence metric is asymmetric and implies an asymmetric relationship between the antecedent and the consequent, which is not always true. Previous studies on using ARM for detecting DDIs also acknowledged the inappropriateness of traditional Apriori scores to evaluate association rules for DDIs. Instead, Harpaz *et al.* [18] proposed using the Relative Reporting Ratio score, and Noguchi *et al.* [15] suggested Lift (or Interest) and Conviction scores to evaluate obtained rules. For the evaluation in this study, we utilised Lift and Conviction measures. Additionally, we incorporated our enhanced version of the lift score to give more accurate consideration to ADEs within the association rules. The newly developed score is a symmetric measure of rule interestingness and can be used to explore drug combinations causing increased readmission rates.

The lift score measures the relative magnitude of the probability of observing an ADE under the condition of $X \cup Y$, compared to the overall probability of observing an ADE. For the association rule drug X and drug $Y \rightarrow ADE$, the lift is calculated in the following way:

$$Lift(X \cup Y \rightarrow ADE) = \frac{Confidence(X \cup Y \rightarrow ADE)}{Support(ADE)} \tag{3}$$

When the lift value is greater than one, the two events $X \cup Y$ and ADE are not independent, and the higher the value, the greater the relevance of the interactions [15].

Conviction measures the implication strength of the rule from statistical independence. It is defined as a comparison of the probability of $X \cup Y$ without ADE if they were dependent on the actual frequency of $X \cup Y$ without ADE [24]

$$Conviction(X \cup Y \rightarrow ADE) = \frac{1 - Support(ADE)}{1 - Conf(X \cup Y \rightarrow ADE)} \quad (4)$$

Like lift, a conviction value higher than one indicates an interesting rule and means that the consequent ADE is highly dependent on the antecedent $X \cup Y$.

Unlike previous studies, this study uses solely prescription and associated hospital readmission data. Hence, we do not use any hypothesis or prior knowledge of ADEs caused by drug-drug interactions. We cannot detect ADE directly between two drugs, however, we can calculate the signal of ADE associated with drug *X*, ADE associated with drug *Y* and ADE associated with $X \cup Y$. An increase of ADE when both drugs are combined is used as a DDI signal. We propose using a customised version of lift to measure lift for drug associations and corresponding ADE associations in the following way:

$$Lift_{\text{drug(X, Y)}} = \frac{Support(X \cup Y)}{Support(X) \times Support(Y)} \quad (5)$$

where $support(X)$ is the frequency of drug *X*, $support(Y)$ is the frequency of drug *Y* and $support(X \cup Y)$ is the frequency of both drugs.

Lift for ADEs is calculated using Eq. 6:

$$Lift_{\text{ADE(X, Y)}} = \frac{Support(X \cup Y|ADE)}{Support(X|ADE) \times Support(Y|ADE)} \quad (6)$$

where support $(X|ADE)$ is the readmission rate of patients taking drug *X* and not drug *Y*, support $(Y|ADE)$ is the readmission rate of patients taking drug *Y* and not drug X, $support(X \cup Y|ADE)$ is the readmission rate of patients taking drug *X* and drug *Y*.

Both metrics are combined into a single metric using harmonic mean:

$$HarmonicMeanofLift = \frac{Lift_{\text{drug(X,Y)}} + Lift_{\text{ADE(X,Y)}}}{Lift_{\text{drug(X,Y)}} \times Lift_{\text{ADE(X,Y)}}} \quad (7)$$

Association rules having both high readmission rates associated with drug combination and high frequency of this drug combination attained the higher harmonic mean of lift. This is an indication of the high relevance of drug interaction.

The overall framework for detecting drug-drug interactions is provided in Fig. 1. We propose forming two groups of drugs with high and low risks of DDIs. To identify combinations that have a higher likelihood of causing ADEs, namely readmission rate. The criteria for such drug combinations are as follows:

1 Minimum support threshold $(X \cup Y)$ is set to 50 to include the drug combinations with sufficient evidence, hence regarded as generalisable.
2 Maximum support threshold $(X \cup Y)$ is set to 15,000 tto exclude the most common combinations whose hospital readmission rates tend to approximate baseline readmission rates. Typically, these medications are meant to be taken in conjunction with one another.
3 Readmission rate associated with the drug should be higher than the overall baseline readmission rate in the sample, as the high-risk association rules should capture only excessive rehospitalisation cases caused by ADEs.
4 The readmission rate associated with both drugs $(X \cup Y)$ should be higher than the readmission rate associated with any one of the drugs.

Drug interactions that are not yet known in the high-risk group, but have high association rule scores, will be identified as potential candidates for further study.

V. Evaluation

In order to assess the effectiveness of the ARM-based method in detecting new drug-drug interactions, we needed to determine if it is capable of identifying existing DDIs. Further, the new candidate DDIs can be suggested using association rule metrics for further exploration. DrugBank database [25] and Drugs.com website [26] websites were used to check the existence of clinically proven DDIs in both low and high-risk groups. In accordance with the aforementioned filtering criteria, a high-risk group is a group with drug that result in an increased hospital readmission rate. Most pharmacovigilance systems classify drug interactions as Major, Moderate, Minor and Unknown. Major drug interactions are highly clinically significant, and the risk of interactions outweighs the benefits. Therefore, such drug interaction should be avoided. Moderate drug interactions are less clinically significant and can be used under particular circumstances. Minor drug interactions have limited clinical effects. Unknown drug interactions in the high-risk group of drug combinations can be the candidate for the new potential DDI. Only indirect methods are available to evaluate whether the potential drug-drug interactions may result in adverse drug events. Some researchers studied social media posts, such as Twitter, to obtain information about adverse events [27]. Other studies investigate the pharmacodynamics and pharmacokinetics of two drugs and predict possible drug-drug interactions [12]. For this study, we used the eHealthMe pharmacovigilance system, which allows the public to participate in post-marketing clinical trials [28]. It contains data about 47,783 drugs and supplements and 20 million patients. Patients can report any adverse drug events they experienced while taking two or more drugs simultaneously. The number of reports is used as an indication of the relevance of DDIs in evaluating the validity of association rules. Association rules in the high-risk group with the high Lift, Conviction and Harmonic mean of Lifts metrics should be used as candidates for new DDIs and may be used as a shortlisting tool for clinical trials.

VI. Results

A total of 9,112 drug combinations were considered in the final model. Out of which, 1,714 were identified as high-risk drug combinations. It can be seen in Fig. 2 that the group of drug combinations identified as high risk of DDI have higher proportions of Minor, Moderate and Major drug interactions. Notably, the number of major DDIs was almost twice as high in the high-risk group of drug combinations. While minor and moderate drug interactions do not normally have a dramatic

impact on a patient's health status, major drug interactions can lead to adverse clinical outcomes. Therefore, a rapid increase in the amount of major DDI in the high-risk group could lead to excessive 30-day readmission due to adverse drug events.

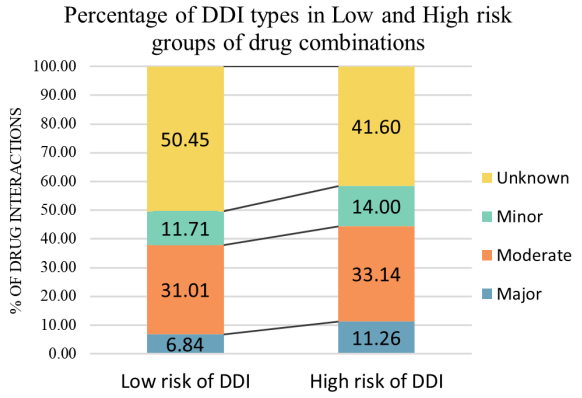

Fig. 2. The Proportion of DDIs in High-risk and Low-risk groups of drug combinations

As shown in Fig. 2, 41.6% of the drugs considered potential drug interactions are unknown yet. Some of these combinations are in the clinical testing phase. Among these unknown drug interactions, those having a higher harmonical mean of interest factor values are the candidates for the new drug interactions. Unknown drug interactions in the high-risk group were examined using the number of patient complaints in the eHealthMe platform. We formed a list of drug combinations using existing association rule metrics (lift and conviction) and our proposed metric (harmonic mean of lift). We set the minimum confidence threshold to 0.10 to exclude low-certainty association rules. We formed the list of high-risk drug combinations with the highest lift, conviction and harmonic mean of lift. This group of drug combinations is expected to have the largest number of ADE reports made by patients. The list of low-risk drug combinations contained drug combinations with the lowest lift, conviction and harmonic mean of lift. The lists of drug combinations using lift and conviction measures were identical. Therefore, two sets of high-risk and low-risk drug combinations were compared, as provided in Tables 1 and 2.

We noticed that the candidate DDIs with the higher harmonic mean of interest factor have more DDI reports on average, and the correlation between the harmonic mean of lift and the number of reports for the high and low-risk groups is 0.9211, which signifies a high positive correlation. Table 1 demonstrates two drug combination groups with strong and weak association rules. The group of drugs with a higher harmonic mean of lift (top-10 strong association rules) have a higher number of DDI reports made on the eHealthMe platform and, on average equals 667 reports in the group.

On the other hand, groups of drugs with weak association rules (top 10 weak association rules) with low harmonic mean of lift have a very low number of DDI reports, on

TABLE I
CORRELATION BETWEEN THE NUMBER OF ADE REPORTS AND HARMONIC MEAN OF LIFT IN STRONG AND WEAK ASSOCIATION RULES.

Drug combination	***Harmonic mean of Lift***	***Number of reports***
Association rules with the highest risk of DDI (Highest harmonic mean of lift)		
Darunavir & ritonavir	8.567×10^{-3}	3,495
Galantamine & memantine	4.238×10^{-4}	111
Clotrimazole & ursodiol	2.903×10^{-4}	135
Amoxicillin & clarithromycin	2.761×10^{-4}	70
Cinacalcet & sevelamer	2.292×10^{-4}	100
Acyclovir & clotrimazole	2.064×10^{-4}	335
Clotrimazole & fluconazole	1.767×10^{-4}	1,213
Calcitriol & febuxostat	1.599×10^{-4}	69
Clotrimazole & nystatin	1.528×10^{-4}	951
Atovaquone & ursodiol	1.144×10^{-4}	191
Average Number of Reports		**667**
Association rules with the lowest risk of DDI (Lowest harmonic mean of lift)		
Amiodarone & aluminium hydroxide	2.896×10^{-6}	8
Hydromorphone & aluminium hydroxide	2.884×10^{-6}	0
Labetalol & aluminium hydroxide	2.867×10^{-6}	0
Hydralazine & aluminium hydroxide	2.601×10^{-6}	0
Hydrochlorothiazide & aluminium hydroxide	2.428×10^{-6}	4
Atenolol & aluminium hydroxide	2.276×10^{-6}	11
Azithromycin & hydromorphone	2.239×10^{-6}	204
Levetiracetam & aluminium hydroxide	2.149×10^{-6}	0
Dexamethasone & isosorbide	1.783×10^{-6}	329
Aluminium hydroxide & quetiapine	1.777×10^{-6}	1
Average Number of Reports		**55.7**
Correlation Between Harmonic Mean of Lift and Number of Reports		**0.920**

average equalling 55.8 reports in the group; however, most often, the number of reports in this group is close to 0. On the contrary, the list of high-risk drug combinations formed with lift and conviction metrics did not show any relevant correlation between the drug combination and the number of ADEs reported by patients (Table 2). The group of low-risk drug combinations had on average 330.5 reports, and the group of high-risk drug combinations had on average 404.8, and the correlation ratio was 0.118.

VII. DISCUSSION

The analysis of results demonstrated that the group of drug combinations identified as a high-risk group indeed had a higher number of all types of drug interactions. However, the highest contrast could be noticed in the number of major drug interactions in the high-risk group. Unlike moderate and minor drug interactions, adverse events caused by major drug interactions are clinically significant and may even lead to life-threatening conditions. Hence, having almost twice the amount of major drug interactions in the high-risk group of drug combinations seems to be related to hospital readmissions caused by adverse drug events. Moreover, analysing minor and moderate drug interactions can be complicated by the effect of positive drug interactions. Some drug interactions can beneficially affect the patient's health, such as increasing

TABLE II
CORRELATION BETWEEN THE NUMBER OF ADE REPORTS AND LIFT IN STRONG AND WEAK ASSOCIATION RULES.

Drug combination	*Lift*	*Number of reports*
Association rules with the highest risk of DDI (Highest lift)		
Pregabalin & sevelamer	3.657	14
Morphine & sevelamer	3.573	178
Fluconazole & ursodiol	3.551	1,325
Atovaquone & ursodiol	3.477	191
Acyclovir & hydroxyzine	3.399	489
Citalopram & ursodiol	3.382	189
Citalopram & phenazopyridine	3.367	198
Calcitriol & morphine	3.226	562
Clotrimazole & folic acid	3.131	391
Mycophenolate & ursodiol	3.113	551
The Average Number of Reports		**404.8**
Association rules with the lowest risk of DDI (Lowest lift)		
Clindamycin & tramadol	1.114	166
Finasteride & rivaroxaban	1.111	253
Rosuvastatin & tamsulosin	1.109	476
Ezetimibe & tamsulosin	1.108	673
Cyanocobalamin & rosuvastatin	1.102	112
Olanzapine & rivaroxaban	1.095	132
Pregabalin & rosuvastatin	1.093	597
Clindamycin & ranitidine	1.082	79
Cyanocobalamin & hydrochlorothiazide	1.075	335
Fluoxetine & pravastatin	1.066	482
The Average Number of Reports		**330.5**
Correlation Between Lift and Number of Reports		**0.118**

the effect of one another. Hence such drug interactions do not result in negative clinical outcomes. The use of the ARM approach with the new harmonic mean of lift metric demonstrated excellent results in identifying candidate DDIs not yet studied by pharmaceutical companies. Traditional lift and conviction metrics were useful in evaluating the strength of interaction between two drugs resulting in ADE. Major DDI type had substantially higher lift and conviction, indicating a stronger interaction between the drugs and ADEs. However, both metrics were not useful for detecting new potential DDIs. Higher lift and conviction scores in unknown drug interactions did not correlate with the high number of patient complaints in public clinical trials run on the eHealthMe website. On the contrary, the correlation was 0.118. In comparison, the harmonic mean of lifts score was highly correlated with the number of patients' complaints and was equal to 0.92 for top high-risk and low-risk drug combinations. Most drug combinations with the lowest harmonic mean of lifts showed 0 or close to 0 ADE reports. And most drug combinations with the highest harmonic mean of lifts showed a large number of ADE reports varying from 69 to 3,495. Such characteristics of the harmonic mean of lifts can be explained by better prioritisation of ADEs in calculating rule interestingness. Post-marketing trials can be costly and can take years to complete. It is also difficult to recruit participants for post-marketing trials, especially if the drug is not widely used. As a result, pharmaceutical companies typically only study the most common and severe drug-drug interactions in post-marketing trials. A data mining-based approach can help to prioritise and shortlist the most important candidate drug combinations for post-marketing trials. This can reduce trial costs and increase drug safety for end users. There are several limitations of this study. Since the analysis only included prescription data, other factors influencing adverse clinical outcomes, such as diseases and comorbidities impact, and demographic and lifestyle factors were not accounted for. These additional factors can be further used to derive more accurate predictions. Also, the pharmacokinetics and pharmacodynamics of the drugs can be analysed to improve the accuracy of prediction. For example, drug combinations with similar pharmacokinetics to known major DDIs can be prioritised and used as an additional evaluation metric. Such additional parameters can improve the model's performance but, at the same time, make it more complicated and resource-consuming. In future, other adverse clinical outcomes can be used as ADE signals, including in-hospital and out-of-hospital mortality, excessive length of stay in the hospital and variations of hospital readmission (7-day, 14-day). When combined with more data from electronic health records of patients, including demographic and disease-related features, stronger drug interactions could be identified.

VIII. CONCLUSION

This paper described a new approach based on association rules for discovering potential drug-drug interactions that lead to adverse clinical outcomes. The results of this study demonstrated that the ARM-based approach can identify high-risk drug combinations from EHR. Using a harmonic mean of lift metric, it is possible to identify candidate drug combinations that are highly likely to have interactions leading to adverse events. With the advance of EHR worldwide, such a simple approach can help identify potential drug-drug interactions and use this as a basis for further investigations and clinical trials. Moreover, this approach is highly customisable and can be used with any other measure of adverse clinical events. In future, these results can be enhanced by including more detailed information about the medications, doses and duration of intake, diagnosis and demographic data in the analysis.

REFERENCES

[1] NHS, "Clinical trials," NHS, 2019. https://www.nhs.uk/conditions/clinical-trials/ (accessed Jun. 12, 2023).

[2] R. L. Maher, J. Hanlon, and E. R. Hajjar, "Clinical consequences of polypharmacy in elderly," Expert Opinion on Drug Safety, vol. 13, no. 1, pp. 57–65, Sep. 2013, doi: https://doi.org/10.1517/14740338.2013.827660.

[3] M. Sheikh-Taha and M. Asmar, "Polypharmacy and severe potential drug-drug interactions among older adults with cardiovascular disease in the United States," BMC Geriatrics, vol. 21, no. 1, Apr. 2021, doi: https://doi.org/10.1186/s12877-021-02183-0.

[4] M. Glans, A. Kragh Ekstam, U. Jakobsson, Å. Bondesson, and P. Midlöv, "Medication-related hospital readmissions within 30 days of discharge—A retrospective study of risk factors in older adults," PLOS ONE, vol. 16, no. 6, p. e0253024, Jun. 2021, doi: https://doi.org/10.1371/journal.pone.0253024.

[5] L. Zerah et al. , "Prevalence of drug-drug interactions in older people before and after hospital admission: analysis from the OPERAM trial," BMC Geriatrics, vol. 21, no. 1, Oct. 2021, doi: https://doi.org/10.1186/s12877-021-02532-z.

[6] M. F. M. Chowdhury and A. Lavelli, "FBK-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information", Proc. 2nd Joint Conf. Lexical Comput. Semantics: Proc. 7th Int. Workshop Semantic Eval., pp. 351-355, 2013.

[7] S. Kim, H. Liu, L. Yeganova and W. J. Wilbur, "Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach", J. Biomed. Inform., vol. 55, pp. 23-30, 2015.

[8] S. Liu, B. Tang, Q. Chen and X. Wang, "Drug-drug interaction extraction via convolutional neural networks", Comput. Math. Methods Med., pp. 1-8, 2016.

[9] X. Sun, L. Ma, X. Du, J. Feng and K. Dong, "Deep convolution neural networks for drug-drug interaction extraction", Proc. IEEE Int . Conf. Bioinf. Biomed., pp. 1662-1668, 2018.

[10] W. Zheng et al., "An attention-based effective neural model for drug-drug interactions extraction", BMC Bioinf., vol. 18, no. 1, 2017.

[11] B. Xu, X. Shi, Z. Zhao and W. Zheng, "Leveraging biomedical resources in bi-LSTM for drug-drug interaction extraction", IEEE Access, vol. 6, pp. 33432-33439, Jun. 2018.

[12] F. Cheng and Z. Zhao, "Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic therapeutic chemical and genomic properties", J. Amer. Med. Informat. Assoc., vol. 21, no. e2, pp. e278-e286, 2014.

[13] A. Kastrin, P. Ferk and B. Leskošek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning", PLoS One, vol. 13, no. 5, 2018.

[14] C. Yan, G. Duan, Y. Pan, F.-X. Wu and J. Wang, "DDIGIP: Predicting drug-drug interactions based on Gaussian interaction profile kernels", BMC Bioinf., vol. 20, no. 15, pp. 1-10, 2019.

[15] Y. Noguchi, T. Tachi, and H. Teramachi, "Review of Statistical Methodologies for Detecting Drug–Drug Interactions Using Spontaneous Reporting Systems," Frontiers in Pharmacology, vol. 10, Nov. 2019, doi: https://doi.org/10.3389/fphar.2019.01319.

[16] D. Zhang, D. Wang, "Relation classification via recurrent neural network," arXiv:1508.01006, 2015.

[17] F. Li, M. Zhang, G. Fu, D. Ji, "A neural joint model for entity and relation extraction from biomedical text," BMC Bioinform., 18, 198, 2017.

[18] R. Harpaz, W. DuMouchel, P. LePendu, P, et al., "Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system," Clin Pharmacol Ther; 93:539–46, 2013.

[19] B. T. Thakrar, S. B. Grundschober, and L. Doessegger, "Detecting signals of drug-drug interactions in a spontaneous reports database," British journal of clinical pharmacology, 64(4), 489–495, 2007.

[20] C. Wang, X-J. Guo, J-F. Xu, C. Wu, Y-L. Sun, X-F. Ye, et al., "Exploration of the association rules mining technique for the signal detection of adverse drug events in spontaneous reporting systems," PLoS ONE 7(7): e40561, 2012.

[21] Y. Susuta and Y. Takahashi, "Safety Risk Evaluation Methodology in Detecting the Medicine Concomitant Use Risk which might Cause Critical Drug Rash," Jpn. J. Pharmacoepidemioly, vol. 19, no. 1, pp. 39–49, Jan. 2014, doi: https://doi.org/10.3820/jjpe.19.39.

[22] A. Johnson, T. Pollard, L. Shen, et al., "MIMIC-III, a freely accessible critical care database," Sci Data 3, 160035, 2016, https://doi.org/10.1038/sdata.2016.35

[23] Centers for Disease Control and Prevention, "Adverse Drug Events in Adults," Jan. 14, 2020. https://www.cdc.gov/medicationsafety/adult_adversedrugevents.html (accessed Dec. 06, 2022).

[24] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," ACM SIGMOD Record, vol. 26, no. 2, pp. 255–264, Jun. 1997, doi: https://doi.org/10.1145/253262.253325.

[25] D. S. Wishart et al., "DrugBank 5.0: a major update to the DrugBank database for 2018," Nucleic acids research, vol. 46, no. D1, pp. D1074–D1082, 2018, doi https://doi.org/10.1093/nar/gkx1037.

[26] "A - Z Drug List from Drugs.com," Drugs.com [Online]. Available: https://www.drugs.com/drug (accessed Feb. 23, 2023).

[27] M. Salman, H. S. Munawar, K. Latif, M. W. Akram, S. I. Khan, and F. Ullah, "Big Data Management in Drug–Drug Interaction: A Modern Deep Learning Approach for Smart Healthcare," Big Data and Cognitive Computing, vol. 6, no. 1, p. 30, Mar. 2022, doi: https://doi.org/10.3390/bdcc6010030.

[28] EHealthme, "eHealthMe: advanced medication management," Ehealthme.com, 2016. https://www.ehealthme.com/ (accessed Feb. 23, 2023).