# DEVES: Interactive Signal Analytics for Drug Safety

Tabassum Kakar[1,2], Xiao Qin[1,2], Andrew Schade[1], Brian McCarthy[1], Huy Quoc Tran[1], Brian Zylich[1], Elke. A. Rundensteiner[1], Lane Harrison[1], Sanjay K. Sahoo[2] and Suranjan De[2]
[1]tkakar,xqin,alschade,bmccarthy,hqtran,bzylich,rundenst,ltharrison@cs.wpi.edu
[2]tabassum.kakar,xiao.qin,sanjay.sahoo,suranjan.de@fda.hhs.gov
[1]Computer Science Department, Worcester Polytechnic Institute
[2]Center for Drug Evaluation and Research, U.S. Food and Drug Administration

## ABSTRACT

Drug-drug interaction related adverse events (DIAE) signals are a major public health issue. Drug safety analysts must sift through thousands of adverse event reports submitted daily to U.S. Food and Drug Administration (FDA) to discover unexpected DIAE signals, which if addressed can lead to life-saving actions. To facilitate the DIAE discovery from these massive data sets, we design several technological innovations that together are integrated into an interactive visual analytics system called DEVES [1]. First, our state-of-the-art DIAE mining algorithm efficiently infers potential DIAE signals from these reports, and then ranks them based on their significance score. For interpretability of these inferred DIAE signals, domain knowledge of adverse events and already known drug interactions is extracted from external authoritative data sources and then seamlessly integrated with the inferred signal set. Guided by this *augmented signal model*, DEVES supports advanced signal analytics – empowering the analyst to interact with linked visualizations offering complementary perspectives into the signal set and its supporting evidence in the form of reports. Our demonstration showcases the technological innovations of DEVES using real-world FDA datasets, demonstrating that DEVES effectively supports the core regulatory tasks from signal screening, signal prioritization to signal validation.

## KEYWORDS

Data Integration; Pattern Mining; Visual Analytics

## ACM Reference Format:

Tabassum Kakar, Xiao Qin, Andrew Schade, Brian McCarthy, Huy Quoc Tran, Brian Zylich, Elke. A. Rundensteiner, Lane Harrison, Sanjay K. Sahoo and Suranjan De. 2018. DEVES: Interactive Signal Analytics for Drug Safety. *In The 27th ACM International Conference on Information and Knowledge Management (CIKM'18), Oct. 22–26, 2018, Torino, Italy. ACM, NY, NY, USA, 4 pages.*
https://doi.org/10.1145/3269206.3269211


*CIKM '18, October 22–26, 2018, Torino, Italy*

ACM ISBN 978-1-4503-6014-2/18/10…$15.00
https://doi.org/10.1145/3269206.3269211

[1]Accessible at: http://diva.wpi.edu:3000/

## 1 INTRODUCTION

**Motivation.** Adverse events caused by drug-drug interactions (DIAEs) are a major cause of emergency room visits and hospitalizations [1]. Unfortunately, identifying all DIAEs in clinical trails is prohibitively expensive and impossible in practice – as it would require exhaustive testing of every possible drug combination on human subjects. For early detection of DIAEs not found during clinical trials, spontaneous reporting databases have been adopted in the U.S. [2] and elsewhere [3] to collect and analyze information on potential adverse events to prevent a major public health crisis.

**Research Challenges.** These reports, however, may not directly capture information about drug interactions. Hence, safety reviewers either engage in the highly demanding and error-prone task of manually sifting through these reports to attempt to put together repeated patterns in their head; or more recently, computational techniques have been applied to attempt to automatically derive potential DIAE signals [3, 7]. These machine generated signals, which serve as a hypothesis of potential undiscovered drug interactions, then must be validated by domain experts to assess their severity and validity before regulatory action can be taken. Unfortunately, such machine learning techniques are currently not used in the review process for several reasons. One, the volume of generated signals is overwhelming, often much larger than the number of reports themselves. For example, $n$ distinct drugs and $m$ unique adverse events across a set of reports result in up to $O(2^{n+m})$ signals in the worst case. Two, it is challenging to interpret the signals output by such algorithms. Therefore, an interactive end-to-end system is required that generates these signals, seamlessly integrates domain knowledge to assist the analysts in their analysis, and provides visual analytics for direct involvement of the analyst in the signal exploration and validation process.

**State-of-the-art.** In contrast, existing drug-drug interaction tools [4] [5, 9] are designed as "fact check" systems, that is, they inform a medical professional about known drug interactions to help guide the prescription process. However, these displays neither provide the means to generate and explore undiscovered potential machine-generated signal candidates, nor do they capture all core relevant information critical for sense making of these potential signals. More importantly, these systems lack the support for interactive analysis tasks core to the drug review process.

**Proposed Solution.** In this demonstration, we now tackle the above challenges in a comprehensive fashion by designing an end-to-end web-based visual analytic system for interactive DIAE signal

[2] https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm
[3]http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/index-eng.php
[4]https://www.drugbank.com

analysis (Fig. 2). The resulting system DEVES is accessible at: http://diva.wpi.edu:3000/. DEVES aims to bridge the gap between machine learning and user-friendly visual interfaces for sense-making of the mining results. DEVES first generates and then ranks DIAE signals using rule mining from FDA's spontaneous reports [7].

To make these signals useful for the drug review process, we design strategies for extracting domain knowledge such as known DIAEs and severe adverse events from external data sources with automatic ETL (Extract, Transform, and Load) techniques both for pre-mining data cleansing and for post-mining result factoring. We also design innovative interlinked visualizations empowering the safety analysts to perform their review tasks. These tasks include: (1) screening and prioritizing a particular signal out of the huge number of machine-generated DIAE signals, (2) in-depth investigation of a selected signal to determine if it indeed warrants to be promoted to be a true signal to take regulatory action by analyzing its adverse events, its associated meta information and the underlying reports all integrated within one powerful analytics platform.

## 2 THE DEVES SYSTEM

The DEVES framework in Fig. 1 consists of two main components, namely, the **Signal Miner** and the **Signal Analyzer** modules. First, each report is converted into a two-set tuple where the first set models the suspected drugs and the other the respective events. This structured information is then fed into the *Signal Generator* that uses a rule learning based DIAE inference algorithm [7] to generate non-spurious signals. The *Signal Augmenter* integrates multiple knowledge bases with the inferred signals such as severity of events and signal status, i.e., whether the signal is known or not. The augmented DIAE signals are then stored into *Signal Database*. *Analytics Executor* executes the data exchange requests from the offline Signal Miner. It also supports the online DEVES interactive visual analytics. Results from frequent interactions are *cached* so that real time user experience is achieved. The DEVES *Signal Analyzer* is a web-based interface consisting of innovative coordinated views that assist analysts in exploring and validating the augmented signals on various levels of abstraction.

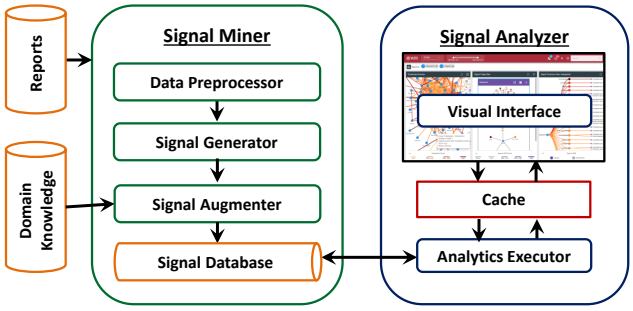


**Figure 1: The DEVES framework.**

## 3 DEVES TECHNICAL INNOVATIONS

Technical innovations for signal analytics are described below.

### 3.1 Non-Spurious Signal Generation & Ranking

FDA adverse event reports contain information about patients demographics, the drugs being taken and the adverse events being observed, along with a text narrative explaining the details of the events. For signal generation using rule mining, we utilize the structured information consisting of drugs and events to find the most significant drug pairs leading to adverse events. These signals are essentially rules with only drug combinations as antecedents and adverse events as consequents. For example, the rule (Aspirin, Warfarin ⇒ *bleeding*) states that taking Warfarin and Aspirin together can lead to bleeding. However, examining the huge number of potential DIAE signals generated by a standard rule mining algorithm can be prohibitively time-consuming. We thus developed an algorithm for not only generating non-spurious signals but also establishing a ranking for the signals based on their likelihood to be promising using the *contrast measure* published in KDD [7].

Non-spurious signals are true representative of the reported drugs and events in the reports and the mining algorithm avoids generation of misleading signals conveying partial associations that are not fully backed by actual reports. The intuition for the *contrast measure* is that if adverse events are triggered by the drug-drug interaction, then they should have less chance of being triggered by any of the individual drugs in the DIAE signal. This contrast score highlights the most statistically significant signals in DEVES. DIAE signals with higher scores have a higher chance of being true signals warranting action, therefore are recommended by DEVES with higher priority to the analysts. Details on non-spurious signal generation and the contrast measure can be found in MARAS [7].

### 3.2 Domain-Based Signal Augmentation

The machine generated signals are merely association rules representing drug pairs and adverse events. These mining techniques [3, 7] are only truly useful when their output can be leveraged during the drug review process to take appropriate actions to cope with potential signals. However, without proper enhancements, the non-technical domain experts barely can interpret let alone use these signals to form a hypothesis. To enhance the interpretability and help analysts make sense of the generated signals, we integrate domain knowledge driven through our interviews with domain experts into these signals. One such domain specific data element is the severity of the adverse events, as some adverse events such as heart attack or kidney failure are more critical then headache or nausea. For this, we leverage the list of designated medical events (DME) used within the FDA to prioritize the severe events during signal detection.

Second, the goal of analysts is to find unknown or novel signals, so that they can take regulatory actions to avoid future incidents. Therefore, identification of already previously known signals is crucial for the drug safety review. To achieve this, we extract information about known signals from the Twosides database [8] and then integrate it with our generated signals to form the augmented signal space. However, different medical naming conventions and terminologies make the integration non-trivial. For instance, due to different adverse event terminologies followed by spontaneous reports and Twosides database, a generated signal with the same drug pair and event may not match. For example, the inferred signal Aspirin, Metoprolol ⇒ *Myocardial Infarction* from DEVES does not match with a known signal form Twosides database Aspirin, Metoprolol ⇒ *Heart Attack*. Although, they both refer to the same fact. This problem is addressed in DEVES by mapping the adverse events from both sources to unified standardized terms using the

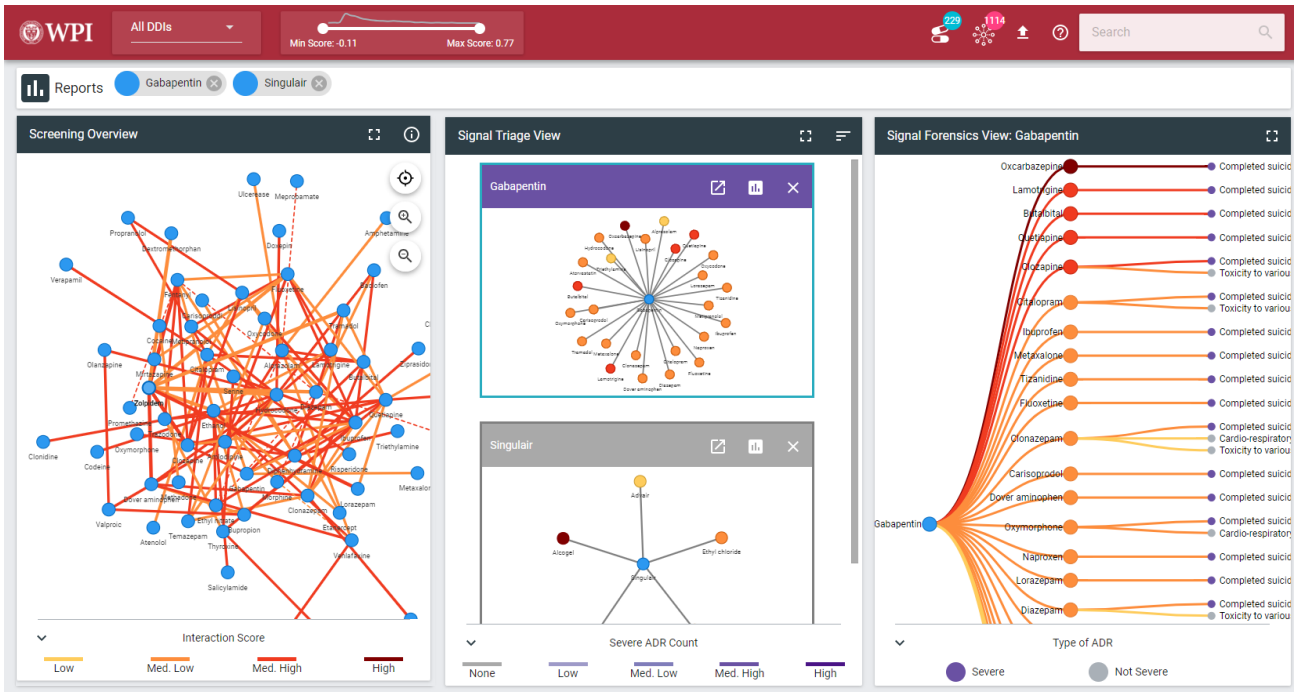


**Figure 2: The user interface of DEVES.**

Medical Dictionary for Regulatory Activities (MedDRA) [5] *preferred term* encodings. In rare cases, where an event can have multiple *preferred terms*, a concensus was made to keep both in the signal by differentiating them by a marker (*).

Similarly, different variants of a drug in the generated signals is another challenge during the integration. For example, "Diazepam" and "DiazepamInjectable" both refer to the same drug, however, with different modes of administration; while "zerit15mgoralcapsule" and "zerit20mgoralcapsule" refer to the same drugs with different dosages. To tackle this, we develop a variant of the *edit distance* [2] algorithm that utilizes the number of edit operations required to transform one string (drugname) into the closest string (drugname) to determine if both drug names can be equated with each other. Constraints are added to limit the number of operations, in cases where the operations might turn a drugname into a different drug using the drugname lexicon. We used the RxNorm [6], a dataset of prescribable drugs managed by the National Library of Medicine as drugname lexicon for mapping names using the edit distance strategy. Using this technique, we got an accuracy of 89.6% on a set of hundred randomly selected drugs after applying this disambiguation technique.

### 3.3 Signal Visual Analytics Techniques

DEVES signal analytics (Figure 2) follows a human-in-the-loop strategy [4], designed based on formative interviews conducted with domain experts at the FDA. DEVES follows the principals of mixed-initiative, where the system and user work collaboratively to meet the user's goals [4]. Existing techniques to visualize association rules [6] are mainly semantics-based, i.e., exploration based on the antecedents or consequents. However, these systems do not focus on tasks related to the investigation of rules to be used to take decisions that have critical impact such as banning a drug from the market.

DEVES 's innovative displays (Fig. 2) support the major drug review tasks: (1) signal screening, (2) signal prioritization, (3) signal forensics, and (4) signal validation. The top panel (Fig. 2) contains different controls to interactively drill down to a set of drugs or signals by searching, filtering or sorting signals based on their significance score [7], or their status being known or unknown.

The Screening panel (Fig. 2-Left) uses a network diagram to display the generated signals. From the screening view, the analyst can interactively screen the drugs of interest to explore further. The Triage panel (Fig 2-Middle) displays the interactions of individual drugs with options to be filtered based on severe events or highly scored signals. It can help an analyst prioritize drugs from a set of hundred possible drugs for further analysis. The Forensics panel (Fig. 2-Right) displays details of all signals related to a selected drug including the adverse events. Finally, the Reports panel (Fig 4) displays the reports used to extract a selected signal. The analyst can easily access and analyze the reports to see if the signal makes sense, or if there are additional factors such as other drugs taken by the patient that might be causing the events. This visual analytics approach completely revolutionize drug safety review.

## 4 DEMONSTRATION SCENARIOS

In our demonstration, the audience will be able to directly interact with DEVES via a visual web interface to understand how DEVES assists the analysts in signal review practices. The case study below will also be shown to demonstrate how DEVES empowers analysts to interpret the inferred DIAE signals and validate them.

**Data Loading and Signal Generation.** With the click of a button on the control panel, adverse event reports are loaded into

[5] https://www.meddra.org/
[6] https://www.nlm.nih.gov/research/umls/rxnorm/

DEVES. The data cleansing and extraction processes are then triggered. Once completed, the signal generation module on the server side produces the DIAE signals.

**Signal Augmentation.** Thereafter, for each signal generated by the DIAE signal generator, DEVES extracts its status from the TwoSide database. These inferred signals are then augmented with their status, such as, known or unknown and stored in the DIAE signal database (Fig. 1).

**Making Sense of the Inferred Signal Space.** To offer an overview of the inferred DIAE signals, the screening view *right panel* (Fig. 2) displays the drug-drug interactions using a network view. This gives a quick glimpse of the inferred signals. The analysts can screen the interesting signals to analyze them further by filtering based on significance score and novelty from the *control panel* (Fig. 2-top). Options to zoom-in and out are provided. Clicking on the node or edge also adds the respective report tabs, which are helpful in retrieving the evidential reports corresponding to a drug (node) or signal (edge). Looking at high scored and unknown signals, the analyst adds the drugs Lansoprozole and Singulair to her Triage view by clicking on the nodes.

**Prioritizing a Drug to Analyze its Signals.** The Triage view *middle panel* (Fig. 2) uses a micro-graph display to help examine a drug related interactions in isolation. The goal of this view is to help prioritize drugs based on number of interesting associated signals. Option of sorting based on multiple criteria is provided. After sorting the selected drugs Lansoprazole and Singulair based on severity and significance score, she finds Lansoprazole having high score severe signals. The analyst wants to know about the adverse events of signals associated with Lansoprazole.

**Honing in On Specific Drug Related Signals.** To further evaluate Lansoprazole, the analyst zooms-in by clicking the zoom button on the micro-graph (Fig. 2) to analyze its corresponding signals including the adverse events using the Forensics view (Fig. 3) She notices that interaction with Digoxin is significant and currently unknown yet results in a severe adverse event 'acute kidney injury' depicted by purple color. The analyst is curious and wants to analyze the raw reports used to generate this signal to find if there are any other factors that are causing the severe event.


**Figure 3: The Forensics view of Lansoprazole.**

**Drilling Down to Evidence Exploration.** The analyst wants to examine the reports related to the Digoxin and Lansoprozole signal. By clicking on the report tab (Fig. 2-top), the corresponding reports are retrieved in the Reports view (Fig. 4). On top (Fig. 4) a histogram-based distribution of drugs reported along with Digoxin and Lansoprazole is displayed. By clicking on the first bar representing the most frequent drug, the rows with the drug 'Furosemide' are highlighted. She adds, "Furosemide is used to treat kidney disorders, the patients who were taking Lansoprozole were also taking Furosemide". That means, they might already be having a kidney disorder and Furosemide was prescribed to them for treatment. She annotates the signal as co-occurence to avoid its repeated examination in the future. The displays and interactions simplified this exploration and investigation of the inferred signals.

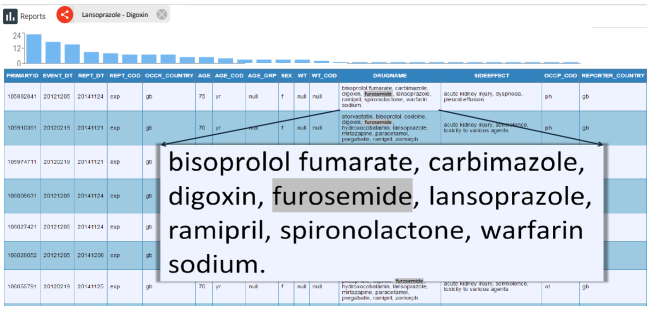


**Figure 4: The Reports View to validate DIAE signal related to Digoxin and Lansoprazole.**

## 5 CONCLUSION

In this demonstration, we showcase DEVES , an end-to-end analytics paradigm that combines computational techniques and visualization in a user-driven manner to tackle the drug safety review problem. DEVES revolutionizes how a safety reviewer approaches signal screening by recommending the most significant signals to the analyst's attention, which they can interactively examine and validate within context supported by critical domain knowledge integrated into the signal store.

This DEVES opens new opportunities for research in machine learning and visual analytics. As the analyst is the main driver of the review process, integrating analysts' feedback into the rule mining algorithm promises to improve the quality of the generated signals, particularly where the interacting drugs belongs to the same disease class. The integration of additional external knowledge both as pre-processing for signal augmentation and data cleaning as well as post-processing to support in-context literature review during signal validation are challenging open problems.

## ACKNOWLEDGEMENTS

The authors are grateful for partial support from Oak Ridge Institute for Science and Education (ORISE) managed for the U.S. Department of Energy (DOE) by Oak Ridge Associated Universities (ORAU) and National Science Foundation IIS-1815866.

## REFERENCES

[1] Matthijs L Becker, Marjon Kallewaard, Peter WJ Caspers, Loes E Visser, Hubert GM Leufkens, and Bruno HCh Stricker. 2007. Hospitalisations and emergency department visits due to drug–drug interactions: a literature review. *Pharmacoepidemiology and drug safety* 16, 6 (2007), 641–651.

[2] Mengmeng Du. 2005. Approximate name matching. *NADA, Numerisk Analys och Datalogi, KTH, Kungliga Tekniska Högskolan. Stockholm: un* (2005), 3–15.

[3] Rave Harpaz, Herbert S. Chase, and Carol Friedman. 2010. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics* 11, S-9 (2010), S7.

[4] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 159–166.

[5] Michael Kuhn, Christian von Mering, Monica Campillos, Lars Juhl Jensen, and Peer Bork. 2008. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Research* 36, suppl 1 (2008), D684–D688.

[6] Xiaotong Liu and Han-Wei Shen. 2016. Association analysis for visual exploration of multivariate scientific data sets. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 955–964.

[7] Xiao Qin, Tabassum Kakar, Susmitha Wunnava, Elke A. Rundensteiner, and Lei Cao. 2017. MARAS: Signaling Multi-Drug Adverse Reactions. In *Proceedings of the 23rd ACM SIGKDD*. 1615–1623.

[8] Nicholas P Tatonetti, P Ye Patrick, Roxana Daneshjou, and Russ B Altman. 2012. Data-driven prediction of drug effects and interactions. *Science translational medicine* 4, 125 (2012), 125ra31–125ra31.

[9] Joachim Von Eichborn, Manuela S Murgueitio, Mathias Dunkel, Soeren Koerner, Philip E Bourne, and et al. 2011. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Research* 39, suppl 1 (2011), D1060–D1066.