








Language-agnostic pharmacovigilant text mining to elicit side effects from clinical notes and hospital medication records

Benjamin Skov Kaas-Hansen^{1,2,3}  | Davide Placido²  |
Cristina Leal Rodríguez²  | Hans-Christian Thorsen-Meyer⁴  |
Simona Gentile⁵  | Anna Pors Nielsen² | Søren Brunak² | Gesche Jürgens¹  |
Stig Ejdrup Andersen¹ 

¹Clinical Pharmacology Unit, Zealand University Hospital, Roskilde, Denmark

²NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

³Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

⁴Department of Intensive Care Medicine, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark

⁵Department of Radiology, Zealand University Hospital, Roskilde, Denmark

Correspondence

Benjamin Skov Kaas-Hansen,
Munkesøvej 18, 4000 Roskilde,
Denmark.
Email: zqw524@ku.dk

Funding information

The authors would like thank Innovation Fund Denmark (5153-00002B) and the Novo Nordisk Foundation (NNF14CC0001 and NNF17OC0027594) for their financial contribution to BigTempHealth without which this study had not been possible. The funders played no role in designing, conducting, interpreting or reporting this study.

Abstract

We sought to craft a drug safety signalling pipeline associating latent information in clinical free text with exposures to single drugs and drug pairs. Data arose from 12 secondary and tertiary public hospitals in two Danish regions, comprising approximately half the Danish population. Notes were operationalised with a fastText embedding, based on which we trained 10 270 neural-network models (one for each distinct single-drug/drug-pair exposure) predicting the risk of exposure given an embedding vector. We included 2 905 251 admissions between May 2008 and June 2016, with 13 740 564 distinct drug prescriptions; the median number of prescriptions was 5 (IQR: 3–9) and in 1 184 340 (41%) admissions patients used ≥ 5 drugs concomitantly. A total of 10 788 259 clinical notes were included, with 179 441 739 tokens retained after pruning. Of 345 single-drug signals reviewed, 28 (8.1%) represented possibly undescribed relationships; 186 (54%) signals were clinically meaningful. Sixteen (14%) of the 115 drug-pair signals were possible interactions, and two (1.7%) were known. In conclusion, we built a language-agnostic pipeline for mining associations between free-text information and medication exposure without manual curation, predicting not the likely outcome of a range of exposures but also the likely exposures for outcomes of interest. Our approach may help overcome limitations of text mining methods relying on curated data in English and can help leverage non-English free text for pharmacovigilance.

KEYWORDS

data mining, machine learning, pharmacovigilance, safety signal detection, safety signal refinement

[Correction added on 8 August 2022, after first online publication: Author Benjamin Skov Kaas-Hansen's email address has been amended and author Stig Ejdrup Andersen's ORCID ID has been added.]

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Basic & Clinical Pharmacology & Toxicology* published by John Wiley & Sons Ltd on behalf of Nordic Association for the Publication of BCPT (former Nordic Pharmacological Society).

1 | INTRODUCTION

Pharmacovigilance usually operates with two qualifications of the common term *side effect*: adverse drug events (ADEs) and adverse drug reactions (ADRs). ADEs are (noxious) medical events occurring while using medicines without assuming causal relationships.¹ ADRs are subsumed by ADEs and constitute outcomes believed or known to be *caused* by exposure to a given medicinal product.^{2,3} ADRs are usually classified in six groups, including *dose-related* and *not dose-related*.⁴ The latter are more unpredictable than the former and tend to be unrelated to the pharmacological effect, making them interesting from a safety signal detection perspective.

ADR signal detection usually revolves around spontaneous case reports, collated nationally (e.g., Danish Medicines Agency) and internationally (e.g., EudraVigilance of the European Medicines Agency and VigiBase of the Uppsala Monitoring Centre⁵). This system suffers from several shortcomings, including the inherit filtering of reports making it into central databases, causing i.a. under-reporting⁶⁻⁹ that may even be biased or otherwise influenced by, for example, media hype or legislation.¹⁰ These weaknesses, and the ever-expanding digitisation of patient data, have sparked much interest in leveraging complementary data sources and technologies for pharmacovigilance, including longitudinal clinical data and natural language processing (NLP), the branch of machine learning for making textual data compatible statistical modelling.^{11,12}

Text mining uses NLP methodologies to extract structured information from inherently unstructured textual data. Its applications in pharmacovigilance often hinge on hand-curated reference sets for named-entity recognition or entity extraction¹³⁻¹⁶; for example, previous work brought about a Danish dictionary of side effects.¹⁷ These tasks focus on assigning labels to free-text terms so they can be codified and used as structured data akin to diagnostic codes recorded in national registers¹⁸ or adverse-event databases.

Creation and maintenance of such gold standards are costly and tedious, which likely explains the limited availability of tools and resources (including corpora) for non-English textual data. For example, the official ADR vocabulary of the Danish Medicines Agency is MedDRA (Medial Dictionary for Regulatory Activities, in English), and submitters of case reports are encouraged to pick from English terms when submitting case reports. When non-standard side effects are entered, these are manually mapped to the English MedDRA afterwards. Thus, it is near-impossible to extract information across languages which would be useful for pharmacovigilant purposes.

We posit that, to leverage clinical free text, complementing existing vocabulary-based approaches to pharmacovigilant NLP with (semi-)automatic information extraction from clinical free text deserves exploration and could facilitate vast screening of clinical free text.

To this end, we report on the creation of one such complementary system: an end-to-end machine learning pipeline associating latent information in clinical free text with medication profiles to highlight potential adverse drug reactions to single drugs and drug pairs. We envision a system that accepts one of several free-text side-effect terms from the user and returns likely prominent exposures to undergo assessments akin to the evaluation of signals in spontaneous case reports.

2 | METHODS AND MATERIALS

Data were obtained from electronic patient record (EPR) systems of 12 secondary and tertiary public hospitals in two Danish regions (Capital Region and Region Zealand), comprising approximately 2.6 million persons (about half the Danish population). We used data from a random sample of 500 000 adult (age ≥ 18 years) patients admitted between 1 January 2006 and 30 June 2016.

The full analytic workflow is depicted schematically in Figure 1 and has five main components (detailed below): deriving doorstep medication profiles (red), training the embedding model (brown), operationalisation of clinical notes (blue), training the signal detection component (green) and evaluating the safety signals (purple).

2.1 | Doorstep medication profiles

We considered only pre-existing medication at start of admission and created one medication vector with one element per distinct single drug and drug pair in the full data set, using their respective anatomical therapeutic chemical (ATC) codes.¹⁹ Medication data were extracted from the electronic patient files and, so, reflect what was registered by a physician at time of admission. Elements corresponding to single drugs and drug pairs used by a given patient at doorstep were set to 1, the rest to 0. We only considered single drugs and drug pairs used in at least 1000 admissions.

2.2 | Embedding model

An embedding packs high-dimensional data into much fewer dimensions. Imagine, for example, one-hot-encoding words^{12,20} in a corpus of clinical notes that

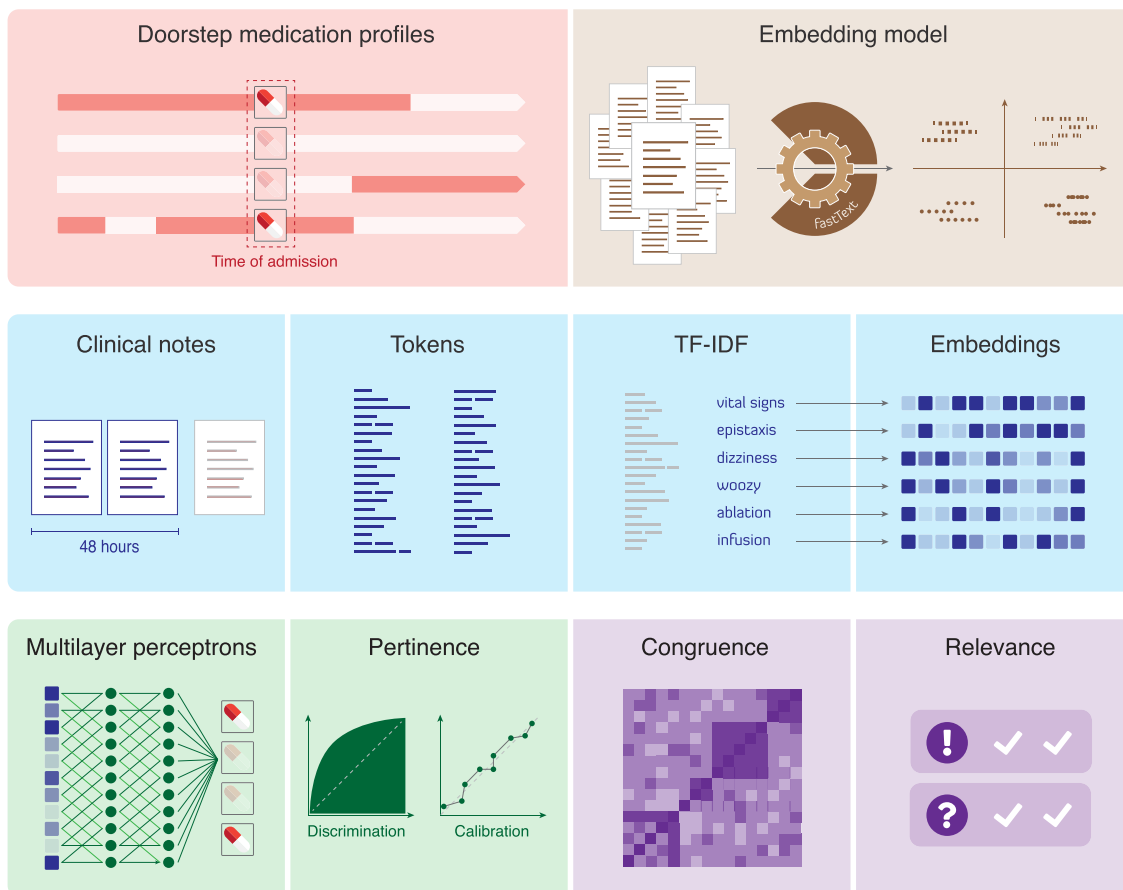


FIGURE 1 Schematic illustration of the end-of-end pipeline; see sections with corresponding headings in main text for details: The blue areas correspond to *operationalisation of clinical notes*, the green to *training the signal detection component* and the purple to *evaluating the safety signals*. The red and blue areas illustrate data capture from a single patient.

collectively contain 345 671 unique words: The presence of a word in a given note could be represented by a (very sparse) vector with 345 670 zeros and a single 1. Learning a 100-dimensional embedding of the words, in contrast, enables us to represent each word by a 100-element vector that also captures latent information in unstructured text.^{12,20} This vector will not be sparse (computationally convenient) and vectors of words with similar meanings will be similar even when lexicographically different (e.g., *headache*, *sore head* and *neuralgia*). Our embedding used tokens (one or several words together that collectively make up a term such as the three terms in *italic* in the previous sentence) and not only single words. See the supporting information for more detailed explanations.

We used fastText²¹ to train the embedding model on the full corpus after slight pruning: Characters other than letters and numbers were removed, as were multiple white spaces. This yielded one white-space separated string of words from each note. Hyperparameters were arbitrary but appropriate for the task at hand; for example, we used a 256-dimensional embedding, sub-word components were allowed to be between three and six

characters long (*minn* and *maxn* fastText settings; ‘dys’ and ‘tonia’ are two sub-word component examples of ‘dystonia’) and tokens were allowed to span up to three words to capture multi-word signals (such as *chest pain* or *sore head*; *wordNgrams* fastText setting; *N*-grams are tokens that consist of *N* words, where *N* is usually one, two and/or three: ‘tremor’ is a unigram, ‘idiopathic tremor’ a digram, and ‘intermittent dystonic tremor’ a trigram). All settings can be found in the analytic code; see below.

2.3 | Operationalisation of clinical notes

The corpus comprised notes recorded within the first 48 h of admission; each note underwent five processing steps. First, the note was split into sentences. Second, within each sentence, we identified negations and for each of these excluded the subsequent five words or until end-of-sentence (heuristic based on Thomas et al.²²). Third, we removed special characters from these non-negated words. Fourth, we retained the pruned words

that were neither Danish stop words (using *nlTK.corpus*²³) nor present in an in-house list of almost 430 000 names used in Denmark. We forewent stemming and lemmatisation¹² to let the model learn from natural words, to facilitate its downstream use (stemming and lemmatisation harmonise the corpus by transforming the words therein to [usually] shorter versions, i.e., their stems and lemmas). Finally, these retained tokens were concatenated by admission, essentially considering each admission one document (an oft-used term in text-mining and information retrieval literature).

We computed the term-frequency/inverse-document-frequency (TF-IDF) as $tf \times \log(N/[1 + df])$ for retained tokens with $10 \leq df \leq 50\,000$ to omit tokens so common or rare that they unlikely contained information of interest.²⁴ The final TF-IDF values were not used to discard tokens at this step; that happened during training; see below.

The final step of this component was converting tokens to their corresponding embedding vectors using the fastText model. This happened during training to not unnecessarily store vectors for tokens many of which were never used due to under-sampling, see below.

2.4 | Training the signal detection component

We constructed one multilayer perceptron (MLP, also called feed-forward neural network) model with two hidden layers of 256 nodes for each of the 12 270 unique drugs and drug pairs in the medication profiles, setting the binary outcome to 1 if that drug (pair) was in the doorstep medication profile and 0 otherwise. Because of the imbalanced nature of the prediction task (Figure S1) and to obtain tolerable runtime, we used random 1:2 under-sampling of the majority class to help the model focus on pertinent signals. We used all tokens for cases and the top 50 tokens based on TF-IDF for controls. Then, the embedding vector for each token and its outcome became one observation for training the MLP model.

We used sigmoid activation functions, the Adam optimiser and regularisation only in the form of early stopping based on area under the receiver operating characteristic curve (AUROC) in the internal validation set. The validation set came about by 80/20 random split-sampling, deemed appropriate as this served solely for regularisation and not validation per se.²⁵

Pertinence was operationalised as signals from well-performing models with respect to discrimination and calibration-in-the-small using the internal validation set. Discrimination was gauged by AUROCs, calibration-in-the-small by the intercepts and slopes of linear

regressions to the calibration curves of decile-binned predicted probabilities and corresponding bin-wise observed outcome proportions.²⁶ Only models with intercepts in $[-0.05, 0.05]$, slopes in $[0.95, 1.05]$ and AUROCs ≥ 0.7 in the validation sets were considered to yield pertinent signals.

2.5 | Evaluating safety signals

2.5.1 | Congruence

To quantify the relevance of the signals, we compared the predicted odds with the odds in the background population and used these odds ratios as the signal scores.

The congruence analysis served to qualitatively assess whether tokens with near-identical or very similar clinical meanings ('clinical cousins') were assigned the same medication profiles regardless of lexicographical (dis)similarity. To this end, we used the terms in Figure 4 (their origin is explained below) and a list of clinical cousins for a total 116 terms. Congruence was, then, assessed visually by plotting pairwise adjusted cosine distances^{24,27} between the signal profiles of all 116 terms, constructed as the union of all exposures in the top 50 of any of the terms.

2.5.2 | Relevance

We used a reference set to gauge the signals' relevance, that is, to what extent signals are meaningful from a clinical and pharmacovigilance point of view. From the several potential reference sets that exist,²⁸ we chose the items in the UKU (Udvalg for Kliniske Undersøgelser, English: Committee for Clinical Investigations) side effect rating scale.²⁹

We manually reviewed the top 5 single-drug and top 5 drug-pair signals for each reference-set term consulting three standard sources in clinical pharmacology, in Denmark: www.pro.medicin.dk (side effects; identical side-effect information as the official Danish summaries of product characteristics [SPCs, available at www.produktresume.dk] with few exceptions), DrugBank (drug-drug interactions; publicly available information; www.drugbank.ca³⁰) and the Danish Interaction Database³¹ (drug-drug interactions). We crafted a helper R package (promedreadr, doi: [10.5281/zenodo.5529817](https://doi.org/10.5281/zenodo.5529817)) to do the heavy lifting when collecting side-effect information from www.pro.medicin.dk. DrugBank kindly made their data (v5.1.8) available to the first author for the purpose of this study.

Each single-drug signal was labelled, in this order, as (a) example of protopathic bias or bias-by-indication,³²

(b) known side effect if reported for at least one product with that ATC code, (c) possible side effect (i.e., biologically plausible) or (d) spurious signal. For drug-pair signals, we labelled each drug according to the single-drug classification and further evaluated the signal from a drug-drug interaction point of view on two axes: whether the two drugs are known to interact (is any interaction described in the Danish Interaction Database and/or DrugBank?) and relevance of signal (three options: known result of interaction, possible result of interaction or not caused by interaction). BSKH, GJ and SEA undertook signal assessment: Each signal was evaluated independently by two assessors, and disagreement (quantified by Cohen's kappa³³) was resolved by consensus.

2.6 | Ethics

This study is part of the BigTempHealth research programme for which approval was granted by the Danish Patient Safety Authority (3-3013-1723, then competent authority for ethical approval), the Danish Data Protection Agency (DT SUND 2016-48, 2016-50, 2017-57) and the Danish Health Data Authority (FSEID 00003724). This report honours relevant items of the RECORD statement.³⁴

3 | RESULTS

The final data set covered the period from 18 May 2008 through 30 June 2016 and comprised 2 905 251 inpatient visits (admissions) of which 1559 685 (54%) were of women. The median age was 58 years (inter-quartile range, IQR: 33–73) and stable throughout the study period. These admissions comprised 10 788 259 clinical notes (18% of these patients' 60 960 247 notes) recorded within 48 h of admission and 13 740 564 doorstep drug prescriptions; the median number of doorstep-profile prescriptions was 5 (IQR: 3–9) and in 1 184 340 (41%) admissions patients used ≥ 5 drugs concomitantly, a common polypharmacy threshold.³⁵ Pruning and filtering left 179 441 739 tokens (per-admission median: 51 [IQR: 29–80]) for training the 10 270 neural-network models of which 3945 (38%) yielded pertinent signals (see Figure S2).

Figure S1 shows the relative frequency of all 571 single-drug exposures and (correspondingly) the top 571 drug-pair exposures. The dominant drug classes were those affecting the nervous system (N, including psychiatric drugs), the alimentary tract and metabolism (A) and the cardiovascular system (C). The same picture emerged from the drug-pair exposures: The most prevalent drug pairs involved these same three drug classes (e.g., AA, AC and AN).

We devised so-called fingerprints for each main UKU term visualising single-drug exposures (Figure 2). These fingerprint plots illustrate that general or vague terms (e.g., depression, nausea and weight gain) are relatively strongly associated with many drug exposures (many wedges in the inner circle are dark) and that fewer drugs, of appropriate drug classes, light up for more specific terms (e.g., amenorrhoea/galactorrhoea and tremor/dystonia/parkinsonism). Also, fingerprints of clinically related terms (e.g., tremor, parkinsonism and dystonia) are similar but clearly distinct from those of other terms.

3.1 | Congruence

We hypothesised that signal profiles of clinical cousins would be similar regardless of lexicographical (dis)similarity. Indeed, as Figure 3 illustrates, signal profiles agreed within UKU terms, within UKU domains and within the mental-neurological spectrum. As expected, the terms in the *Other* domain did not agree well, likely because this domain comprises very different side effects not fitting in elsewhere. Agreement was imperfect, which can be seen from, e.g., the light stripes representing terms with signal profiles distinct from all other terms. Several UKU terms have synonyms identical to those of other UKU terms so these will of course show perfect congruence, even if across UKU domains.

3.2 | Relevance

Agreement between the three assessors (BSKH, GJ and SEA) was moderate, with four values of Cohen's kappa (κ): relevance of drug 1 ($\kappa = 0.49$), relevance of drug 2 ($\kappa = 0.72$), whether the two drugs were known to interact in any way ($\kappa = 1.0$) and relevance of interaction ($\kappa = 0.73$); see pairwise κ values in Figure S3.

The consensus assessments in Figure 4 show that the method picked up pertinent information. There were 345 single-drug/potential-reaction pairs (Figure 4, caption). Of these, 28 (8.1%) represented possible relationships between drug exposure and the reaction in question (Figure 4B, light green). For 186 (54%) signals, the reactions were either possible, known or due to protopathic or indication bias, all clinically meaningful relationships (Figure 4B, green and dark grey). Sixteen (14%) of the 115 drug-pair signals were possible interactions; two (1.7%) were known and the rest not attributable to the drugs interacting (Figure 4C). Table S1 contains a selection of clinically interesting signals of possibly undocumented relationships between exposures and reactions.

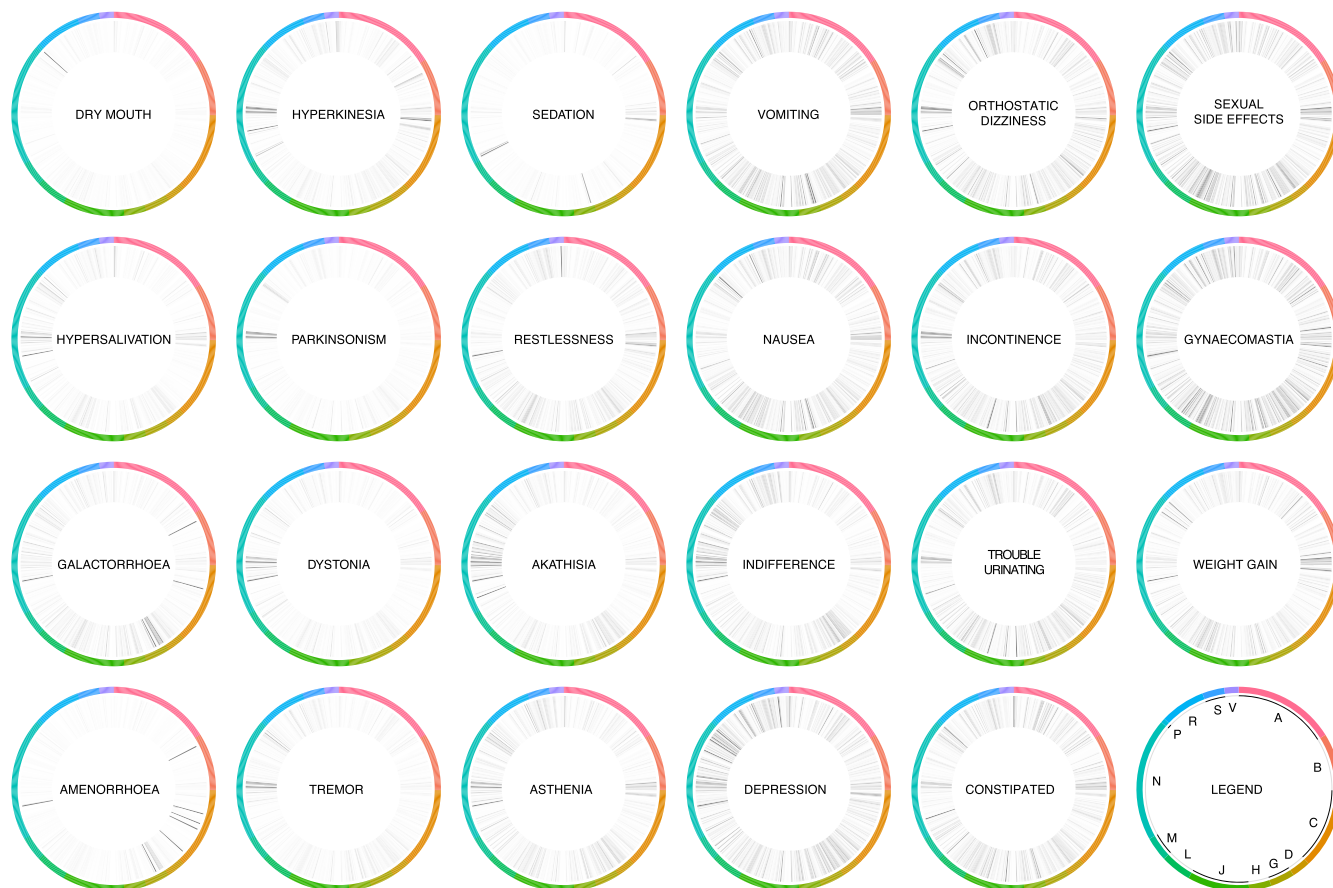


FIGURE 2 Fingerprint plots of the 23 main UKU terms and their 571 single-drug signals. Inner circles: Each wedge represents one drug and transparency the signal score. Outer circles: Colours represent anatomical drug classes (ATC level 1); see legend. A, alimentary tract and metabolism; B, blood and blood forming organs; C, cardiovascular system; D, dermatologicals; G, genito-urinary system and sex hormones; H, systemic hormonal preparations, excluding sex hormones and insulins; J, anti-infectives for systemic use; L, antineoplastic and immunomodulating agents; M, musculo-skeletal system; N, nervous system; P, antiparasitic products, insecticides and repellents; R, respiratory system; S, sensory organs; V, various [Correction added on 8 August 2022, after first online publication: Figure 2 has been corrected.]

4 | DISCUSSION

With a novel, language-agnostic approach using word embeddings, we successfully built an end-to-end machine learning pipeline to elicit potential side effects of out-of-hospital drug exposure; the method may well complement existing safety signal detection and refinement. Using side effects from the psychiatric domain with (somewhat) well-defined pharmacological properties, we illustrated that this method may offer genuine utility: manual review of signals for clinically relevant side effects illustrated the ability of the pipeline to highlight pertinent signals, with the ‘hit rate’ in the same order of magnitude as that of signal detection in spontaneous case reports.³⁶

The novelty of our approach hinders direct comparisons with the published literature. Indeed, we try to fill a gap in the three-axis categorisation of pharmacovigilance NLP: using non-English text, overcoming the reliance on

annotated data and leveraging EHR data. The number of published NLP applications in pharmacovigilance is growing: a review from 2012 included but seven studies, most of which used either simplistic keyword searches or more elaborate NLP methodologies (MediClass and MedLEE), predominantly in discharge summaries with relatively old data (1995 through 2008).³⁷ More recently, a review from 2017 included 48 studies and emphasised the need for side-effect detection methods to handle also polypharmacy-related side-effects,³⁸ an issue intimately related to drug-drug interactions.

Side-effect signal detection generally occurs in three types of data (spontaneous case reports, online forums including social media and longitudinal patient data) with the analytical approaches somewhere along two axes (modelling complexity and structuredness of the data). The long-standing signal detection in spontaneous case reports rests on several large database (e.g., FAERS, EudraVigilance and VigiBase) collecting reports from



FIGURE 3 Mean-adjusted cosine similarities between signal pairs. Rows and columns show pairwise similarities between signal profiles for specific terms. Dark blue squares signify agreement between blocks of terms (red represent disagreement). Black and white margin bars represent UKU side-effect terms, and columns/rows within the span of one bar are synonyms. The cosine similarity of two identical signals equals 1 (e.g., the diagonal). See the supporting information for more detailed explanation.

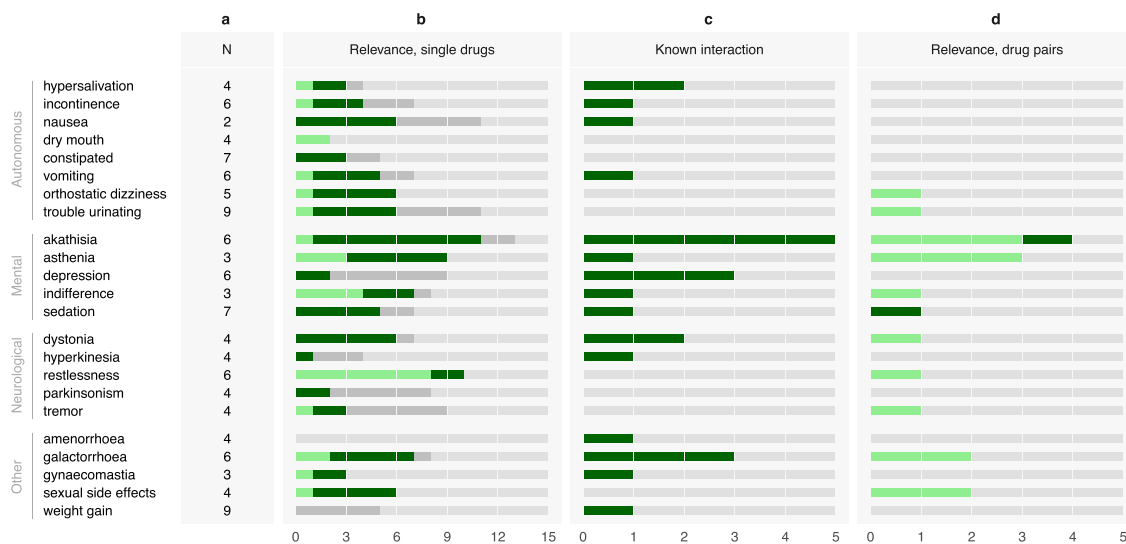


FIGURE 4 Main UKU terms by domain. (A) The number of terms used in congruence analysis (total = 116). (B) All 345 single-drug assessments (23 terms \times 5 single-drug signals = 115; 23 terms \times 5 drug-pair signals \times 2 drugs per pair = 230). Light green: undocumented reaction possibly caused by single-drug (B) or drug-pair (D) exposure. Dark green: known reaction (B + D) or interaction (C). Dark grey: protopathic or indication bias. Light grey: spurious signal. Horizontal scales in panels B–D are counts.

healthcare staff, patients and pharmaceutical companies across the globe. The mainstay of this system has been disproportionality analytic³⁹ with attempts at assessing DDIs,⁴⁰ although NLP applications exist.^{41–44} Several

attempts at leveraging online content for pharmacovigilance have come about,^{45–49} especially using Twitter posts^{50–59} with examples of trying to disentangle temporality of exposure-event pairs.⁶⁰

Although pharmacovigilant text mining in non-English corpora is not the norm, examples do exist. A Danish dictionary of side effects was created and used for mining psychiatric patient files, relying on ontologies against which terms found in the clinical text were compared^{17,61,62} and, thus, different in scope than ours. Oroz et al.⁶³ sought to create a gold standard from EMR notes in Spanish that had been annotated by pharmacologists and pharmacists, with particular focus on medicines and diagnoses, while Segura-Bedmar and Martinez⁶⁴ sought to extract drug effects, both beneficial and noxious, from a Spanish online health forum. Another study used Japanese online platforms to evaluate basic characteristics of medicine users,⁶⁵ and Ujiie et al.⁶⁶ used medical articles, manually annotated by a medical engineer, in Japanese articles published for post-marketing surveillance. Usui et al.⁶⁷ devised a system to automatically assign ICD-10 codes to Japanese free-text patient complaints recorded by pharmacists when dispensing prescription medicines.

These examples all share the foundational characteristic that they rely on curated ontologies for annotating their corpora. This eases evaluation as the curation process establishes a ground truth against which to compare the algorithm's output. Nevertheless, real-life clinical corpora are moving targets, and the constant expansion and morphing of ontologies require continual and costly updating of annotation rules. Our approach stands in contrast to this: It is an end-to-end pipeline that requires no annotation of specific documents but acts a simple signal detection engine whose signals should then undergo expert review and can underpin evaluation of signals from other systems. With text embeddings at its core, the method allows for data augmentation without hand-tuning⁶⁸; we did not, however, venture down this path.

Data mining models generally carry no causal meaning, and an oft-raised issue of NLP is the need for (often large) annotated corpora which requires much work and continuous updating to remain relevant, the very thing we attempted to circumvent by reversing the prediction direction. Others have used word embeddings to operationalise free text in a non-annotated manner. For example, Workman et al.⁶⁹ showed that word embeddings can help overcome the problems of misspelling in a pharmacovigilance application; the RedMed model was trained on Reddit posts to extract *health entities* therein and performed reasonably well in such consumer-generated content⁷⁰; and combining pre-trained word embeddings and conditional random fields could have flagged potential cutaneous adverse reactions to two chemotherapy classes in internet content before they were reported in the scientific literature.⁴⁸

We trained one model per drug exposure for a total of 10 270 individual models. Although multi-label architectures sometimes aid learning,⁷¹ we found this to drown pertinent signals in models with thousands of outputs nodes in a single network. This probably happens because the model can only optimise a single loss value, and we found no good way to automatically up- or down-weight contributions from different outputs. Further, in a multi-label feed-forward architecture, all weights are shared except those between the last hidden layer and the outputs, and there seems to be no good reason that predicting the risk of, say, exposure to metformin should be so intimately linked to that of olanzapine. One potentially viable alternative might have been a factorial-like design in which each model had four mutually exclusive outcome nodes: exposure to none of the drugs, drug 1 only, drug 2 only and both drugs.

As mentioned, several options exist for the reference sets in the relevance evaluation. Among these, we chose UKU side effect rating scale for three principal reasons. First, the UKU items were originally developed in a Nordic setting, so English-Danish translations are readily available. Second, the UKU items were developed to gauge the side-effect load of psychotropics, and so their (somewhat) well-defined pharmacological mechanisms aid the assessment of biological plausibility of signals. Third, our results are readily put in a scientific context because the UKU scale has been used for several years and in different contexts,⁷²⁻⁷⁴ ensuring transparency with respect to and confidence in the translations for readers unfamiliar with the Danish language.

When designing our approach, we had institutional/regulatory pharmacovigilance in mind, but alternative use cases exist, such as patient-level decision-making support and drug repurposing research. Including patient characteristics (e.g., age, sex and comorbidities) would enable clinical staff to query the method for single drugs or drug combinations potentially explaining the symptoms of their patients. Instead of looking at drugs given disproportionately often for a given term, we could focus on those given more rarely (so with the odds ratio of <1) potentially eliciting interesting novel target conditions for existing treatments similar in spirit to, e.g., Kessing et al.⁷⁵

Combinatorial explosion is a well-known challenge for the study of DDIs: A person using seven different medicines is exposed to 21 two-way drug combinations. This challenge is only exacerbated if higher order combinations are considered. So, instead of modelling this explicitly, one could consider higher order interactions (e.g., three- or four-way) by piecing together two-way combinations that yield predicted probabilities above a certain threshold when multiplied, i.e., using a simplistic approximation to the predicted joint probability.

An alternative approach, and indeed research question, would have been to compare new in-hospital exposures with terms in subsequent days for immediate side effects. To be feasible, this would likely require a much larger data set to have sufficient exposure-outcome pairs. It might, however, be less unwieldy as such an approach could focus on new(er) drugs drastically reducing the number of labels (and, thus, models to be trained).

4.1 | Strengths and limitations

Our approach has six principal strengths. First, its unsupervised nature drastically reduces the need for manual work. This sets it apart from most other published studies using NLP in pharmacovigilance that tend to hinge on manual curation. Second, the method is language-agnostic owing to its unsupervised nature, so that it does not rely on a vocabulary for looking up words. This renders the approach potentially useful for pharmacovigilance in also smaller languages. Third, our corpus is quite large, a natural consequence of its non-reliance on curated data. Fourth, skipgrams (i.e., using sub-word information) enable embedding of also i.a. word bigrams, misspellings and out-of-vocabulary words. Fifth, the crude and almost reductionist nature of our approach circumvents many difficulties posed by NLP because we break documents down to basic components and use them without modelling semantics and syntax. Finally, using the UKU side effect rating scale (i.e., a Nordic, translated, pharmacology-based and widely used tool in Denmark) aids in contextualising the results. Even though the UKU side effect rating scale target psychotropics, interesting signals emerged also for somatic drugs (Table S1).

This study, however, is subject to several limitations. First, the apparently well-defined temporality obtained using doorstep medication profiles does not necessarily guarantee that what is reported in the text occurred after start of exposure. This potential problem, and source of protopathic bias,³² is not unique to our approach but rather necessitates cautious interpretation of any signal detection method, in longitudinal and case-report settings alike. Second, we do not actually have data on prescriptions from the primary sector but rely on the doorstep registration of pre-existing medication. Physicians are obliged to record these doorstep medication profiles, and we expect they generally be accurate despite occasional exceptions. Third, we considered exposure a binary notion and, due to the nature of the data, do not have well-defined start-of-exposure. Doses could be considered, perhaps on an ordinal scale, if the interest revolves around dose-related ADRs; the lack of well-

defined exposure time could be mitigated if doorstep medication profiles were based on data from the Danish Drug Statistics Register⁷⁶ (unavailable to us when conducting this study). Fourth, word embeddings are powerful but not magical: The method clearly links clinical terms with similar meanings (even if lexicographically very different) to similar medications profiles, but the embedding model has difficulties with i.a. rare variations. These yield different embedding vectors resulting in noisy signal profiles fitting poorly with clinical expectations. However, rarity of terms also hampers other kinds of association-mining or disproportionality-analytic techniques, and our method might even be less prone because few mentions could suffice to at least hint at relevant clinical cousins. Fifth, even if the doorstep medication profiles are correct, we have no records of exposure to over-the-counter and herbal drugs, and we have to assume patients be compliant, just as any study using secondary data. Finally, we only had data on inpatients who were not, generally, admitted due to side effects although this is common.^{77–80} Inpatients are not representative of the general population, and so, with the data at our disposal, the safety signals might be somewhat conditional on frailty although this could be mitigated by focusing on specific sub-populations (e.g., elderly or oncological patients).

5 | CONCLUSION

Combining various flavours of machine learning and data scientific tools, we have built an end-to-end pipeline for mining associations between free-text information and medication exposure without the need for manual curation. We achieve this by turning things upside down, predicting not the likely outcome of a range of exposures but also the likely exposures for one or several outcomes of interest.

The congruence analysis suggests that the method pick up pertinent information, even when supplied with synonyms, and with 8% of single-drug and 14% of drug-pair signals being possibly undocumented side effects, it provides a hit rate appropriate for its purpose: shortlisting few relevant signals from thousands of noisy ones.²⁸ These shortlists would then undergo review by pharmacologists, pharmacists or other pharmacovigilance experts^{5,28} to elicit truly unknown side effects (safety signal detection) or aid substantiating/refuting suspected side effects emerging from, e.g., spontaneous case reports (safety signal refinement).

Our approach is original in the field of side effect detection and helps overcome many limitations of NLP methods relying on curated data including being language-agnostic. Crucially, this makes our method

appealing in settings that must make sense of non-English free text for pharmacovigilance while lending itself well to alternative use cases, e.g., patient-level decision-making support and drug repurposing.

ACKNOWLEDGEMENTS

The authors would also like to thank DrugBank for granting access to their database. Figure 1 contains various Font Awesome icons (<https://fontawesome.com/license>).

CONFLICT OF INTEREST

SB reports ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, and managing board memberships in Proscion A/S and Intomics A/S outside the submitted work. All other authors report no competing interests.

DATA AVAILABILITY STATEMENT

The permissions do not allow us to offer access or to share the EHR data with third parties. Application for access to relevant authorities can be made. Trained on the full corpus, the embedding model contains sensitive information and so cannot be shared with third parties. This study's codebase is available online (doi: [10.5281/zenodo.5598068](https://doi.org/10.5281/zenodo.5598068)).

ORCID

Benjamin Skov Kaas-Hansen  <https://orcid.org/0000-0003-1023-0371>

Davide Placido  <https://orcid.org/0000-0002-8327-8838>

Cristina Leal Rodriguez  <https://orcid.org/0000-0002-3133-0630>

Hans-Christian Thorsen-Meyer  <https://orcid.org/0000-0001-7002-3089>

Simona Gentile  <https://orcid.org/0000-0003-3212-2865>

Gesche Jürgens  <https://orcid.org/0000-0002-3317-8118>

Stig Ejdrup Andersen  <https://orcid.org/0000-0002-1914-4720>

REFERENCES

- Delamothe T. Reporting adverse drug reactions. *Br Med J*. 1992;304(6825):465-465.
- Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Ann Intern Med*. 2004;140(10):795-801. doi:[10.7326/0003-4819-140-10-200405180-00009](https://doi.org/10.7326/0003-4819-140-10-200405180-00009)
- Human Medicines Evaluation Unit. Clinical safety data management: definitions and standards for expedited reporting. European Agency for the Evaluation of Medicinal Products. 1995. Technical report. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-15.pdf. Accessed 8 June 2022.
- Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet*. 2000;356(9237):1255-1259. doi:[10.1016/S0140-6736\(00\)02799-9](https://doi.org/10.1016/S0140-6736(00)02799-9)
- Lindquist M. VigiBase, the WHO global ICSR database system: basic facts. *Drug Inf J*. 2008;42(5):409-419. doi:[10.1177/009286150804200501](https://doi.org/10.1177/009286150804200501)
- Edwards IR. An agenda for UK clinical pharmacology: pharmacovigilance. *Br J Clin Pharmacol*. 2012;73(6):979-982. doi:[10.1111/j.1365-2125.2012.04249.x](https://doi.org/10.1111/j.1365-2125.2012.04249.x)
- Alvarez-Requejo A, Carvajal A, Bégaud B, Moride Y, Vega T, Arias LH. Under-reporting of adverse drug reactions. Estimate based on a spontaneous reporting scheme and a sentinel system. *Eur J Clin Pharmacol*. 1998;54(6):483-488. doi:[10.1007/s002280050498](https://doi.org/10.1007/s002280050498)
- Moride Y, Haramburu F, Requejo AA, Bégaud B. Under-reporting of adverse drug reactions in general practice. *Br J Clin Pharmacol*. 1997;43(2):177-181. doi:[10.1046/j.1365-2125.1997.05417.x](https://doi.org/10.1046/j.1365-2125.1997.05417.x)
- Patrignani A, Palmieri G, Ciampani N, Moretti V, Mariani A, Racca L. Under-reporting of adverse drug reactions, a problem that also involves medicines subject to additional monitoring. Preliminary data from a single-center experience on novel oral anticoagulants. *Giornale Italiano di Cardiologia (Rome)*. 2018; 19(1):54-61.
- Danish Medicines Agency. Bivirkningsindberetninger om afhængighed ved tramadol: Gennemgang og analyse af danske indberetninger. 2018. Online. Available at: <https://laegemiddelstyrelsen.dk/da/nyheder/2018/ny-rapport-om-bivirkningsindberetninger-om-den-smertestillende-medicin-tramadol>. Accessed 8 June 2022.
- Aggarwal CC. *Data Mining—The Textbook*. Springer; 2015.
- Goldberg Y. *Neural Network Methods in Natural Language Processing*. 1st ed. Morgan & Claypool Publishers; 2017.
- Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012;4(125):125ra31. doi:[10.1126/scitranslmed.3003377](https://doi.org/10.1126/scitranslmed.3003377)
- Iyer SV, Lependu P, Harpaz R, Bauer-Mehren A, Shah NH. Learning signals of adverse drug-drug interactions from the unstructured text of electronic health records. *AMIA Jt*. 2013; 2013:83-87.
- Iyer SV, Harpaz R, Lependu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc*. 2014;21(2):353-362. doi:[10.1136/amiajnl-2013-001612](https://doi.org/10.1136/amiajnl-2013-001612)
- Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. *J Am Med Inform Assoc*. 2020;27(1):39-46. doi:[10.1093/jamia/ocz101](https://doi.org/10.1093/jamia/ocz101)
- Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. *J Am Med Inform Assoc*. 2013;20(5):947-953. doi:[10.1136/amiajnl-2013-001708](https://doi.org/10.1136/amiajnl-2013-001708)
- Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;11(7):449-490. doi:[10.2147/CLEP.S91125](https://doi.org/10.2147/CLEP.S91125)
- WHO Collaborating Centre for Drug Statistics Methodology. WHOCC—home. 2021. Available at: <https://www.whocc.no>. Accessed 8 June 2022.

20. Goldberg Y. A primer on neural network models for natural language processing. *J Artif Intell Res.* 2016;57:345-420. doi:[10.1613/jair.4992](https://doi.org/10.1613/jair.4992)
21. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. *Trans Assoc Comput Linguist.* 2017;5:135-146. doi:[10.1162/tacl_a_00051](https://doi.org/10.1162/tacl_a_00051)
22. Thomas CE, Jensen PB, Werge T, Brunak S. Negation scope and spelling variation for text-mining of Danish electronic patient records. In: Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi); 2014:64-68.
23. Bird S, Loper E, Klein E. *Natural language processing with python.* United States of America: O'Reilly Media Inc.; 2009.
24. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval.* New York NY, USA: Cambridge University Press; 2008.
25. Steyerberg EW, Harrell FEJ. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;1(69):245-247. doi:[10.1016/j.jclinepi.2015.04.005](https://doi.org/10.1016/j.jclinepi.2015.04.005)
26. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35(29):1925-1931.
27. Leon SJ. *Linear Algebra With Applications.* 7th ed. Pearson Prentice Hall; 2006.
28. Trifirò G, Pariente A, Coloma PM, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf.* 2009;18(12):1176-1184. doi:[10.1002/pds.1836](https://doi.org/10.1002/pds.1836)
29. Lingjærde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K. The UKU side effect rating scale: a new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand.* 1987;76(s334):1-100. doi:[10.1111/j.1600-0447.1987.tb10566.x](https://doi.org/10.1111/j.1600-0447.1987.tb10566.x)
30. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2017;46(D1):D1074-D1082. doi:[10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037)
31. Aagaard L, Kristensen MB. The national drug interactions database [in Danish]. *Ugeskr Laeger.* 2005;167(35):3283-3286.
32. Faillie J. Indication bias or protopathic bias? *Br J Clin Pharmacol.* 2015;80(4):779-780. doi:[10.1111/bcp.12705](https://doi.org/10.1111/bcp.12705)
33. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37-46. doi:[10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)
34. Benchimol EI, Smeeth L, Guttman A, et al. The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 2015;12(10):e1001885 doi:[10.1371/journal.pmed.1001885](https://doi.org/10.1371/journal.pmed.1001885)
35. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. *BMC Geriatr.* 2017;17(1):230. doi:[10.1186/s12877-017-0621-2](https://doi.org/10.1186/s12877-017-0621-2)
36. Hult S, Sartori D, Bergvall T, et al. A feasibility study of drug-drug interaction signal detection in regular pharmacovigilance. *Drug Saf.* 2020;43(8):775-785.
37. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *Br J Clin Pharmacol.* 2012;73(5):674-684. doi:[10.1111/j.1365-2125.2011.04153.x](https://doi.org/10.1111/j.1365-2125.2011.04153.x)
38. Luo Y, Thompson WK, Herr TM, et al. Natural language processing for EHR-based pharmacovigilance: a structured review. *Drug Saf.* 2017;40(11):1075-1089. doi:[10.1007/s40264-017-0558-6](https://doi.org/10.1007/s40264-017-0558-6)
39. Juhlin K, Star K, Norén GN. A method for data-driven exploration to pinpoint key features in medical data and facilitate expert review. *Pharmacoepidemiol Drug Saf.* 2017;26(10):1256-1265. doi:[10.1002/pds.4285](https://doi.org/10.1002/pds.4285)
40. Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. *Stat Med.* 2008;27(16):3057-3070. doi:[10.1002/sim.3247](https://doi.org/10.1002/sim.3247)
41. Polepalli Ramesh B, Belknap SM, Li Z, Frid N, West DP, Yu H. Automatically recognizing medication and adverse event information from Food and Drug Administration's adverse event reporting system narratives. *JMIR Med Inform.* 2014;2(1):e10. doi:[10.2196/medinform.3022](https://doi.org/10.2196/medinform.3022)
42. Maciejewski M, Lounkine E, Whitebread S, et al. Reverse translation of adverse event reports paves the way for de-risking pre-clinical off-targets. *elife.* 2017;6. doi:[10.7554/eLife.25818](https://doi.org/10.7554/eLife.25818)
43. Dewulf P, Stock M, de Baets B. Cold-start problems in data-driven prediction of drug-drug interaction effects. *Pharmaceuticals (Basel).* 2021;14(5):429. doi:[10.3390/ph14050429](https://doi.org/10.3390/ph14050429)
44. Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction. *BMC Bioinform.* 2021;22(1):385. doi:[10.1186/s12859-021-04298-y](https://doi.org/10.1186/s12859-021-04298-y)
45. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *J Biomed Inform.* 2016;62:148-158. doi:[10.1016/j.jbi.2016.06.007](https://doi.org/10.1016/j.jbi.2016.06.007)
46. Chen X, Deldossi M, Aboukhamis R, et al. Mining adverse drug reactions in social media with named entity recognition and semantic methods. *Stud Health Technol Inform.* 2017;245:322-326.
47. Rezaallah B, Lewis DJ, Pierce C, Zeilhofer H, Berg B. Social media surveillance of multiple sclerosis medications used during pregnancy and breastfeeding: content analysis. *J Med Internet Res.* 2019;21(8):e13003. doi:[10.2196/13003](https://doi.org/10.2196/13003)
48. Nikfarjam A, Ransohoff JD, Callahan A, et al. Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JPHS.* 2019;5(2):e11264. doi:[10.2196/11264](https://doi.org/10.2196/11264)
49. Gavrielov-Yusim N, Kürzinger M, Nishikawa C, et al. Comparison of text processing methods in social media-based signal detection. *Pharmacoepidemiol Drug Saf.* 2019;28(10):1309-1317.
50. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing twitter annotations to identify first-hand experiences of prescription drug use. *J Biomed Inform.* 2015;58:280-287. doi:[10.1016/j.jbi.2015.11.004](https://doi.org/10.1016/j.jbi.2015.11.004)
51. Alvaro N, Miyao Y, Collier N. TwiMed: twitter and PubMed comparable corpus of drugs, diseases, symptoms, and their relations. *JPHS.* 2017;3(2):e24. doi:[10.2196/publichealth.6396](https://doi.org/10.2196/publichealth.6396)
52. Jiang K, Chen T, Calix RA, Bernard GR. Prediction of personal experience tweets of medication use via contextual word representations. *Conf Proc IEEE Eng Med Biol Soc.* 2019;2019:6093-6096. doi:[10.1109/EMBC.2019.8856753](https://doi.org/10.1109/EMBC.2019.8856753)
53. Liu J, Zhao S, Zhang X. An ensemble method for extracting adverse drug events from social media. *Artif Intell Med.* 2016;6(70):62-76. doi:[10.1016/j.artmed.2016.05.004](https://doi.org/10.1016/j.artmed.2016.05.004)
54. Emadzadeh E, Sarker A, Nikfarjam A, Gonzalez G. Hybrid semantic analysis for mapping adverse drug reaction mentions

- in tweets to medical terminology. *AMIA Ann Symp Proc.* 2017; 2017:679-688.
55. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J Am Med Inform Assoc.* 2017;24(4):813-821.
 56. Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. In: SHB '12: Proceedings of the 2012 international workshop on Smart health and well-being; 2012:25-32.
 57. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? *JMIR Public Health Surveill.* 2017;3(2):e36. doi:[10.2196/publichealth.6577](https://doi.org/10.2196/publichealth.6577)
 58. Carbonell P, Mayer MA, Bravo À. Exploring brand-name drug mentions on twitter for pharmacovigilance. *Stud Health Technol Inform.* 2015;210:55-59.
 59. Gattepaille LM. How far can we go with just out-of-the-box BERT models? In: Proceedings of the 5th Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; 2020:95-100.
 60. Eshleman R, Singh R. Leveraging graph topology and semantic context for pharmacovigilance through twitter-streams. *BMC Bioinform.* 2016;17(Suppl 13):335. doi:[10.1186/s12859-016-1220-5](https://doi.org/10.1186/s12859-016-1220-5)
 61. Eriksson R, Werge T, Jensen LJ, Brunak S. Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population. *Drug Saf.* 2014;37(4):237-247. doi:[10.1007/s40264-014-0145-z](https://doi.org/10.1007/s40264-014-0145-z)
 62. Sørup FKH. Exploring associations between text mined adverse events and antipsychotic drug use (PhD thesis). 2019.
 63. Oronoz M, Gojenola K, Perez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: mining adverse drug reactions. *J Biomed Inform.* 2015;56:318-332. doi:[10.1016/j.jbi.2015.06.016](https://doi.org/10.1016/j.jbi.2015.06.016)
 64. Segura-Bedmar I, Martinez P. Pharmacovigilance through the development of text mining and natural language processing techniques. *J Biomed Inform.* 2015;58:288-291. doi:[10.1016/j.jbi.2015.11.001](https://doi.org/10.1016/j.jbi.2015.11.001)
 65. Matsuda S, Aoki K, Tomizawa S, et al. Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JPHS.* 2017;3(1):e10. doi:[10.2196/publichealth.6872](https://doi.org/10.2196/publichealth.6872)
 66. Ujiie S, Yada S, Wakamiya S, Aramaki E. Identification of adverse drug event-related Japanese articles: natural language processing analysis. *JMIR Med Inform.* 2020;8(11):e22661. doi:[10.2196/22661](https://doi.org/10.2196/22661)
 67. Usui M, Aramaki E, Iwao T, Wakamiya S, Sakamoto T, Mochizuki M. Extraction and standardization of patient complaints from electronic medication histories for pharmacovigilance: natural language processing analysis in Japanese. *JMIR Med Inform.* 2018;6(3):e11021. doi:[10.2196/11021](https://doi.org/10.2196/11021)
 68. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data.* 2019;6(1):60. doi:[10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0)
 69. Workman TE, Divita G, Shao Y, Zeng-Treitler Q. A proficient spelling analysis method applied to a pharmacovigilance task. *Stud Health Technol Inform.* 2019;264:452-456. doi:[10.3233/SHTI190262](https://doi.org/10.3233/SHTI190262)
 70. Lavertu A, Altman RB. RedMed: extending drug lexicons for social media applications. *J Biomed Inform.* 2019;99:103307. doi:[10.1016/j.jbi.2019.103307](https://doi.org/10.1016/j.jbi.2019.103307)
 71. Chollet F. *Deep Learning With Python.* New York, USA: Manning Publications Co.; 2018.
 72. Jürgens G, Andersen SE, Rasmussen HB, et al. Effect of routine cytochrome P450 2D6 and 2C19 genotyping on antipsychotic drug persistence in patients with schizophrenia: a randomized clinical trial. *JAMA Netw Open.* 2020;3(12):e2027909. doi:[10.1001/jamanetworkopen.2020.27909](https://doi.org/10.1001/jamanetworkopen.2020.27909)
 73. Seitz DP, Adunuri N, Gill SS, Gruneir A, Herrmann N, Rochon P. Antidepressants for agitation and psychosis in dementia. *Cochrane Database Syst Rev.* 2011;(2):CD008191. doi:[10.1002/14651858.CD008191.pub2](https://doi.org/10.1002/14651858.CD008191.pub2)
 74. Bock MS, Achter ONV, Dines D, et al. Clinical validation of the self-reported Glasgow antipsychotic side-effect scale using the clinician-rated UKU side-effect scale as gold standard reference. *J Psychopharmacol.* 2020;34(8):820-828. doi:[10.1177/0269881120916122](https://doi.org/10.1177/0269881120916122)
 75. Kessing LV, Rytgaard HC, Ekstrøm CT, Torp-Pedersen C, Berk M, Gerds TA. Antihypertensive drugs and risk of depression. *Hypertension.* 2020;76(4):1263-1279. doi:[10.1161/HYPERTENSIONAHA.120.15605](https://doi.org/10.1161/HYPERTENSIONAHA.120.15605)
 76. Gregersen R, Wiingreen R, Rosenberg J. Health-related register-based research in Denmark (in Danish). *Ugeskr Laeger.* 2018;180(43):e36.
 77. Lombardi N, Crescioli G, Bettiol A, et al. Italian emergency department visits and hospitalizations for outpatients' adverse drug events: 12-year active pharmacovigilance surveillance (the MEREAFaPS study). *Front Pharmacol.* 2020;11:412. doi:[10.3389/fphar.2020.00412](https://doi.org/10.3389/fphar.2020.00412)
 78. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. *N Engl J Med.* 2011;365(21):2002-2012. doi:[10.1056/NEJMsa1103053](https://doi.org/10.1056/NEJMsa1103053)
 79. Lombardi N, Bettiol A, Crescioli G, et al. Risk of hospitalisation associated with benzodiazepines and z-drugs in Italy: a nationwide multicentre study in emergency departments. *Intern Emerg Med.* 2020;15(7):1291-1302. doi:[10.1007/s11739-020-02339-7](https://doi.org/10.1007/s11739-020-02339-7)
 80. Crescioli G, Bettiol A, Bonaiuti R, et al. Risk of hospitalization associated with cardiovascular medications in the elderly Italian population: a Nationwide multicenter study in emergency departments. *Front Pharmacol.* 2021;11. doi:[10.3389/fphar.2020.611102](https://doi.org/10.3389/fphar.2020.611102)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kaas-Hansen BS, Placido D, Rodríguez CL, et al. Language-agnostic pharmacovigilant text mining to elicit side effects from clinical notes and hospital medication records. *Basic Clin Pharmacol Toxicol.* 2022;131(4): 282-293. doi:[10.1111/bcpt.13773](https://doi.org/10.1111/bcpt.13773)