



OPEN ACCESS

EDITED BY
Marco Tuccori,
University of Pisa, Italy

REVIEWED BY
Giulia Valdiserra,
University of Pisa, Italy
Michele Fusaroli,
University of Bologna, Italy

*CORRESPONDENCE
Thomas Lieber,
✉ t.lieber@lareb.nl

SPECIALTY SECTION
This article was submitted to Advanced Methods in Pharmacovigilance and Pharmacoepidemiology, a section of the journal Frontiers in Drug Safety and Regulation

RECEIVED 09 December 2022
ACCEPTED 24 January 2023
PUBLISHED 07 February 2023

CITATION
Lieber T, Gosselt HR, Kools PC, Kruijssen OC, Van Lierop SNC, Härmark L and Van Hunsel FPAM (2023), Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment. *Front. Drug Saf. Regul.* 3:1120135. doi: 10.3389/fdsfr.2023.1120135

COPYRIGHT
© 2023 Lieber, Gosselt, Kools, Kruijssen, Van Lierop, Härmark and Van Hunsel. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment

Thomas Lieber^{1*}, Helen R. Gosselt¹, Pelle C. Kools², Okko C. Kruijssen², Stijn N. C. Van Lierop², Linda Härmark¹ and Florence P. A. M. Van Hunsel¹

¹Netherlands Pharmacovigilance Centre Lareb, 's-Hertogenbosch, Netherlands, ²Faculty of Social Sciences, Radboud Universiteit, Nijmegen, Netherlands

Objective: To improve a previously developed prediction model that could assist in the triage of individual case safety reports using the addition of features designed from free text fields using natural language processing.

Methods: Structured features and natural language processing (NLP) features were used to train a bagging classifier model. NLP features were extracted from free text fields. A bag-of-words model was applied. Stop words were deleted and words that were significantly differently distributed among the case and non-case reports were used for the training data. Besides NLP features from free-text fields, the data also consisted of a list of signal words deemed important by expert report assessors. Lastly, variables with multiple categories were transformed to numerical variables using the weight of evidence method.

Results: the model, a bagging classifier of decision trees had an AUC of 0.921 (95% CI = 0.918–0.925). Generic drug name, info text length, ATC code, BMI and patient age. were most important features in classification.

Conclusion: this predictive model using Natural Language Processing could be used to assist assessors in prioritizing which future ICSRs to assess first, based on the probability that it is a case which requires clinical review.

KEYWORDS

prediction model, machine learning, natural language processing, drug safety reports, pharmacovigilance

1 Introduction

Before a medicine receives marketing authorization, evidence of its safety and efficacy is limited to the results from clinical trials, where patients are selected carefully and followed up under controlled conditions for a limited period of time. This means that at the time of a medicine's authorisation, the information about its safety profile is limited. After authorisation the medicine may be used by a large number of patients, for a long period of time, in patients with comorbidities that warrants the use of other medicines. New information on adverse drug reactions (ADRs) may emerge in such circumstances (Stricker and Psaty, 2004; European Medicines Agency, 2022). It is therefore essential that the safety of all medicines is monitored throughout their lifecycle. Pharmacovigilance in this perspective is the *science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine-related problem* (European Medicines Agency, 2022).

A safety signal in this context is *information on a new or known adverse event that may be caused by a medicine and requires further investigation* (European Medicines Agency, 2021). Safety signals can be detected from a wide range of sources, such as spontaneous reports, clinical studies and scientific literature. Spontaneous reporting systems remain the most efficient and fastest way to get insight in the safety profile of drugs and vaccines (Raine et al., 2007; Lester et al., 2013; Klungel and Pottegård, 2021; Lo Re et al., 2021; Rudolph et al., 2022). Recently, for the COVID-19 vaccines important signals such as thrombosis with thrombocytopenia syndrome for the viral vector vaccines were signalled based on cases (Lane and Shakir, 2022).

Signal detection solely based on the review of individual case safety reports (ICSRs), also called case-by-case analysis or qualitative signal detection, has proven its value (Egberts, 2007). However, it is becoming increasingly time consuming given the growing volumes of data in pharmacovigilance and can become more complex as co-variables may play a role (Egberts, 2007). In the past decades a lot of experience has been gained with the use of different types of disproportionality analyses in signal detection (Bate et al., 1998; van Puijenbroek et al., 2002; Orre et al., 2005; Seabroke et al., 2016). These methods have earned their place in the signal detection process, although they have their limitations. First, they do not take into account the clinically relevant parameters present in the individual reports and should always be followed by clinical review of cases (Scholl, 2022). Secondly, they are prone to certain types of bias (de Geaaf et al., 2003; Pariente et al., 2007) and lastly, a minimum number of cases is needed before a statistical signal can be detected.

Netherlands Pharmacovigilance Centre Lareb relies on both case-by-case analyses and statistical based methods in order to find potential signals. A statistical screening method based on a prediction model is used to identify potential combinations of drugs and adverse drug reactions (ADRs) which require further review (Scholl et al., 2018). In addition, methods such as time-to-onset analysis (Van Holle et al., 2014; Scholl and van Puijenbroek, 2016; Scholl et al., 2019) and topic modelling (Löscher et al., 2022) have been employed, with varying success in finding new signals. The majority of signals is currently still found by cases-by-case analyses in which trained pharmacovigilance assessors, mostly medical doctors and pharmacists, review ICSRs and discuss them during a weekly Signal Detection Meeting (SDM). In order to lower the burden of having to assess all individual cases, Lareb has relied on methods such as triage to decrease the set of reports which needs clinical review by assessors. This triage is performed manually by trained assessors; however, automation of this process could further reduce the workload.

Previously, others have attempted to tackle identifying important cases for signal detection in an automated manner; Munoz et al. (Muñoz et al., 2020). Developed and validated a model predictive of an ICSR's pharmacovigilance utility based on the United States Food and Drug Administration (FDA) database. The strongest predictors of ICSR inclusion in this study were reporting of a designated medical event (DME) and positive dechallenge. Their validated model showed modest discriminative ability (Muñoz et al., 2020), (Cherkas et al., 2022). Developed a machine learning-based model that can predict the likelihood of a causal association of an observed drug-reaction combination in an ICSR. The model performed

well in predicting the causality assessment of drug-event pairs compared with clinical judgment. It should be noted that a causal relationship for a drug-event combination in an ICSR is not the same as having signal value. ICSRs with signal value need to have a certain degree of causality but not all ICSRs with a causal relationship are potential signals. The principle of using elements of information relating to causality assessment in an automated manner was also tested on the French pharmacovigilance database. However, the authors looked at drug-event pairs and not individual ICSRs (Berbain et al., 2020).

Recently, Lareb has published their first efforts to develop a prediction model to identify ICSRs that require clinical review. This was defined by identifying reports which were taken to the Signal Detection Meeting (SDM) and investigating the features of these ICSRs [26]. Most important features in this prediction model were: "absence of ADR in the Summary of product characteristics," "ADR reported as serious," "ADR labelled as an important medical event," "ADR reported by physician" and "positive rechallenge." An AUC of 0.75 (0.73–0.77) was obtained, which can be seen as moderate model performance (Gosselt et al., 2022).

The aim of this study is to improve a previously developed prediction model that could assist in the triage of individual case safety reports using the addition of features designed from free text fields using natural language processing.

2 Materials and methods

2.1 Source of data and ICSRs

ICSR reports received from 20-03-2019 to 30-04-2022 were extracted from Lareb's ICSR database. This time period was chosen since different criteria were used to label reports as case or non-case in earlier reports. Reports on vaccines were excluded from our analysis because also for these reports different criteria were used.

2.1.1 Outcome

The binary outcome is defined as ICSRs marked as 'case' or 'non-case' by scientific assessors. Cases are ICSRs that need in depth clinical review whereas non-case reports can be coded and directly stored in the database without clinical review.

2.2 Features

2.2.1 Structured features

The features were chosen from Lareb's ICSR database based on how informative they were perceived by expert assessors. Of these features all were included into the training data except for: 1) non-informative features, such as the safety report identification number and 2) features with small variances (≤ 0.0001) with respect to cases or non-cases, as these were unlikely to be useful to discriminate between cases and non-cases.

2.2.2 Natural language processing features

The "additional information" text field contained Dutch descriptions of the ADR that patients experienced and any

potential other relevant information such as treatment, contact with a healthcare professional or thoughts on causality made by the reporter (Rolfes et al., 2015). Because the semantic content of this text field could be informative for the prediction, NLP features were extracted from this text field using sklearn (Pedregosa et al., 2011) and nltk (Bird et al., 2009), which are libraries for predictive data analysis and statistical NLP developed for the python programming language.

First, a bag-of-words model was used with tokenization without stemming to assess NLP features, considering the frequency of the words mentioned. Stop words were excluded and a maximum word frequency of 0.2 was set to exclude structured text that was present in many ICSRs. Stop words refer to words that occur commonly in reports, but hold little semantic value. Secondly, the 2000 most frequent words (excluding stop words) in the additional info text field of case and non-case reports were compared. The words that were present in case - but not in non-case reports were subsequently stored on a list. These words were assessed on their usefulness by two independent Lareb pharmacovigilance experts. Also, words that were not yet on the list, but which at least one of the experts considered important, were added. Subsequently, the list was used as a vocabulary for the bag-of-words model and resulting features were appended to the existing feature set.

Lastly, the feature “info_length” was added. This feature represented the number of characters in the additional text field and was based on the assumption that cases would have more text in the additional text field as compared to non-cases.

2.2.3 Weight of evidence feature engineering

Categorical variables that could be equal to more than two different levels were changed into numerical variables using the weight of evidence (WOE) method. WOE is a feature engineering method and has been proven useful for predictive modelling (Dahal et al., 2008; Cao et al., 2021).

The benefits of using the WOE method is that you have one variable instead of multiple dummy variables for each level. This makes it easier to check the feature importance as you have to check only one variable opposed to multiple dummy variables.

The WOE is calculated for each level of a variable. For example, the variable primary_source has multiple levels: {physician, pharmacist, ..., consumer or other non – health professional}. The WOE can then be calculated using the following formula:

$$WOE = \ln \left\{ \frac{P(\text{Case})}{P(\text{Non Case})} \right\}$$

where $P(\text{Case})$ and $P(\text{Non Case})$ refer to the proportion of case and non-case reports respectively. If a level has a positive WOE value then the proportion of case reports is higher than the proportion of non-case reports.

2.3 Sample size

The dataset contained 49928 reports of which 18,236 (36.5%) were cases. Various resampling strategies (random upsampling of the minority class, random downsampling of the majority class and Synthetic Minority Over-sampling Technique [SMOTE; generating synthetic samples from the minority

class]) were tested using imblearn library (Lematre et al., 2017) to balance the number of samples of both classes in the training set.

2.4 Missing data

Relatively many features were compulsory for the ICSR reporter to answer, hence little data was missing. Missing data fields can be explained by reports send on paper/letter, or from other origins such as registers. Also, new features have been introduced to the reporting form over time and others have been removed, resulting in features with missing values. Variables containing missing data where age in years, weight in kg and length in cm. These variables were imputed using k nearest neighbor imputation ($k = 2$ neighbors) using the KNNImputer function from the sklearn package (Pedregosa et al., 2011).

2.5 Statistical analyses

2.5.1 Models

Multiple models were trained and tested, including logistic regression, Extreme Gradient Boosting (XGBoost), support vector machine, random forest, decision tree classifier and bagging classifier of decision trees. The bagging classifier was optimized using the parameters : *nestimators*, *max features*, and *criterion* where *nestimators* = {50, 100, 200, 500, 1000} refers to the amount of base estimators; *max features* = {10, 50, 100, 120} refers to the number of features to draw from the training data to train each base estimator. Lastly; *criterion* = {Gini, entropy, log loss} refers to the criterion to be optimized by the model. Best model parameters were chosen using 5-fold cross-validation and found using the RandomizedSearchCV function from the scikit-learn package (Pedregosa et al., 2011).

2.5.1.1 Model training and test set

Firstly, the data was divided into a training set (70%) and a test set (30%) where a fixed proportion case reports/non-case reports in each set was ensured using the RandomOverSampler from the imblearn package (Lematre et al., 2017). Model performance was assessed on the test set using accuracy, precision, recall, f1-score and area under the curve (AUC). The threshold for which the precision and recall were reported was chosen such that Youden’s index was the highest. Feature importance was determined by looking at what features were selected in every base classifier and how important that feature was for the prediction. This was provided by the bagging classifier from scikit-learn package (Pedregosa et al., 2011). These results were then summed for all base classifiers to get an idea of overall feature importance.

All data analysis was performed using Python 3.10.

3 Results

The training set contained 34,589 ICSR reports (70%) of which 12,634 (36.5%) were cases. Random upsampling of the minority class in the train set resulted in the best performance (data not shown) and was therefore applied to train the models.

TABLE 1 Features in the final data set.

| Feature name | Description |
|--|--|
| Features extracted from database used for prediction | |
| adr_count | The number of adverse drug reactions (ADR) reported |
| ATCode | Classification for drugs. The more similar the code, the more closely-related the drugs |
| comedication_count | The number of drugs that are used in addition to the suspect drug(s) |
| Dechallenge | If the ADR stops after stopping medication |
| drug_action_taken | The action that was taken corresponding to the drug when the side effects occurred |
| drug_brand_name | The brand name of a drug |
| drug_generic_name | The generic name of a drug |
| drug_suspect_interacting_count | The number of drugs that are suspected to have caused the adverse drug reaction(s) |
| Indication | The reason for taking a drug |
| IME | Important Medical Event, adverse drug reactions that deserve extra attention |
| Impact based on a Likert scale 1–5 | Severity of the adverse drug reaction |
| latency_start | If the time between starting medication and the onset of the ADR is known, or can be calculated |
| latency_stop | If the latency between stopping/reducing medication and the onset of ADR is known, or can be calculated |
| patient_sex | Sex of the patient |
| patient_age_year | Age of the patient |
| patient_body_weight_kg | Weight of the patient |
| patient_height_cm | Height of the patient |
| primary_source | The professional or consumer that reported a signal |
| primary_source_function | Specific role of the professional that reported the signal |
| reaction_outcome | The outcome of the side effect |
| rechallenge | If after a dechallenge medication is taken again and the same ADR occurs again |
| Spc | If the ADR is in the Summary of Product Characteristics of the drug |
| seriousness_case_report_level | Whether the adverse drug reaction was considered to be serious or non-serious. If multiple ADRs were reported, the seriousness of the most serious ADR was reported here |
| treatment | If the ADR has been treated |
| medication_previously_used | Whether the drug was used before by the patient according to the additional info text field |
| primary_source_reaction (1 feature per unique ADR, 99 in total) | The primary ADR of the patient as reported in the additional info text field. Because this is written in natural language text, every type of ADR is a separate feature: if that ADR occurs in the additional info text field or not (hence 1 feature per ADR) |
| seriousness_death | Whether seriousness_death was set to true in the additional info text field |
| seriousness_disabling | Whether seriousness_disabling was set to true in the additional info text field |
| seriousness_hospitalization | Whether seriousness_hospitalization was set to true in the additional info text field |
| Newly created features | |
| expert-identified signal words (1 feature per word, 33 in total) | For every expert-identified signal word, if it occurs in the additional info field (hence 1 feature per signal word) |
| info_length | The total number of characters in the additional info text field |

TABLE 2 Performance of the bagging classifier on the test set.

| Measure | |
|---------------|---------------------|
| Accuracy (%) | 84.8 |
| Precision (%) | 89.3 |
| Recall (%) | 86.4 |
| F1 Score | 0.89 |
| AUC (95% CI) | 0.921 (0.918–0.925) |

The list of content-related words from the additional_info text that were frequently present in cases and not in non-cases did not improve classification performance of the model. Most words on the list that were acquired using this method were considered irrelevant by two Lareb pharmacovigilance experts. The experts added another 33 words

that they considered important in the classification of cases and non-cases. An overview of the final list of words is given in [Supplementary Table S1](#).

In total 175 features were used for prediction. An overview of all the features in the final feature set and a short description of every feature are given in [Table 1](#).

The best-performing model was a bagging classifier of decision trees. The optimal number of base estimators was 100, where each estimator selected a maximum of 100 features out of the 175 for prediction. Performance was good with an f1 score of 0.89 and AUC of 0.921 (95% CI = 0.918–0.925) ([Table 2](#)). The ROC curve as well as precision-recall curve showing all cut-off values are presented in [Figure 1](#).

The top 50 most important features for prediction are shown in [Figure 2](#).

The plot shows the relative importance of the variables in the model. The direction of the relation is “present.” This means that for example the variable “ziekenhuis” (hospital) can take the values 1 or

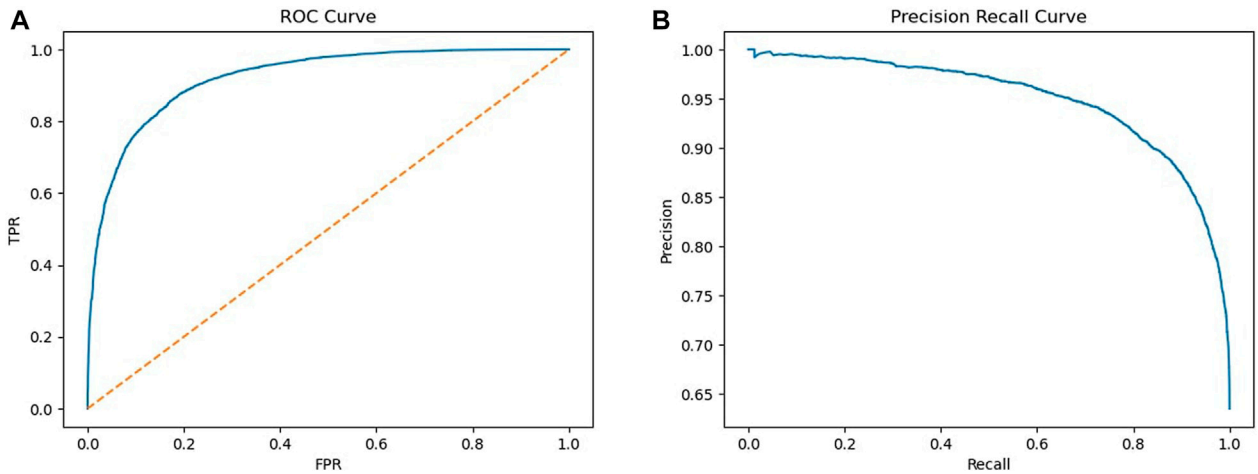


FIGURE 1
ROC curve (A) and precision recall curve (B).

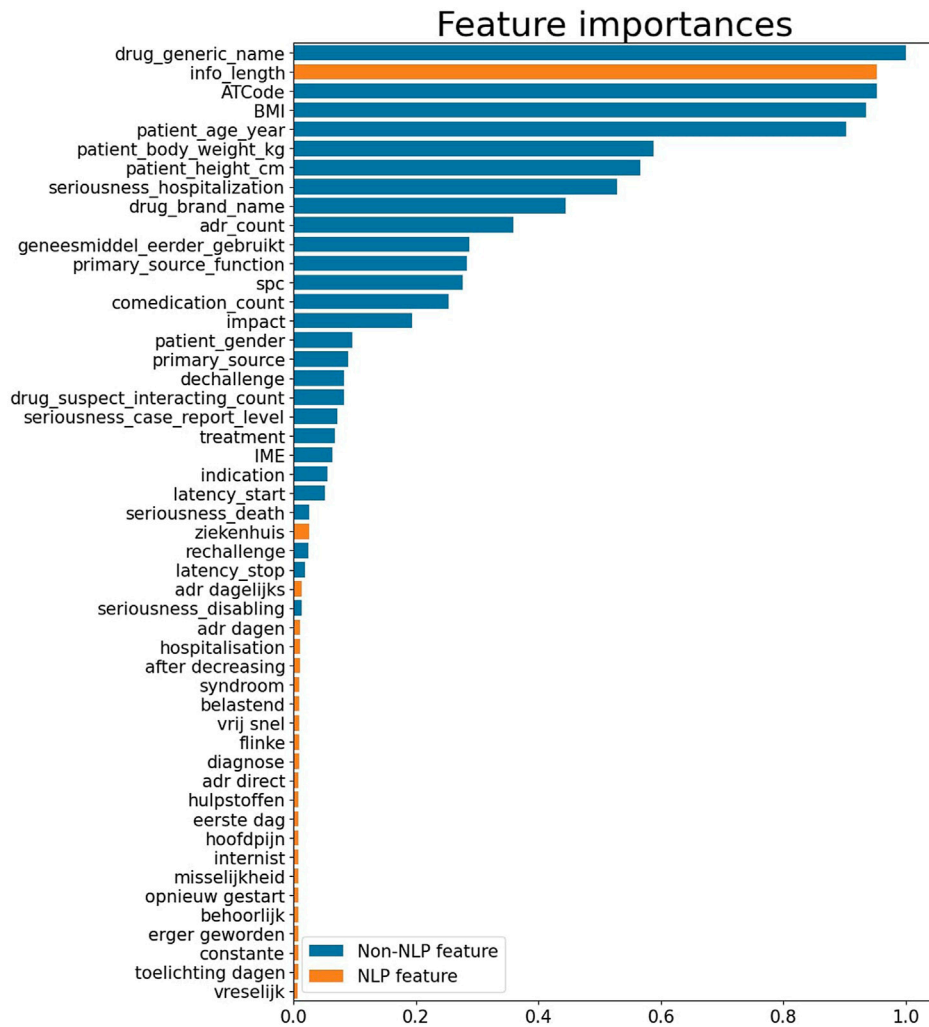


FIGURE 2
Feature importances of top 50 features. A longer bar indicates greater importance.

0 where 1 means that the presence of the word “ziekenhuis” in the report is predictive of whether a report is a case or non-case. For “patient_gender” a value of 1 refers to male and a 0 to female. The most important features comprised structured fields (non-NLP), describing patient characteristics (e.g., patient BMI or age), or drug descriptions (e.g., drug name, ATC code). (Figure 2). Other structured features that were important were related to the reporter (reporter function) or the ADR (number of ADRs, seriousness, latency, rechallenge/dechallenge). Additionally, the NLP feature “info_length” was of importance. Cases seem to have longer text lengths in the additional information field as compared to non-cases. Other NLP features, or expert signal words also contributed to classification, but were less important.

4 Discussion

Predictive modeling can be used for the identification of previously unrecognized risks of medicines in pharmacovigilance ICSRs. These methods have been applied on association level, for instance in “Vigirank” which was developed by the Uppsala Monitoring Centre as a data-driven predictive model for emerging safety signals (Caster et al., 2017). We have developed a prediction model using a combination of structured data fields and unstructured text fields to distinguish case reports from non-case reports with good performance [AUC of 0.92 (95% CI = 0.92–0.93), Precision of 0.89 and recall of 0.86]. The model could be used to assist clinical assessors in ranking the reports on importance. For future reports, the model can predict the probability that the report is a case report based on the different features in the model. A report with long texts in the additional information fields, about a patient with multiple ADRs that are not mentioned in the SmPC, for instance will have a higher probability of being a case report than a report with no additional information, reporting a single ADR that is labelled in the SmPC of the reported drug. Reports with higher probabilities could then be manually reviewed first. Moreover, the cut-off value for which probability the reports should be manually reviewed could be adjusted to the number of reports received. In case the number of reports is increasing and resources are the same, only reports with very high probabilities of being a case report could be reviewed. This is in line with a recent study where a machine learning-based model was developed with the aim to predict the likelihood of a causal association of an observed drug–reaction combination in an ICSR and thus assisting in pharmacovigilance ICSR case processing (Cherkas et al., 2022).

Current model with NLP performed better than our previously developed model that had an AUC of 0.75 (95% CI: 0.73–0.77). In addition, upon visual inspection of the precision and recall plot it was found that the current model also outperformed the previous model which was only based on structured fields (non-NLP features) (Gosselt et al., 2022). From the feature importance plot it can be seen that the feature ‘info_length’ seemed of high importance to distinguish case reports from non-case reports. This is not surprising, since also the follow-up information is added to the additional information, which is more often requested for interesting cases. The list of words defined by pharmacovigilance experts on the other hand, only minimally contributed to the model, however these still contributed more

than other techniques that we have tried including: extracting words and compare the frequency between cases and non-cases. This indicates the relevance of domain knowledge for this model. To better understand why some features are high in the feature importance plot, it would be interesting to further assess which reports are classified as case reports. For instance, ATC code or generic drug name are quite high in the list, which indicate that certain drugs are more often reported as case. These could be new drugs for example. Also, in contrast to our previous model, BMI is high in the list. An explanation for this could be that BMI may be related to other features in the dataset, such as the seriousness of an ADR, since high BMI is a risk factor for many complications (Alomar, 2014; Modesto et al., 2020). However, it is also possible that BMI was selected by chance, hence further assessment is required to better understand this result.

Beside the addition of NLP features, the outcome was slightly different defined as compared to our previous model. In our previous model, all reports discussed at a signal detection meeting were classified as “case.” In current study only reports that were potentially important for signal detection were marked as case. This may have created more homogeneous groups of reports which may also have improved the model in classifying the reports. Besides, current dataset contained more reports (49928 as compared to 30424) and the outcome was more balanced with 36.5% cases as compared to 4.7% in our previous study. This led to a higher number of training instances, which may have also improved the model performances. To limit the scope and complexity of the model, we excluded the “receive_date” of an ICSR in the feature set and made the model ‘static’ for the current time period. However, given that classification criteria may also change over time, it would be interesting to explore the use of continual learning models in future work (Lee and Lee, 2020).

The development of this method is part of a variety of methods such as auto-coding (Létinier et al., 2021; Martin et al., 2022) and automated case classification (Ball et al., 2018) to improve in the processing and signal detection of large amounts of ICSRs in pharmacovigilance databases.

5 Limitations

A limitation of our study is that we assessed single extracted NLP words. These single words were not assessed in relation to other words in the text and were taken out of context. It would be interesting to explore part of speech tagging, as this considers the context of the extracted words (Chiche and Yitagesu, 2022). Also, the use of sentiment analysis, which is a tool to extract opinions, perceptions or sentiments from free text (Birjali et al., 2021; Marcec and Likic, 2022). This could be an interesting tool to further explore, because the sentiment of the reporter could be of importance for signal detection. Lastly, the NLP was applied to one unstructured field. The Lareb database has more unstructured fields so in future modelling exercises these other fields could be explored using NLP as well.

Another limitation is that, even though, the definition ‘case report’ as defined by scientific assessors is more specific as compared to all reports discussed at a signal detection meeting, this is still a subjective measure and heterogeneity between assessors may exist. Scientific assessors decide whether the report is a case or

non-case based on their interpretation of the report using their expert knowledge. This process is hard to capture using a machine-learning model, which is why this is the best measure we have so far. That is also why the tool could be used to assist the assessors in ranking the reports on importance, while the reports still need clinical review by the assessors.

In conclusion, Natural Language Processing only minimally contributed to the predictive performances of the model. Yet, current model could be used to assist assessors in prioritizing which future ICSRs to assess first, based on the probability that it is a case which requires clinical review.

Data availability statement

The dataset for this study is not publicly available due to Lareb's data protection policy. Access to the data will be granted on reasonable request by the last author.

Author contributions

The concept for the study was initiated by LH with input from the other authors. TL, PK, OK, and SvL performed the data-analysis. LH, HG, and FH supervised the analysis. The first draft of the manuscript was written by FvH, HG, and TL. All authors were involved in the finalization of the manuscript and have approved the final version.

References

- Alomar, M. J. (2014). Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharm. J.* 22 (2), 83–94. doi:10.1016/j.jsps.2013.02.003
- Ball, R., Toh, S., Nolan, J., Haynes, K., Forshee, R., and Botsis, T. (2018). Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System. *Pharmacoepidemiol Drug Saf.* 27 (10), 1077–1084. doi:10.1002/pds.4645
- Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A., et al. (1998). A Bayesian neural network method for adverse drug reaction signal generation. *Eur. J. Clin. Pharmacol.* 54 (4), 315–321. doi:10.1007/s002280050466
- Berbain, T., Pariente, A., Miremont-Salamé, G., Grandvuillemin, A., Micallef, J., Chouchana, L., et al. (2020). Contribution of causality assessment for an automated detection of safety signals: An example using the French pharmacovigilance database. *Drug Saf.* 43 (3), 243–253. doi:10.1007/s40264-019-00887-2
- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language processing with Python*. Sebastopol, CA: O'Reilly Media Inc.
- Birjali, M., Kasri, M., and Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Syst.* 226, 107134. doi:10.1016/j.knosys.2021.107134
- Cao, Y., Wei, X., Fan, W., Nan, Y., Xiong, W., and Zhang, S. (2021). Landslide susceptibility assessment using the weight of evidence method: A case study in xunyang area, China. *PLoS One* 16 (1), e0245668. doi:10.1371/journal.pone.0245668
- Caster, O., Sandberg, L., Bergvall, T., Watson, S., and Norén, G. N. (2017). vigiRank for statistical signal detection in pharmacovigilance: First results from prospective real-world use. *Pharmacoepidemiol Drug Saf.* 26 (8), 1006–1010. doi:10.1002/pds.4247
- Cherkas, Y., Ide, J., and van Stekelenborg, J. (2022). Leveraging machine learning to facilitate individual case causality assessment of adverse drug reactions. *Drug Saf.* 45 (5), 571–582. doi:10.1007/s40264-022-01163-6
- Chiche, A., and Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. *J. Big Data* 9 (1), 10. doi:10.1186/s40537-022-00561-y
- Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., and Paudyal, P. (2008). Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence. *Geomorphology* 102, 496–510. doi:10.1016/j.geomorph.2008.05.041
- de Geaaf, L., Fabius, M. A., Diemont, W. L., and van Puijenbroek, E. P. (2003). The weber-curve pitfall: Effects of a forced introduction on reporting rates and reported

Acknowledgments

We thank Ingrid Oosterhuis for her role as a Lareb content expert and knowledge on the dataset.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdsfr.2023.1120135/full#supplementary-material>

adverse reaction profiles. *Pharm. World Sci.* 25 (6), 260–263. doi:10.1023/b:phar.0000006518.22231.ea

Egberts, T. C. (2007). Signal detection: Historical background. *Drug Saf.* 30 (7), 607–609. doi:10.2165/00002018-200730070-00006

European Medicines Agency (2022). Pharmacovigilance: Overview 2022. Available from: <https://www.ema.europa.eu/en/human-regulatory/overview/pharmacovigilance-overview>.

European Medicines Agency (2021). Safety signal 2021. Available from: <https://www.ema.europa.eu/en/glossary/safety-signal>.

Gosselt, H. R., Bazelmans, E. A., Lieber, T., van Hunsel, F., and Härmak, L. (2022). Development of a multivariate prediction model to identify individual case safety reports which require clinical review. *Pharmacoepidemiol Drug Saf.* 31, 1300–1307. doi:10.1002/pds.5553

Klungel, O. H., and Pottegård, A. (2021). Strengthening international surveillance of vaccine safety. *Bmj* 374, n1994. doi:10.1136/bmj.n1994

Lane, S., and Shakir, S. (2022). Assessing case fatality on cases of thrombosis with concurrent thrombocytopenia following COVID-19 vaccine AstraZeneca (vaxzevria) in the United Kingdom: A review of spontaneously reported data. *Drug Saf.* 45 (9), 1003–1008. doi:10.1007/s40264-022-01217-9

Lee, C. S., and Lee, A. Y. (2020). Clinical applications of continual learning machine learning. *Lancet Digit. Health* 2 (6), e279–e281. doi:10.1016/S2589-7500(20)30102-3

Lematre, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* 18 (1), 559–563. doi:10.48550/arXiv.1609.06570

Lester, J., Neyarapally, G. A., Lipowski, E., Graham, C. F., Hall, M., and Dal, P. G. (2013). Evaluation of FDA safety-related drug label changes in 2010. *Pharmacoepidemiol Drug Saf.* 22 (3), 302–305. doi:10.1002/pds.3395

Létiniér, L., Jouganous, J., Benkebil, M., Bel-Létoile, A., Goehrs, C., Singier, A., et al. (2021). Artificial intelligence for unstructured healthcare data: Application to coding of patient reporting of adverse drug reactions. *Clin. Pharmacol. Ther.* 110 (2), 392–400. doi:10.1002/cpt.2266

Lo Re, V., 3rd, Klungel, O. H., Chan, K. A., Panozzo, C. A., Zhou, W., and Winterstein, A. G. (2021). Global Covid-19 vaccine rollout and safety surveillance-how to keep pace. *Bmj* 373, n1416. doi:10.1136/bmj.n1416

Lösch, L., Brown, P., and van Hunsel, F. (2022). Using structural topic modelling to reveal patterns in reports on opioid drugs in a pharmacovigilance database. *Pharmacoepidemiol Drug Saf.* 31 (9), 1003–1006. doi:10.1002/pds.5502

- Marcec, R., and Likic, R. (2022). Using twitter for sentiment analysis towards AstraZeneca/oxford, pfizer/BioNTech and moderna COVID-19 vaccines. *Postgrad. Med. J.* 98 (1161), 544–550. doi:10.1136/postgradmedj-2021-140685
- Martin, G. L., Jouganous, J., Savidan, R., Bellec, A., Goehrs, C., Benkebil, M., et al. (2022). Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Saf.* 45 (5), 535–548. doi:10.1007/s40264-022-01153-8
- Modesto, A. C. F., Silveira, E. A., Santos, A. S. E. A. D. C., Rodrigues, A. P. D. S., Lima, D. M., Provin, M. P., et al. (2020). Prevalence of adverse drug events in severely obese adults and associated factors: Clinical trial baseline results. *Sci. Pharm.* 88 (4), 41. doi:10.3390/scipharm88040041
- Muñoz, M. A., Dal Pan, G. J., Wei, Y.-J. J., Delcher, C., Xiao, H., Kortepeter, C. M., et al. (2020). Towards automating adverse event review: A prediction model for case report utility. *Drug Saf.* 43 (4), 329–338. doi:10.1007/s40264-019-00897-0
- Orre, R., Bate, A., Norén, G. N., Swahn, E., Arnborg, S., and Edwards, I. R. (2005). A bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. *Int. J. Neural Syst.* 15 (3), 207–222. doi:10.1142/S0129065705000219
- Pariante, A., Gregoire, F., Fourrier-Reglat, A., Haramburu, F., and Moore, N. (2007). Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: The notoriety bias. *Drug Saf.* 30 (10), 891–898. doi:10.2165/00002018-200730100-00007
- Pedregosa, F., Varoquaux, G. L., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-Learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi:10.48550/arXiv.1201.049
- Raine, J. (2007). “Risk management: A European regulatory view,” in *Pharmacovigilance*. Editors R. M. Mann, and E. B. Andrews (Hoboken, NJ: John Wiley & Sons Ltd), 553–558.
- Rolfes, L., van Hunsel, F., Wilkes, S., van Grootheest, K., and van Puijenbroek, E. (2015). Adverse drug reaction reports of patients and healthcare professionals—differences in reported information. *Pharmacoepidemiol. Saf.* 24 (2), 152–158. doi:10.1002/pds.3687
- Rudolph, A., Mitchell, J., Barrett, J., Sköld, H., Taavola, H., Erlanson, N., et al. (2022). Global safety monitoring of COVID-19 vaccines: How pharmacovigilance rose to the challenge. *Ther. Adv. Drug Saf.* 13, 20420986221118972. doi:10.1177/20420986221118972
- Scholl, J. H. G. (2022). Signal detection in pharmacovigilance: Time for a new era? Thesis. Groningen: University of Groningen.
- Scholl, J. H. G., van Hunsel, F., Hak, E., and van Puijenbroek, E. P. (2018). A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in The Netherlands. *Pharmacoepidemiol. Drug Saf.* 27 (2), 199–205. doi:10.1002/pds.4364
- Scholl, J. H. G., van Hunsel, F., Hak, E., and van Puijenbroek, E. P. (2019). Time to onset in statistical signal detection revisited: A follow-up study in long-term onset adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* 28 (10), 1283–1289. doi:10.1002/pds.4790
- Scholl, J. H. G., and van Puijenbroek, E. P. (2016). The value of time-to-onset in statistical signal detection of adverse drug reactions: A comparison with disproportionality analysis in spontaneous reports from The Netherlands. *Pharmacoepidemiol. Drug Saf.* 25 (12), 1361–1367. doi:10.1002/pds.4115
- Seabroke, S., Candore, G., Juhlin, K., Quarcio, N., Wisniewski, A., Arani, R., et al. (2016). Performance of stratified and subgrouped disproportionality analyses in spontaneous databases. *Drug Saf.* 39 (4), 355–364. doi:10.1007/s40264-015-0388-3
- Stricker, B. H., and Psaty, B. M. (2004). Detection, verification, and quantification of adverse drug reactions. *BMJ* 329 (7456), 44–47. doi:10.1136/bmj.329.7456.44
- Van Holle, L., Tavares Da Silva, F., and Bauchau, V. (2014). Signal detection based on time-to-onset: Extending a new method from spontaneous reports to observational studies. *Pharmacoepidemiol. Drug Saf.* 23 (8), 849–858. doi:10.1002/pds.3669
- van Puijenbroek, E. P., Bate, A., Leufkens, H. G., Lindquist, M., Orre, R., and Egberts, A. C. (2002). A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* 11 (1), 3–10. doi:10.1002/pds.668