biologit

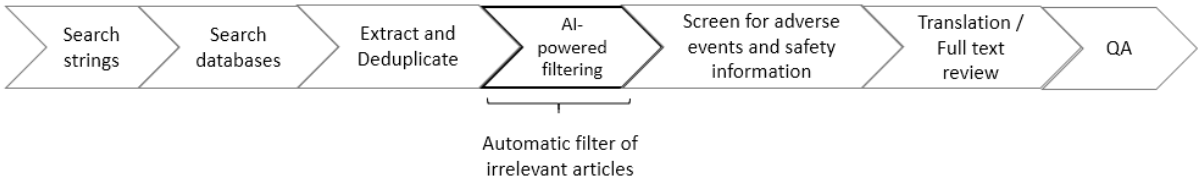
Model Fact Sheet

Suspect Adverse Event Detector | Rel. 1.1.14.4 | Feb 2023

1. Overview

Biologit MLM-AI is an integrated scientific literature monitoring platform for pharmacovigilance. It includes machine learning-based capabilities that help reduce the screening volume of scientific literature articles. This fact sheet describes the intended use and technical specification of biologit MLM-AI's main adverse detector model.

The machine learning model reflects how assessments are made by pharmacovigilance specialists following a typical literature monitoring process where an initial assessment is made by screening the title and abstract of a citation. By using predictions from the machine learning model as a first stage screening step, the overall screening volume requiring human inspection can be reduced to only the most relevant articles.



2. Intended Use

The system uses articles title and abstracts from the scientific literature as input and issues a prediction on whether the article contains one or more **suspect adverse events**. The prediction can be used to rank or filter abstracts before they reach human specialists for further screening. The following intended uses are envisaged for the suspect adverse prediction:

- Assist in the ranking and filtering of abstracts for the detection of individual case safety reports (ICSR) from the literature.
- Assist in the ranking and filtering of abstracts from the literature for aggregate safety data reporting or signal detection.

2.1 Target Domains

The system is intended to be used on article abstracts from the scientific literature, specifically on biomedical studies addressing a drug, compound or therapy to be used in human treatments.

2.2 Level of Supervision

To support different validation and risk management requirements of pharmacovigilance processes, the system supports various degrees of human supervision ranging from human-in-the-loop with full validation of results to higher levels of automation combined with human verification - ie. human-in-the-loop processes with the ability to audit and override AI decisions as needed.

The desired level of supervision is determined by customer requirements. Please consult [1,3] for further details on implementing various levels of supervision of the adverse model using biologit MLM-AI.

2.3 Adaptiveness

The suspect adverse model is trained following biologit's engineering and validation processes and uses a training dataset curated and labelled by biologit's in-house team of pharmacovigilance specialists. Any model and dataset updates follow the same process.

At present, the model does not dynamically learn from user input and can be considered a static model. Static AI models more closely follow current AI validation bets practice and can leverage existing methodologies for computer system validation [4].

2.4 Inputs and Outputs

Attribute	Expected Input
Input format	Title and abstract of a scientific article in plain text format (UTF8 encoding)
Domain	Biomedical text typical in academic articles from the scientific literature
Language	English

2.5 Operating Envelope

Several rule-based safeguards are in place to prevent the model from making predictions on conditions not observed in training. The following scenarios are automatically flagged as a suspect adverse event (so they are presented to users) irrespective of the model predictions:

- Abstracts in languages other than English.
- Articles with no abstract, containing only "errata" description, or containing only very short informative notes.
- Abstracts with several tokens greater than 1000 or less than 5.

Untested adjacent domains

Although the current model may present accurate results in domains related to human medicinal products, they are not yet represented in the training set. These domains may be included in future releases, we currently recommend their use coupled with extensive validation. Examples include medical devices, cosmetics and veterinary products.

3. Training Data

The data set is built from a selection of publicly available articles (title and abstracts) from the scientific literature. To achieve broad coverage articles were selected based using pre-defined keyword searches that (1) reflected compounds belonging to one or more MeSH categories for therapeutic uses, (2) were part of the European Medicines Agency list of products under surveillance for medical literature monitoring, or (3) were retrieved using commonly used adverse terms or terms for special situations (paediatric, off-label, pregnancy-related terms, etc).

3.1 DATASET STATISTICS

TRAINING SET	
- Total abstracts	24,444
- Label: Suspect adverse	9,254
- Label: Not suspect adverse	15,190
TEST+VALIDATION SET	
- Total abstracts	2,717
- Label: Suspect adverse	1,032
- Label: Not suspect adverse	1,685
Total abstracts	27,161
Test set % of total	10%

3.2 Data Labelling Protocol

The labelling protocol is designed to reflect a typical literature monitoring workflow, where a first-pass decision is made upon screening the abstract and title of a citation. An abstract is labelled as a suspected adverse event when (a) safety events are explicitly mentioned in the article abstract (applying to any product or treatment), or (b) contains implicit mentions of a safety event that may be fully described in the article full text (again independent of product or treatment).

All labelled data undergoes sampled quality check by a second annotator. The inter-annotator agreement is monitored and additional revisions are triggered should it fall below an accepted threshold.

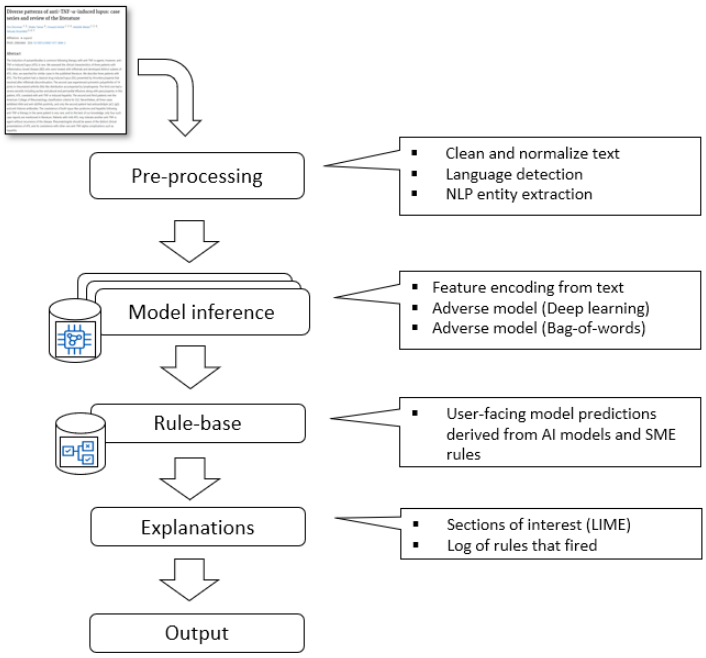
3.3 DATA SPLITS

The training set is built by randomly sampling 90% of all available labels (n=24,444) with the remaining 10% (n=2,717) further split evenly into validation and test sets. Sampling is stratified across the adverse and non-adverse labels. The training set (n=24,444) is used to train the core machine learning models; the validation set (n=1,358) is used for hyperparameter tuning. The test set (n=1,359) is not used in any training stages and is used to report model performance results.

4. Model Description

The inference pipeline supporting the adverse model comprises a pre-processing stage where input text is cleaned and tokenized. This stage also performs language detection and entity extraction (patient mentions, medication names, etc).

The model inference stage encodes the normalized text into features and runs the prediction step of the machine learning models. Next, a post-processing rules-based stage produces the final predictions and an explanation step computes additional metadata that can be used to help users understand the model output.



4.1 Machine Learning Models

The inference stage combines the output of two separate models:

- **Neural Model** - The neural model employs a multi-layer neural network architecture organized as follows: an initial embedding layer converts tokens into vector representations using a combination of pre-trained word embeddings built with a biomedical text corpus [5] and additional trainable embedding layers derived from part-of-speech tags and dependency parsing tags. The embeddings are combined and processed by a series of convolutional layers followed by an LSTM recurrent layer and an attention layer. Regularization is applied across the network architecture by using drop-out during training and the use of batch normalization layers.
- **Bag-of-words -** The neural model is supplemented by a bag-of-words model using 1-gram and 2-grams as features and trained with a random forest estimator.

Issuing Predictions

During the rule-based inference stage, the neural and bag-of-word model predictions are combined and subject to override rules authored in conjunction with pharmacovigilance experts. At present the following overrides are in place and will produce a suspect adverse prediction irrespective of input:

- Input is outside of operating envelope (see Section 2.5)
- Input contains an explicit identifiable patient mention (ex: “60-year-old female patient”, etc)

4.2 Performance Metric

Machine learning predictions for the literature monitoring of adverse events have an asymmetric risk profile: abstracts falsely identified as a safety event (false positive) incur incremental screening effort, while articles falsely identified as not a safety event (false negative) potentially mask a safety issue.

Therefore false negatives predictions are riskier and it is of paramount importance that this metric is minimized, even if at the expense of additional effort (more false positives).

To ensure the rate of false negatives remains low, the adverse event model is parametrized for a desired target recall level. With this value set, the goal is to minimize a metric that reflects the additional effort caused by false positives. We use the false positive rate, defined as the ratio of false positives (FP) to the number of ground truth negative examples (N) given by:

$$\frac{FP}{N} = \frac{FP}{FP + TN}$$

Where FP is the number of false positives and TN is the number of true negatives. Thus, the performance target is the minimization of the false positive rate at the desired target recall.

4.3 Experimental Results

Test set results tuned for a 95% desired recall are shown below. All metrics are for the suspect adverse class.

Metric (adverse class)	Value
Recall	96.5%
False Positive Rate	27.2%
Precision	68.4%
f1 score	0.8

Table 1- Test set performance (adverse class)

4.3 Development Approach and Validation

The approach for model development, testing and validation used by biologit follows guidance from [4] and other sources, and is publicly available in the technical paper:

- “Validation and Transparency in AI systems for pharmacovigilance: a case study applied to the medical literature monitoring of adverse events” – December 2021, arXiv pre-print.

5. References

[1] Achieving Faster Literature Screening with AI - biologit Blog (November 2020).

[2] Reducing screening workload in medical literature monitoring with machine learning - DIA Regulatory Science Forum (September 2020).

[3] AI-based screening workflows - biologit MLM-AI product documentation.

[4] Huysentruyt, K., Kjoersvik, O., Dobracki, P. et al. *Validating Intelligent Automation Systems in Pharmacovigilance*, Drug Safety v. 44 (2021) - https://doi.org/10.1007/s40264-020-01030-2

[5] SciSpacy – *SpaCy models for biomedical text processing* - https://allenai.github.io/scispacy/