A Machine Learning Approach to Classification of Drug Reviews in Russian

Ilseyar Alimova
Kazan (Volga Region) Federal University
Kazan, Russian Federation
Email: alimovailseyar@gmail.com

Elena Tutubalina
Kazan (Volga Region) Federal University
Kazan, Russian Federation
Email: ELVTutubalina@kpfu.ru

Julia Alferova
N.V. Sklifosovsky Research Institute of
Emergency Care, Moscow, Russian Federation

Guzel Gafiyatullina
Kazan (Volga Region) Federal University
Kazan, Russian Federation

***Abstract*—The automatic extraction of drug side effects from social media has gained popularity in pharmacovigilance. Information extraction methods tailored to medical subjects are essential for the task of drug repurposing and finding drug reactions. In this article, we focus on extracting information about side effects and symptoms in users' reviews about medications in Russian. We manually develop a real-world dataset by crawling user reviews from a health-related website and annotate a set of reviews on a sentence level. The paper addresses the classification problem with more than two classes, comparing a simple bag-of-words baseline and a feature-rich machine learning approach.**

***Index Terms*—drug side effects, text mining, social media, machine learning, Russian**

I. Introduction

One of the main tasks of pharmacology is drug repurposing. Drug repurposing is the use of existing drugs for the treatment of diseases which is not included initially in indications for use of this drug. This approach is cheaper and faster than developing a new drug since clinical trials are simplified in that case [1], [2]. According to the USA Food and Drug Administration[1], 279 drugs from 6733 were repurposed from 1998 to 2016 years. The example of such cases is the use of erythromycin antibiotic for the treatment of gastric motility disorder or use of an anti-emetic drug thalidomide for the treatment of multiple myeloma. For drug repurposing, it is necessary to identify new side effects from drugs, both beneficial and adverse effects. This is a challenging task since it is impossible to try all of existing drugs for each disease due to the lack of necessary information and the need for a large amount of time [3], [4]. In addition, the detection of new possible adverse drug reactions is also an important task, since unexpected adverse effects can harm the patient's health or life.

One of the methods for detecting new side effects for drugs is the analysis of the texts from social media with using natural language processing (NLP) methods [5], since patients may discuss their problems after taking drugs in various health-related forums and social networks. The studies of extracting new side effects from social media are mainly devoted to extracting new adverse drug reactions (ADR) [6], [7], [8], [9], [10], [11], [12], [13] and are seldom extends to detect beneficial effects of drugs (BNE) [14], [15]. Basically, these studies are conducted for English texts. In this work, we begin to fill this gap, providing first results for Russian health social media.

We created real world dataset by crawling the forum Otzovik and manually annotated each sentence of review's text with one of the labels: beneficial effect, adverse effect, symptom, and the other and considered this problem as a classification task. Classification methods can remove noise information automatically and detect if the text contains mentions of side effects or not. We also propose an approach for classifying sentences based on machine learning with a set of features that improve the classification results.

II. Related work

There are many studies on the extraction of adverse drug reactions from reviews in social networks, as evidenced by a number of reviews on this topic [5], [16], [17], [18]. In these studies, various approaches are used to identify adverse reactions in user reviews. The most widely used method is a dictionary-based approach [6], [7], [8], [10], [19], [20]. Dictionaries include adverse drug reactions extracted from instructions for the use of drugs, records of clinical trials and user reviews from health-related forums. The most well-known and frequently used dictionaries for English language are COSTART[2], CHV[3], MedEffect[4], UMLS[5], MedDRA[6], SIDER[7]. There are also rule-based methods [21], [22] that identify the most common constructs of sentences that may

[1] https://www.fda.gov/Drugs/

[2] http://bioportal.bioontology.org/ontologies/COSTART
[3] http://www.consumerhealthvocab.org/
[4] http://www.hc-sc.gc.ca/dhp-mps/medeff/index-eng.php
[5] http://www.nlm.nih.gov/research/umls/
[6] http://www.meddra.org/
[7] http://sideeffects.embl.de/

indicate a description of adverse drug reactions. Most of the studies describe approaches with using machine learning techniques. In [8], [14], [17], [23], [24], [25], [26], [27], authors applied Support Vector Machine (SVM) In [11], [12], [21], [28], Conditional Random Fields and random forest method were used. The most popular features for machine learning are n-grams, parts of speech tags, semantic types from UMLS, the number of negated contexts, the belonging lexicon-based features for ADRs, drug names, and word embeddings. Existing studies show that features based on sentiment analysis, subjectivity analysis, topic modeling, and polarity classification can be used to improve results of ADR detection [29]. Recently, several studies have employed convolutional neural networks for ADR classification [13], recurrent neural networks for extraction of ADRs [30], and a encoder-decoder network for ICD coding [31]. In [32], [33], [34], authors explored convolutional neural networks for predictions of demographic attributes based specifically on medical reviews.

For the development and testing of methods, there are several publicly available corpora were created: a corpus of adverse drug event annotations (CADEC) [35], a corpus of tweets [29], [36], a dataset of reviews about treatment for breast cancer [37], a corpus of clinical notes in Russian [38]. CADEC consists of AskaPatient (http://askapatient.com) medical forum posts in English. It includes the following annotations: Drug, Symptom, Adverse effect, Disease, and Finding. The Twitter corpus was developed by Diego Lab and contains 10,822 tweets about health in English. Each tweet was labeled whether tweets text contains the information about adverse drug reactions. In [37], authors annotated the dataset of user reviews for five commonly used breast cancer drugs from three health-related sites: askapatient.com, drugs.com, and drugratingz.com. The reviews were annotated for ADRs. Annotated corpus of clinical notes in Russian was created by Institute for System Analysis of Russian Academy of Sciences [38]. The corpus consists of clinical free-text notes in Russian of patients with allergic and pulmonary disorders. It includes the following annotations: Disease, Symptom, Drug, Treatment, Body Location, Severity, Course and related attributes such as Negation, Conditional, Effect, etc. However, there is no information about side effects experienced by patients during treatment.

There are a number of studies by Russian research groups on corpus-based linguistic resources in Russian or processing of medical texts. In [39], Solovyev and Ivanov defined a set of necessary linguistic resources for event extraction task from Russian texts. In [40], a system that extracts data on the patient's condition from the texts of clinical trials and results of analyses is described. The system use rule-based approach. The article [41] is devoted to the problem of extracting aspect terms from user reviews about medicines in Russian. The authors created the annotated corpus and applied neural networks for experiments. However, their system and the corpus are not described in detail and not publicly available.

III. DATA

We created the annotated dataset of patient reviews about drugs collected from the Otzovik forum[8]. The annotation was made by specialists in the field of medicine. The corpus is based on 580 reviews about antiviral, immunosuppressive, sedative, soporific drugs and nasal sprays. We based on annotation instruction from [14]. We splitted the texts of reviews on sentences with using Texterra system[9] [42] and marked each of them with one of a label: *Indication*, *Beneficial effect* (*BNE*), *Adverse drug reaction* (*ADR*), *Other*. The sentences with label *Indication* contain the symptoms and diseases of patients, that was the reason to take the drug. Indications contain the necessary information since the presence of various diseases in a patient could be the reason for side effects of taking a drug. *BNE* describes the effects of patient and recovery cases after taking a drug. Sentences with *ADR* labels describe deterioration in the state of patient health. *Other* sentences are not characterized as one of the above. For further evaluation, we removed sentences that could be annotated with two labels: *Indication* + *ADR*, *Indication* + *BNE* or *ADR* + *BNE* due to a small number of examples (69 cases). The statistics of the dataset and examples of sentences from different classes are presented in Table I and II. In future work, we plan to annotate the data on the entity-level and after that count the agreement.

Table I
THE NUMBER OF SENTENCES WITH EACH LABEL AND REVIEWS, CONTAINS SENTENCES WITH EACH LABEL (ONE REVIEW MAY CONTAIN SENTENCES FROM DIFFERENT CLASSES).

	Indication	**BNE**	**ADR**	**Other**	**Total**
Sentences	646	335	279	4 488	5 748
Reviews	409	239	185	579	580

Table II
THE EXAMPLES OF SENTENCES BELONGING TO DIFFERENT CLASSES WITH TRANSLATION TO ENGLISH.

Label	Example
Indication	Я пользуюсь свечами по 1 000 000 МЕ как для профилактики, так и для лечения ОРВИ.
	I use candles for 1 000 000 ME for both prevention and treatment of cold.
BNE	Прием мелаксена помог наладить сон.
	Taking of Melaksen helped to establish a sleep.
ADR	Стала спокойной даже чересчур, на работе стала тупить, коллеги сказали что я какая то заторможенная, все время клонит в сон.
	I became calm even too much, at work began to blunt, colleagues said that I am a little bit inhibited, want to sleep all the time.
Other	Время использования: месяц Стоимость: 150 руб.
	Duration of use: month. Cost: 150 rub.
	Общее впечатление : Снотворное донормил может вас удивить, но не всегда приятно
	General impression: sleeping pill Donormyl can surprise you, but not always pleasantly

[8]http://otzovik.com
[9]https://api.ispras.ru/products

IV. Approach for Sentence classification

We applied a machine learning model based on Linear SVM with different sets of features described below to classification task. The methods were implemented with LinearSVC class from the scikit-learn library [10] with parameter class weight='auto' due to the class disbalance in annotated data. We applied the following set of features:

- Bag of Words (BOW): we used unigrams and bigrams with tf-idf transformation.
- Part of Speech (pos): we identified the part of speech of each word with using Texterra and counted the number of nouns, verbs, adverbs, and adjectives.
- Word embedding (emb): we used ruscorpora vector representation from RusVectores resource[11] [43]. We calculated the average of the vectors for each token of a sentence and applied it as a feature.
- Sentiment feature (sent): we used RuSentiLex [44] lexicon and applied features for sentiment analysis described in [45].
- Polarity feature (pol): we defined the sentiment polarity of the sentence with using Texterra[12] and gave for positive sentences value 1, for negative value -1 and 0 for neutral polarity sentences as a feature.
- Disease lexicon (dis): we collected manually a list of diseases and symptoms from various resources, including medication instructions and used the presence of these diseases in a sentence as a feature.
- Drug name presence (drug): we collected the list of drug names from different internet resources and applied the presence of the name of the drug from these list in a sentence as a feature.
- Pointwise mutual information (pmi): we counted PMI for the large corpus of user reviews collected from Otzovik forum and used the PMI score, maximum and minimum values as the feature.

Word embedding vectors have a dimension 300 and obtained by training on Russian National Corpus[13] with using the Continuous Skip-Gram algorithm. It contains 184 973 words. In order to count PMI we used 80698 unlabeled reviews from forum otzovik.com preprocessed with using MyStem library[14]. Following state-of-the-art approaches for sentiment analysis, we compute a score for each token w with using 80698 unlabeled reviews from forum otzovik.com.

$$score(w) = PMI(w, pos) - PMI(w, neg),$$

$$PMI(w, neg) = \log \frac{p(w, neg)}{p(w) * p(neg)},$$

where PMI is the pointwise mutual information, pos and neg denote all reviews with star ratings greater or less than 3, respectively.

[10] http://scikit-learn.org/
[11] http://rusvectores.org/ru/
[12] https://api.ispras.ru/products
[13] http://www.ruscorpora.ru/index.htm
[14] https://tech.yandex.ru/mystem/

V. Results and Discussion

We performed pre-processing by lower-casing and lemmatization all words with using Texterra. We tested the method on the 5-folds cross-validation and computed macro-averaged recall (R), precision (P) and F1-measures (F). Table III presents the results of classification for 4 classes for different sets of features. According to the results, the most significant features are word2vec, polarity and disease lexicon. The combination of them obtained the maximum value of the macro F-measure 56.4%. We also tried different types of binary classification. The results for binary classification *BNE*, *ADR*, *Indication* versus *Other* are presented in Table IV. In that case, the most significant features are part of speech tags, word2vec, disease lexicon and PMI and the greatest value of F-measure is 73.3%. The results of other binary classifications with the best set of features are performed in Table V. We note that our classification algorithms may be limited in their performance on our highly unbalanced corpus. Class *Other* contains 78% of the total number of sentences.

Table III
5-fold cross-validation performances of Linear SVM model with different groups of features for 4 classes.

Features	Folds (F_1-measure)					av. F
	1	2	3	4	5	
bow	.567	.543	.537	.569	.502	.544
bow, pos	.564	.550	.542	.560	.509	.544
bow, dis	.556	.550	.550	.586	.524	.550
bow, w2v	.574	.553	.551	.551	**.551**	.559
bow, pol	.566	.545	.538	.538	.511	.547
bow, sent	.567	.546	.537	.569	.502	.545
bow, drug	.565	.543	.538	.570	.501	.544
bow, pmi	.555	.532	.535	.564	.508	.538
bow, dis, w2v	.556	.553	.542	.586	.541	.556
bow, dis, w2v, pos	.581	.545	.551	.589	.522	.558
bow, w2v, pos	**.588**	.553	.547	.566	.530	.557
bow, w2v, pol, dis	.564	.557	.552	.589	.538	.561
bow, w2v, pol, dis, pmi	.562	**.558**	**.558**	**.594**	**.551**	**.565**
all features	.582	.557	.547	.590	.540	.564

Table IV
5-fold cross-validation performances of our method with different set of features for binary classification (*Indication*, *BNE*, *ADR*) versus (Other).

Features	Folds (F_1-measure)					av. F
	1	2	3	4	5	
bow	.715	.713	.723	.719	.718	.718
bow, pos	.724	.726	.713	.711	.722	.720
bow, dis	.720	.724	.727	.720	**.741**	.727
bow, w2v	.724	.727	.724	.722	.723	.724
bow, pol	.711	.715	.724	.715	.719	.717
bow, sent	.713	.713	.722	.721	.718	.718
bow, drug	.716	.713	.725	.718	.719	.718
bow, pmi	.716	.714	.721	.724	.722	.719
bow, dis, w2v	.719	.731	.726	**.730**	.730	.728
bow, dis, w2v, pos	.733	**.733**	**.731**	.726	.732	.731
bow, dis, w2v, pos, pmi	**.739**	.730	**.731**	.724	.740	**.733**
bow, dis, w2v, pos, pmi, drug	.736	.730	.725	.728	.740	.732
all features	.733	.730	.728	.724	.744	.732

We investigated the effectiveness of different lexicons and resources, which was used to generate features. The lexicons

Table V

THE 5-FOLD CROSS-VALIDATION RESULTS FOR DIFFERENT TYPES OF BINARY CLASSIFICATIONS WITH THE BEST SETS OF FEATURES.

Method	Feature set	av. F1
(BNE, ADR) vs. (Indication, Other)	bow, pmi, dis, pos, w2v, pol	.702
	bow	.695
(BNE) vs.(ADR, Indication, Other)	bow, pmi, pos, w2v, pol	.681
	bow	.670
(ADR) vs. (BNE, Indication, Other)	bow, w2v, pmi	.680
	bow	.674
(Indication) vs. (ADR, BNE, Other)	bow, pmi, sent	.729
	bow	.727

Table VI

SUMMARY OF STATISTICS OF CONSIDERED LEXICONS.

Resources	RuSentiLex	Disease lexicon	Drug name list	Ruscorpora word2vec model	All
unique unigrams	849	61	41	3976	6209
unigrams	13466	820	329	34108	65505

coverage of corpus is presented in Table VI. As it can be seen the largest coverage has Ruscorpora word2vec model the others have a smaller coverage. This is the reason that the features based on these resources give a small increase in a task of classification. We also created domain-specific vector representation and compared its effectiveness with different models of word embedding from RusVectores resource. We collected 127 840 unlabeled user drug reviews from three resources: otzovik.com, meduniver.com, and rusmedserv.ru. Each review was lemmatized and lowercased. We used word2vec from Gensim library to train embeddings on the obtained dataset. We applied Continuous Bag of Words model with different parameters and selected the best set of them: vector size of 100, the length of the local context of 10, the negative sampling of 10, vocabulary cutoff of 5. We refer to our pre-trained vectors as RusDrugReviewsVec. The results of binary classification (*Indication*, *BNE*, *ADR)* versus (*Other*) for different word embedding models are presented in Table VII. The results are quite close to each other. The greatest value of F-measure was obtained with ruscorpora model and our vectors, however, the combination of applied features with ruscorpora vectors in classification task outperformed the results of combination with RusDrugReviewsVec.

We have also performed the qualitative analysis of the most representative n-grams for different classes.The results for *BNE* and *ADR* classes are presented in table VIII in descending order of feature weights. From the results, we can see that for *ADR* class the most significant features contain negation, diseases and the word 'again' indicating most likely the return of the disease. The most significant n-grams for *BNE* class contain words indicating the efficiency of the drug ('help', 'effective') and words with a positive tonality ('good', 'improve').

Table VII

THE EVALUATION OF DIFFERENT WORD EMBEDDING MODEL FROM RUSVECTORIZERS RESOURCE FOR BINARY CLASSIFICATION (*Indication*, *BNE*, *ADR*) VS. (*Other*) ON 5-FOLDS CROSS-VALIDATION.

	Dictionary	Vector dim.	av. F1
bow, ruwikiruscorpora	392339	300	.721
bow, web	267540	300	.722
bow, news	194058	300	.722
bow, ruscorpora	184973	300	**.724**
bow, ruscorpora 2017	173816	600	.722
bow, RusDrugReviewsVec	95013	100	**.724**

VI. CONCLUSION

In this work, we have made the first steps in the direction of extracting drug reactions and medical conditions from user-generated texts in Russian. We have presented the sentence level annotated dataset of user reviews about medications and the classification method based on supervised machine learning. In order to improve classification performance over a simple bag-of-words baseline, we have utilized different features including word embedding, disease terms, drug names, and sentiment lexicons. We have improved the results on 1.5% and 2.1% for binary and multi-class classification tasks, respectively. Our evaluation results indicate the need for a strategy of handling class imbalance and the development of domain-specific resources for analysis of medical texts in Russian. We have made our resources and the source code available at the GitHub repository[15]. Future work will focus on the development of linguistic resources in Russian. We believe that new corpora and medical dictionaries will lead to a better understanding of the informal language of patients on social media.

ACKNOWLEDGEMENTS

The work on problem definition, corpus creation, word embeddings, and paper preparation was supported by the Russian Science Foundation grant no. 15-11-10019. Other parts of this work were performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

[1] T. Oprea and J. Mestres, "Drug repurposing: far beyond new targets for old drugs," *The AAPS journal*, vol. 14, no. 4, pp. 759–763, 2012.

[2] J. L. Medina-Franco, M. A. Giulianotti, G. S. Welmaker, and R. A. Houghten, "Shifting from the single to the multitarget paradigm in drug discovery," *Drug discovery today*, vol. 18, no. 9, pp. 495–501, 2013.

[3] H.-M. Lee and Y. Kim, "Drug repurposing is a new opportunity for developing drugs against neuropsychiatric disorders," *Schizophrenia research and treatment*, vol. 2016, 2016.

[4] G. Jin and S. T. Wong, "Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines," *Drug discovery today*, vol. 19, no. 5, pp. 637–644, 2014.

[5] R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell, and M. Pirmohamed, "Social media and pharmacovigilance: a review of the opportunities and challenges," *British journal of clinical pharmacology*, vol. 80, no. 4, pp. 910–920, 2015.

[15] https://github.com/Ilseyar/adr_detection_russian

Table VIII
REPRESENTATIVE SVM FEATURES (LEMMATIZED) FOR TWO CLASSES (BINARY CLASSIFICATION TASK).

ADR		BNE	
ацикловир не	acyclovir no	быстродействующий	high-speed
впечатление ацикловир	impression of acyclovir	быстродействующий эффективный	high-speed effective
впечатление лекарство	impression of drug	он быстродействующий	it is high-speed
лечить калечить	treat maim	впечатление мягкий	soft impression
опять	again	мягкий седативный	mild sedative
зрение	vision	взрослый хорошо	good for adult
впечатление мы	impression we	препарат хорошо	the drug is good
думать это	think this	заболевание не	disease is not
впечатление аллергия	impression of an allergy	не разболодиться	do not get angry
не подходить	does not fit	расходиться	diverge
не помогать	not help	не расходиться	not diverge
эффективный вызывать	efficient	он только	it only
вызывать	cause	он действовать	it works
совсем не	absolutely not	отлично помогать	excellent help
потница	prickly heat	впечатление помогать	the impression of helping
это потница	this is a prickly heat	впечатление повышать	impression of raise
сильно помогать	to help a lot	эффективно работать	work effectively
небольшой температура	low temperature	улучшать	improve
быть небольшой	be small	не заболевать	do not get sick
мы совсем	we are absolutely	он помогать	he is helping

[6] A. Benton, L. Ungar, S. Hill, S. Hennessy, J. Mao, A. Chung, C. E. Leonard, and J. H. Holmes, "Identifying potential adverse effects using the web: A new approach to medical hypothesis generation," *Journal of biomedical informatics*, vol. 44, no. 6, pp. 989–996, 2011.

[7] C. C. Yang, H. Yang, L. Jiang, and M. Zhang, "Social media mining for drug safety signal detection," in *Proceedings of the 2012 international workshop on Smart health and wellbeing*. ACM, 2012, pp. 33–40.

[8] X. Liu and H. Chen, "Azdrugminer: an information extraction system for mining patient-reported adverse drug events in online patient forums," in *International Conference on Smart Health*. Springer, 2013, pp. 134–150.

[9] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on twitter? mining tweets for adverse drug reactions," in *AMIA annual symposium proceedings*, vol. 2014. American Medical Informatics Association, 2014, p. 924.

[10] C. C. Yang, H. Yang, and L. Jiang, "Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media," *ACM Transactions on Management Information Systems (TMIS)*, vol. 5, no. 1, p. 2, 2014.

[11] E. Aramaki, Y. Miura, M. Tonoike, T. Ohkuma, H. Masuichi, K. Waki, and K. Ohe, "Extraction of adverse drug effects from clinical records." in *MedInfo*, 2010, pp. 739–743.

[12] M. Rastegar-Mojarad, R. K. Elayavilli, Y. Yu, and H. Liu, "Detecting signals in noisy data-can ensemble classifiers help identify adverse drug reaction in tweets," in *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing*, 2016.

[13] T. Huynh, Y. He, A. Willis, and S. Rüger, "Adverse drug reaction classification with deep neural networks," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 877–887.

[14] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez, "Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks," in *Proceedings of the 2010 workshop on biomedical natural language processing*. Association for Computational Linguistics, 2010, pp. 117–125.

[15] A. Cassels, M. A. Hughes, C. Cole, B. Mintzes, J. Lexchin, and J. P. McCormack, "Drugs in the news: an analysis of canadian newspaper coverage of new prescription drugs," *Canadian Medical Association Journal*, vol. 168, no. 9, pp. 1133–1137, 2003.

[16] R. Harpaz, A. Callahan, S. Tamang, Y. Low, D. Odgers, S. Finlayson, K. Jung, P. LePendu, and N. H. Shah, "Text mining for adverse drug events: the promise, challenges, and state of the art," *Drug safety*, vol. 37, no. 10, pp. 777–790, 2014.

[17] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, "Utilizing social media data for pharmacovigilance: A review," *Journal of biomedical informatics*, vol. 54, pp. 202–212, 2015.

[18] J. Lardon, R. Abdellaoui, F. Bellet, H. Asfari, J. Souvignet, N. Texier, M.-C. Jaulent, M.-N. Beyens, A. Burgun, and C. Bousquet, "Adverse drug reaction identification and extraction in social media: a scoping review," *Journal of medical Internet research*, vol. 17, no. 7, 2015.

[19] S. Yeleswarapu, A. Rao, T. Joseph, V. G. Saipradeep, and R. Srinivasan, "A pipeline to extract drug-adverse event pairs from multiple data sources," *BMC medical informatics and decision making*, vol. 14, no. 1, p. 13, 2014.

[20] C. C. Freifeld, J. S. Brownstein, C. M. Menone, W. Bao, R. Filice, T. Kass-Hout, and N. Dasgupta, "Digital drug safety surveillance: monitoring pharmaceutical products in twitter," *Drug safety*, vol. 37, no. 5, pp. 343–350, 2014.

[21] A. Nikfarjam and G. H. Gonzalez, "Pattern mining for extraction of mentions of adverse drug reactions from user comments," in *AMIA Annual Symposium Proceedings*, vol. 2011. American Medical Informatics Association, 2011, p. 1019.

[22] J.-C. Na, W. Y. M. Kyaing, C. S. Khoo, S. Foo, Y.-K. Chang, and Y.-L. Theng, "Sentiment classification of drug reviews using a rule-based linguistic approach," in *International Conference on Asian Digital Libraries*. Springer, 2012, pp. 189–198.

[23] Y. Niu, X. Zhu, J. Li, and G. Hirst, "Analysis of polarity information in medical text," in *AMIA annual symposium proceedings*, vol. 2005. American Medical Informatics Association, 2005, p. 570.

[24] B. W. Chee, R. Berlin, and B. Schatz, "Predicting adverse drug events from personal health messages," in *AMIA Annual Symposium Proceedings*, vol. 2011. American Medical Informatics Association, 2011, p. 217.

[25] J. Bian, U. Topaloglu, and F. Yu, "Towards large-scale twitter mining for drug-related adverse events," in *Proceedings of the 2012 international workshop on Smart health and wellbeing*. ACM, 2012, pp. 25–32.

[26] M. Yang, X. Wang, and M. Y. Kiang, "Identification of consumer adverse drug reaction messages on social media." in *PACIS*, 2013, p. 193.

[27] I. Alimova and E. Tutubalina, "Automated detection of adverse drug reactions from social media posts with machine learning," in *Proceedings of International Conference on Analysis of Images, Social Networks and Texts*, 2017, pp. 1–12.

[28] Z. S. Miftahutdinov, E. Tutubalina, and A. Tropsha, "Identifying disease-related expressions in reviews using conditional random fields," in *Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog*, vol. 1, 2017, pp. 155–167.

[29] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training," *Journal of biomedical informatics*, vol. 53, pp. 196–207, 2015.

[30] E. Tutubalina and S. Nikolenko, "Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug

reactions from user reviews," *Journal of Healthcare Engineering*, vol. 2017, 2017.
[31] Z. Miftakhutdinov and E. Tutubalina, "Kfu at clef ehealth 2017 task 1: Icd-10 coding of english death certificates with recurrent neural networks." CLEF, 2017.
[32] E. Tutubalina and S. Nikolenko, "Automated prediction of demographic information from medical user reviews," in *International Conference on Mining Intelligence and Knowledge Exploration*. Springer, 2016, pp. 174–184.
[33] ——, "Exploring convolutional neural networks and topic models for user profiling from drug reviews," *Multimedia Tools and Applications*, Nov 2017. [Online]. Available: https://doi.org/10.1007/s11042-017-5336-z
[34] ——, "Demographic prediction based on user reviews about medications," *Computación y Sistemas*, vol. 21, no. 2, 2017.
[35] S. Karimi, A. Metke-Jimenez, M. Kemp, and C. Wang, "Cadec: A corpus of adverse drug event annotations," *Journal of biomedical informatics*, vol. 55, pp. 73–81, 2015.
[36] R. Ginn, P. Pimpalkhute, A. Nikfarjam, A. Patki, K. O'Connor, A. Sarker, K. Smith, and G. Gonzalez, "Mining twitter for adverse drug reaction mentions: a corpus and classification benchmark," in *Proceedings of the fourth workshop on building and evaluating resources for health and biomedical text processing*, 2014.
[37] A. Yates and N. Goharian, "Adrtrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites," in *European Conference on Information Retrieval*. Springer, 2013, pp. 816–819.
[38] A. Shelmanov, I. Smirnov, and E. Vishneva, "Information extraction from clinical texts in russian," in *Computational Linguistics and Intellectual Technologies: Annual International Conference "Dialog"*, vol. 17, 2015.
[39] V. Solovyev and V. Ivanov, "Knowledge-driven event extraction in Russian: corpus-based linguistic resources," *Computational intelligence and neuroscience*, vol. 2016, p. 16, 2016.
[40] M. V. Lapayev, A. B. Vodyakho, Aleksandr Ivanovichand Smirnov, and N. A. Zhukova, "Sistema obrabotki tekstovykh meditsinskikh dannykh."
[41] Y. D. Izotova and D. S. Tarasov, "Izvlecheniye farmatsevticheski znachimykh aspektnykh terminov model'yu rekurrentnoy neyronnoy seti iz tekstov na yestestvennom yazyke pri malykh vyborkakh," in *NEYROINFORMATIKA-2016*, 2016, pp. 65–73.
[42] D. Y. Turdakov, N. A. Astrakhantsev, Y. R. Nedumov, A. A. Sysoev, I. A. Andrianov, V. D. Mayorov, D. G. Fedorenko, A. V. Korshunov, and S. D. Kuznetsov, "Texterra: A framework for text analysis," *Programming and Computer Software*, vol. 40, no. 5, pp. 288–295, 2014.
[43] A. Kutuzov and E. Kuzmenko, *WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models*. Cham: Springer International Publishing, 2017, pp. 155–161. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-52920-2_15
[44] N. V. Loukachevitch and A. Levchik, "Creating a general russian sentiment lexicon." in *LREC*, 2016.
[45] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *Journal of Artificial Intelligence Research*, vol. 50, pp. 723–762, 2014.