IberSPEECH 2021
24-25 March 2021, Valladolid, Spain

Adverse Drug Reaction extraction on Electronic Health Records written in Spanish: A PhD Thesis overview

Sara Santiso González

IXA group, University of the Basque Country (UPV/EHU)
Manuel Lardizabal 1, 20018, Donostia
sara.santiso@ehu.eus

Abstract

The aim of this work is the automatic extraction of Adverse Drug Reactions (ADRs) in Electronic Health Records (EHRs) written in Spanish. From Natural Language Processing (NLP) perspective, this is approached as a relation extraction task in which the drug is the causative agent of a disease, the adverse reaction. This would help to increase the reporting of ADRs and their earliest possible detection, helping to improve the health of the patients.

ADR extraction from EHRs involves major challenges. First, drugs and diseases found in an EHR are often unrelated or sometimes related as treatment, but seldom as ADRs. This implies the inference of a predictive model from samples with skewed class distribution. Second, EHRs contain both standard and nonstandard abbreviations and misspellings. All this leads to a high lexical variability. Third, the Spanish count with few resources and tools to apply NLP. To cope with these challenges, we explored several ADR detection algorithms (Random Forest and Joint AB-LSTM) and representations (symbolic and dense) to characterize the ADR candidates. In addition, we assessed the tolerance of the ADR detection model to external noise such as the incorrect detection of the medical entities involved in the ADR extraction.

Index Terms: Adverse Drug Reactions, Electronic Health Records, Text mining, Supervised machine learning

1. Introduction

An ADR is defined by the World Health Organization (WHO) as ‘a response to a medicine which is noxious and unintended, and which occurs at doses normally used in man’ [1]. The WHO informed about the importance of reporting ADRs to understand and treat the diseases caused by drugs and, as a result, improve the patients care [2]. However, ADRs are still heavily under-reported, which makes their prevention difficult. This was the **motivation** to automatically extract ADRs on Electronic Health Records (EHRs). Given that information stored digitally by the hospitals is growing, Natural Language Processing (NLP) techniques can be used to create a system that helps the doctors to analyze the ADRs of the patients in a given EHR, facilitating the decision making process and alleviating the work-load. As a consequence, the patients’ health could improve and the pharmaco-surveillance service would be informed about the detected ADRs. The ADR extraction was defined as a relation extraction task. That is, the aim is to detect ADR relations between the entities (drugs and diseases) recognized in a given text. For the ADR extraction developed in this work, we distinguished the two steps involved in this task, which were developed using a pipeline approach:

1. Medical Entity Recognition (MER) to find “drug” entities and “disease” entities. The “drug” entity encompasses either a brand name, a substance or an active ingredient and the “disease” entity encompasses either a disease, a sign or a symptom.
2. ADR detection to discover the relations between “drug” entities and “disease” entities that correspond to ADRs. The “drug“ entity would be the causative agent and the “disease” entity would be the caused adverse reaction.

In the ADR extraction process, we had to overcome some **challenges** that make this supervised classification task difficult. On the one hand, the ADRs are minority relations because generally the drug and the disease are either unrelated or related as treatment and, thus, the ADRs are rare cases. This implies the inference of a predictive model from samples with skewed class distribution. On the other hand, the EHRs show multiple lexical variations. EHRs are written by experts under time pressure, employing rich medical jargon together with colloquial expressions, not always grammatical, and it is not infrequent to find misspellings and both standard and nonstandard abbreviations. In addition, ***our EHRs are written in Spanish whereas the majority of biomedical NLP research has been done in English***. The Spanish and other languages different to English count with few resources and tools to apply NLP in the medical domain. In this line, it is remarkable the recent interest in developing NLP tools for languages other than English [3]. To cope with these challenges, we explored several ADR detection algorithms, Random Forest (RF) [4] and Joint Attentive Bidirectional Long Short-Term Memory (AB-LSTM) [5], and representations to characterize the ADR candidates, symbolic and dense.

The main **objective** of this work is the creation of a model able to detect automatically ADRs in EHRs written in Spanish. This, in turn, encompasses the sub-objectives stated below:

- Detect ADRs by discovering relations between the causative drug and the caused diseases.

  The aim is to detect drug-disease pairs related as ADRs and not only the disease caused by the drug. Indicating explicitly the entities involved in an ADR can result more useful for their study.

- Discover approaches to overcome the class imbalance.

  Given that ADRs are rare events, it is frequent to find the class imbalance problem in this task. Machine learning algorithms tend to expect balanced class distributions and learning the minority class is difficult for them. For

This manuscript presents a summary of the thesis developed by Sara Santiso and supervised by Arantza Casillas and Alicia Pérez. It was defended on June 13 2019 obtaining excellent grade with Cum Laude mention.

10.21437/IberSPEECH.2021-34

this reason, our intention is to explore different techniques that could help to tackle this issue improving the ADR detection or find approaches that could be robust against imbalanced distributions of the class.

- Discover robust representations to cope with the lexical variability and the data sparsity.

  This is a challenge goal due to two factors. First, the EHRs are written during consultation time and each doctor uses different terms or expressions, producing lexical variations. Second, due to confidentiality issues, there is a lack of available EHRs. Then, our intention is to explore different representations in order to make the most of the annotated corpus.

2. Related work

One of the differentiating factors in related works dealing with ADR extraction is the definition of the task itself. In our case, it is approached as a relation extraction task between drugs and diseases (the adverse reactions). From the NLP perspective this consists in the extraction of cause-effect events between the entities. There are relevant works following this approach [6, 7, 8, 9]. Alternatively, some authors considered ADR extraction as the identification of the caused disease (the adverse reaction), that is, a sub-class of MER [10, 11, 12, 13] or refer to ADR extraction as the detection of presence (or absence) of ADRs in a document [14, 15, 16].

ADR extraction was applied to several textual genres such as social media, scientific publications or EHRs. Among the works developed with EHRs we distinguished three approaches. The earliest attempts made use of symbolic features (e.g. word-forms, lemmas and POS tags) to represent the ADRs and employed traditional classifiers [17, 18]. Next, with the growing distributional semantics, word-embeddings were introduced and main trends in characterizations turned from symbolic into dense spaces [19]. Finally, deep neural networks re-emerged as state-of-the-art classification approaches [20, 21].

For Spanish, we found the SpanishADRCorpus [10] labeled with drugs and effects as entities and drug indications and adverse drug reactions as relations. It is composed by 400 documents gathered from ForumClinic, a health network website in Spanish. Note that this corpus was employed for ADR extraction with techniques based on rules and unsupervised methods [22, 23, 24].

3. Methods

3.1. Corpora

In this work we employed three annotated corpora that involve EHRs written in Spanish from two hospitals within Osakidetza. The gold standard corpus (IxaMed-GS) contains 75 EHRs from the Galdakao hospital [25]. The cross hospital corpus (IxaMed-CH) contains 267 EHRs from the Galdakao and Basurto hospitals. It contains the EHRs of IxaMed-GS. The extended corpus (IxaMed-E) contains 463 EHRs from the Galdakao and Basurto hospitals. It contains some EHRs of IxaMed-CH but not of IxaMed-GS. In order to infer and evaluate the models, each corpus was divided in train, development and test sets randomly selected without replacement. Parameters were tuned by training with the train set and evaluating on the dev set (train vs dev). With those parameters the model was trained with the union of the train and dev sets and evaluated on the test set (train∪dev vs test). While the positive ADR relations were those manually annotated by the experts, the negative relations were created by combining all the Disease group and Allergy entities with all the Drug group entities present in each document. Table 1 shows the quantitative description of these corpora: number of documents, word-forms, vocabulary, Out-Of-Vocabulary (OOV) words and medical entities that the experts manually tagged together with the number of ADR relations of each class. Note that the OOVs are words of the evaluation set that were not seen in the training set. The OOV of the Dev set are calculated with respect to the vocabulary of the Train set and the OOV of the Test set are calculated with respect to the vocabulary of the Train and Dev sets.

Table 1: *Quantitative description of the corpora. Positive relations ($\oplus$) refer to ADRs while negative relations ($\ominus$) refer to non-ADRs.*

Corpus			Partition		
			Train	**Dev**	**Test**
IxaMed-GS	**Documents**		41	17	17
	Word-forms		20,689	11,246	9,698
	Vocabulary		4,934	–	–
	OOV		–	1,526	979
	Entities	**Drug**	503	346	354
		Disease	1,341	737	629
	Relations	$\oplus$	53	30	27
		$\ominus$	231	134	173
IxaMed-CH	**Documents**		157	55	55
	Word-forms		91,088	34,004	33,171
	Vocabulary		13,809	–	–
	OOV		–	2,628	2,280
	Entities	**Drug**	2,436	943	887
		Disease	6,828	2,328	2,473
	Relations	$\oplus$	197	79	62
		$\ominus$	2,162	366	559
IxaMed-E	**Documents**		279	92	92
	Word-forms		138,695	47,487	43,858
	Vocabulary		18,003	–	–
	OOV		–	3,182	2,735
	Entities	**Drug**	3,474	1,128	1,122
		Disease	10,894	3,831	3,387
	Relations	$\oplus$	332	113	82
		$\ominus$	12,877	5,312	3,756

3.2. ADR detection approaches

Firstly, we used three approaches for ADR extraction, with alternative characterizations and classification algorithms, using the IxaMed-GS corpus.

- ADR detection with symbolic representations and RF

  We explored symbolic characterizations with the traditional classifier RF [26, 27, 28, 29, 30]. First, we tackled the detection of intra-sentence as well as inter-sentence ADRs. In order to overcome the class imbalance we tried different techniques: sampling, cost-sensitive learning, ensemble learning and one-class classification. Finally, we restricted the task to the same sentence to reduce the level of imbalance. Note that we also explored two approaches to detect negated entities automatically [31, 32]. These would be used to discard negative ADR candidates.

- ADR detection with dense representations and RF

  We explored dense characterizations created from embeddings that were used together with the RF classifier overcoming the class imbalance [33]. Instead of using an embedding for each word, we built a single context-aware embedding (dense representation created taking into account the embeddings of the context-words). Furthermore, we proposed different smoothing techniques that were applied to the dense representations to improve the proximity between semantically related words. These techniques are direction cosines, truncation, Principal Component Analysis (PCA) and clustering.

- ADR detection with dense representations and Joint AB-LSTM

  We used dense characterization automatically inferred by the Joint AB-LSTM network [34]. Specifically, in the Joint AB-LSTM two Bi-LSTM are trained. We compared different pooling strategies such as max, average and attention pooling separately and also their combinations. We explored the use of word-forms and lemmas as core-features. Furthermore, we explored the techniques to overcome the class imbalance suited for neural networks (re-sample, re-sample per batch and cost-sensitive learning).

3.3. Tolerance of ADR detection to noise

Finally, with the best performing approach, we used different corpora (IxaMed-GS, IxaMed-CH and IxaMed-E) with a higher number of examples and slightly variations in the sub-domains (EHRs from different hospitals with different services or specializations). In addition, we also analyzed the influence in the ADR extraction of a real system for the automatic detection of medical entities. Specifically, we employed Conditional Random Fields (CRF) [35] as classifier.

4. Results

4.1. Results for the ADR detection approaches

Developing the ADR detection with symbolic representations and RF, the best results were obtained at sentence level with the application of re-sample. Restricting the ADR detection to sentence level alleviated drastically the class imbalance problem, reducing the imbalance ratio from 1:222 to 1:4. Among all the hand-crafted features used in our symbolic representation, the 20 most relevant features for the intra-sentence scope were the word-forms and lemmas of the entities and their contexts. By contrast, the distances are the most relevant ones when inter- and intra-sentence scope is considered.

With dense representations and RF the best performing model was created with the application of director cosines, truncation and PCA. The embeddings were generated with a corpus of EHRs using GloVe. We compared the results obtained using the concatenation of words with those obtained replacing the words by their corresponding embeddings and it led to significant improvements. We also observed that the smoothing techniques outperformed their corresponding non-smoothed counterpart.

In the case of dense representations and Joint AB-LSTM, the best results were obtained without tackling the class imbalance, using a lemmatized version of the embeddings, Batch Normalization and combining max pooling and attentive pooling. We used a Feed Forward Neural Network (FFNN) as baseline, a simplified version of the Joint AB-LSTM that skips the Bi-LSTM layer. This was outperformed by the Joint AB-LSTM. We found that Batch Normalization was helpful and lemmatization was effective. According to the results, it seemed as if max and attention pooling complimented each other.

Table 2 shows these results, which demonstrate that the dense representation resulted useful since the f-measure of the positive class obtained with the symbolic representation improved in all the cases. Furthermore, the abstract representation automatically inferred by the Joint AB-LSTM was even better. For more detailed see [30, 34, 34].

Table 2: *Results of each best performing approach (symbolic + RF, dense + RF, dense + Joint AB-LSTM) for the IxaMed-GS corpus.*

		train vs dev			**train$\cup$dev vs test**			Class
		P	**R**	**F**	**P**	**R**	**F**	
symbolic	RF	54.5	40.0	46.2	34.0	59.3	43.2	$\oplus$
		87.3	92.5	89.9	92.8	82.1	87.1	$\ominus$
		81.3	82.9	81.9	84.9	79.0	81.2	W. Avg.
		82.9	82.9	82.9	79.0	79.0	79.0	Micro Avg.
		70.9	66.3	68.0	63.4	70.7	65.2	Micro Avg.
dense	RF	54.8	76.7	63.9	47.4	66.7	55.4	$\oplus$
		94.3	85.8	89.8	94.4	88.4	91.3	$\ominus$
		87.0	84.1	85.1	88.1	85.5	86.5	W. Avg.
		84.1	84.1	84.1	85.5	85.5	85.5	Micro Avg.
		74.5	81.2	76.9	70.9	77.6	73.4	Micro Avg.
	Joint AB-LSTM	87.2	67.8	**76.3**	72.4	71.4	**71.9**	$\oplus$
		93.2	97.8	95.4	95.3	95.5	95.4	$\ominus$
		92.1	92.3	91.9	92.0	92.1	92.0	W. Avg.
		92.3	92.3	92.3	92.1	91.8	92.7	Micro Avg.
		90.2	82.8	85.8	83.8	83.4	83.6	Micro Avg.

4.2. Results for the tolerance of ADR detection to noise

Analyzing the tolerance of ADR detection to noise we observed that despite of increasing the class imbalance and the sub-domains, the results improved as the size of the corpus increased. This is shown in Table 3, where the best results were obtained with IxaMed-E, the largest corpus.

In addition, Table 4 gives the results obtained with CRF as entity recognizer. The f-measure of the positive class suggests that ADR relations are scarce and missing some entities does not have an impact in the results while they are important in clinical practice.

5. Discussion

In this work we made a step ahead in the development of NLP methods that deal with ADR extraction defined as relation extraction task between a causative drug and the adverse reaction. We observed that the combination of approaches to tackle the high class imbalance, precisely sampling and cost-sensitive learning, was beneficial in the context of inter- and intra-sentence ADR extraction. We also observed that class imbalance can be, somehow, tackled in intra-sentence ADR extraction.

We experimentally corroborated that to deal with lexical variability, context-aware embeddings are useful to preserve the lexical nuances in this domain. Furthermore, to alleviate the influence that the lack of training samples might have in the

Table 3: *Results of the best performing approach (dense + Joint AB-LSTM) with each corpus (IxaMed-GS, IxaMed-CH, IxaMed-E).*

	train vs dev			**train∪dev vs test**			Class
	P	**R**	**F**	**P**	**R**	**F**	
IxaMed-GS	87.2	67.8	76.3	72.4	71.4	71.9	⊕
	93.2	97.8	95.4	95.3	95.5	95.4	⊖
	92.1	92.3	91.9	92.0	92.1	92.0	W. Avg.
	92.3	92.3	92.3	92.1	91.8	92.7	Micro Avg.
	90.2	82.8	85.8	83.8	83.4	83.6	Micro Avg.
IxaMed-CH	89.3	69.2	77.9	76.0	70.9	73.3	⊕
	94.1	98.3	96.2	96.1	96.9	96.5	⊖
	93.2	93.4	93.1	93.7	93.8	93.7	W. Avg.
	93.4	93.4	93.4	93.8	93.8	93.8	Micro Avg.
	91.6	83.7	87.0	86.1	83.9	84.9	Micro Avg.
IxaMed-E	90.3	71.8	**79.9**	74.4	76.0	**75.2**	⊕
	94.7	98.5	96.6	96.5	96.2	96.3	⊖
	94.0	94.2	93.9	93.7	93.6	93.7	W. Avg.
	94.2	94.2	94.2	93.6	93.6	93.6	Micro Avg.
	92.6	85.2	88.3	85.4	86.1	85.8	Micro Avg.

Table 4: *Results of the best performing approach (dense + Joint AB-LSTM) with the IxaMed-E corpus, evaluated using the automatic entities.*

train vs dev			**train∪dev vs test**			Class
P	**R**	**F**	**P**	**R**	**F**	
96.4	60.7	74.5	86.2	53.1	65.7	⊕
92.9	99.5	96.1	93.6	98.8	96.1	⊖
93.5	93.3	92.6	92.6	93.0	92.3	W. Avg.
93.3	93.3	93.3	93.0	93.0	93.0	Micro Avg.
94.7	80.1	85.3	89.9	75.9	80.9	Micro Avg.

quality of the inferred dense representations, we proposed the use of smoothing techniques. We observed that dense spaces of lemmas also helped to tackle the lexical variability. In fact, lemmatization was particularly effective in the neural networks used for ADR extraction.

In addition, we corroborated that the Joint AB-LSTM is able to cope with these types of noise although, naturally, there is a small decrease in its performance due to the missed entities involved in the ADR pairs.

The main **contribution** of this work is that the ADR extraction was developed using EHRs written in Spanish. To the best of our knowledge, for ADR extraction in texts written in Spanish, we are the first employing EHRs. In related works we observed that there is a corpus written in Spanish, SpanishADR-Corpus [10], that was employed to develop several works for ADR detection. However, this corpus is not composed by EHRs and the authors employed techniques based on rules and unsupervised methods instead of supervised learning.

6. Conclusions and future work

ADRs are rare events, then, supervised classification algorithms tend to be biased and learning to predict the minority class is complex. The application of approaches to overcome the **class imbalance** improves the performance of the ADR detection model to find **inter-** as well as **intra-sentence** ADRs. However, the results are considerably better in the intra-sentence scope than in the inter-sentence scope.

A key issue in the extraction of ADRs is the operative **characterization** of events. With regard to initial symbolic characterizations, if both inter-sentence and intra-sentence relations are taken into account, features related to the distances between the entities involved result relevant for the task. If the ADR detection is focused on intra-sentence ADRs, the word-forms and the lemmas of the entities and their contexts are more relevant. NLP rapidly evolved towards dense characterizations. Dense representations have the strength of exploiting semantic relatedness in dense low dimensional spaces. This is an important factor in our task to cope with lexical variability. We corroborated that dense representations outperform symbolic ones and it seemed as if the model gains generalization ability.

Another important factor is the **classification approach**. In this work we compared a traditional supervised classification approach (RF) and an emerging technique based on deep neural networks (Joint AB-LSTM) and found that Joint AB-LSTM outperformed RF. We speculated about the reasons behind. An outstanding difference between traditional and neural approaches rests on the generation of the inherent characterization for the instances. While traditional approaches make use of hand-crafted features (either in their symbolic or embedded as dense representations), neural approaches infer, automatically, abstract features. Nevertheless, we found that FFNN did not outperform the RF when the instances were characterized with smoothed embeddings. Our hypothesis to explain that Joint AB-LSTM outperform RF is that the information captured from the context is crucial in relation extraction. While RF exploits the context in a static way, Joint AB-LSTM can leverage the context dynamically. Furthermore, we observed, empirically, that Joint AB-LSTM networks are less sensitive to class imbalance than RF.

Variations in the size and domain of the corpus have an **effect in the performance** of the ADR detection model. To be precise, the larger the corpus the better the results. Regarding the variations associated to different sub-domains introduced by the use of EHRs of different hospitals, Joint AB-LSTM resulted robust. Needless to say, the errors propagated from the MER step affect the ADR detection. Missing entities lead to undiscovered relations. However, the drop in performance is not as dramatic as we expected.

ADR extraction is a major issue in pharmaco-surveillance and documentation. So far, the systems tend to focus on the detection of drug-disease pairs located in the same sentence (intra-sentence relations). Nevertheless, EHRs have implicit information that might reveal underlying relations (e.g. information in the antecedents might be relevant to guess the causes for an adverse event). That is, as **future work** an effort should be made to detect inter-sentence relations both, explicitly and implicitly stated. Moreover, we can work in developing entity recognition and relation extraction simultaneously, using a joint model to avoid the propagation of pipeline errors.

7. Acknowledgements

The authors would like to thank the staff of the Pharmacy and Pharmacovigilance services of the Galdakao-Usansolo and Basurto hospitals. This work was partially funded by the Spanish Ministry of Science and Innovation (PROSAMED: TIN2016-77820-C3-1-R) and the Basque Government (BERBAOLA: KK-2017/00043, Predoctoral Grant: PRE 2018 2 0265).

8. References

[1] World Health Organization, "Safety of medicines: a guide to detecting and reporting adverse drug reactions: why health professionals need to take action," pp. 1–16, 2002.

[2] ——, "The importance of pharmacovigilance," pp. 1–52, 2002.

[3] A. Névéol, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, "Clinical natural language processing in languages other than English: opportunities and challenges," *Journal of Biomedical Semantics*, vol. 9, no. 1, pp. 1–13, 2018.

[4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] S. K. Sahu and A. Anand, "Drug-drug interaction extraction from biomedical texts using long short-term memory network," *Journal of Biomedical Informatics*, vol. 86, pp. 15–24, 2018.

[6] H. Gurulingappa, A. Mateen-Rajpu, and L. Toldo, "Extraction of potential adverse drug events from medical case reports," *Journal of Biomedical Semantics*, vol. 3, no. 1, pp. 1–10, 2012.

[7] F. Li, D. Ji, X. Wei, and T. Qian, "A transition-based model for jointly extracting drugs, diseases and adverse drug events," in *2015 IEEE International Conference on Bioinformatics and Biomedicine*, 2015, pp. 599–602.

[8] J. Legrand, Y. Toussaint, C. Raïssi, and A. Coulet, "Syntax-based transfer learning for the task of biomedical relation extraction," in *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, 2018, pp. 149–159.

[9] B. He, Y. Guan, and R. Dai, "Classifying medical relations in clinical text via convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 93, pp. 43–49, 2019.

[10] I. Segura-Bedmar, R. Revert, and P. Martínez, "Detecting drugs and adverse events from Spanish social media streams," in *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*, April 2014, pp. 106–115.

[11] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *Journal of the American Medical Informatics Association*, vol. 22, no. 3, pp. 671–681, 2015.

[12] G. Stanovsky, D. Gruhl, and P. Mendes, "Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, vol. 1, April 2017, pp. 142–151.

[13] S. Gupta, S. Pawar, N. Ramrakhiyani, G. K. Palshikar, and V. Varma, "Semi-supervised recurrent neural network for adverse drug reaction mention extraction," *BMC Bioinformatics*, vol. 19, no. 8, pp. 1–7, 2018.

[14] I. Karlsson, J. Zhao, L. Asker, and H. Boström, "Predicting adverse drug events by analyzing electronic patient records," in *Conference on Artificial Intelligence in Medicine in Europe*, 2013, pp. 125–129.

[15] J. Zhao, A. Henriksson, L. Asker, and H. Boström, "Predictive modeling of structured electronic health records for adverse drug event detection," *BMC Medical Informatics and Decision Making*, vol. 15, no. 4, pp. 1–15, 2015.

[16] S. Friedrich and H. Dalianis, "Adverse drug event classification of health records using dictionary based pre-processing and machine learning," in *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, 2015, pp. 121–130.

[17] 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.

[18] Y. Miura, E. Aramaki, T. Ohkuma, M. Tonoike, D. Sugihara, H. Masuichi, and K. Ohe, "Adverse-effect relations extraction from massive clinical records," in *Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era*, 2010, pp. 75–83.

[19] A. Henriksson, M. Kvist, H. Dalianis, and M. Duneld, "Identifying adverse drug event information in clinical notes with distributional semantic representations of context," *Journal of Biomedical Informatics*, vol. 57, pp. 333–349, 2015.

[20] Y. Luo, "Recurrent neural networks for classifying relations in clinical notes," *Journal of Biomedical Informatics*, vol. 72, pp. 85–95, 2017.

[21] D. Raj, S. Sahu, and A. Anand, "Learning local and global contexts using a convolutional recurrent network model for relation classification in biomedical text," in *Proceedings of the 21st Conference on Computational Natural Language Learning*, 2017, pp. 311–321.

[22] S. de la Peña, I. Segura-Bedmar, P. Martínez, and J. L. Martínez, "ADRSpanishTool: a tool for extracting adverse drug reactions and indications," *Procesamiento del Lenguaje Natural*, vol. 53, pp. 177–180, 2014.

[23] I. Segura-Bedmar, S. de la Peña, and P. Martınez, "Extracting drug indications and adverse drug reactions from Spanish health social media," in *Proceedings of BioNLP*, 2014, pp. 98–106.

[24] I. Segura-Bedmar, P. Martínez, R. Revert, and J. Moreno-Schneider, "Exploring Spanish health social media for detecting drug effects," *BMC Medical Informatics and Decision Making*, vol. 15, no. 2, pp. 1–9, 2015.

[25] M. Oronoz, K. Gojenola, A. Pérez, A. Díaz de Ilarraza, and A. Casillas, "On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions," *Journal of Biomedical Informatics*, vol. 56, pp. 318–332, 2015.

[26] S. Santiso, A. Casillas, A. Perez, M. Oronoz, and K. Gojenola, "Adverse drug event prediction combining shallow analysis and machine learning," in *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*, 2014, pp. 85–89.

[27] S. Santiso, A. Casillas, A. Pérez, M. Oronoz, and K. Gojenola, "Document-level adverse drug reaction event extraction on electronic health records in Spanish," *Procesamiento del Lenguaje Natural*, no. 56, pp. 49–56, 2016.

[28] A. Casillas, A. Pérez, M. Oronoz, K. Gojenola, and S. Santiso, "Learning to extract adverse drug reaction events from electronic health records in Spanish," *Expert Systems with Applications*, vol. 61, pp. 235–245, 2016.

[29] A. Casillas, A. D. de Ilarraza, K. Fernandez, K. Gojenola, M. Oronoz, A. Pérez, and S. Santiso, "IXAmed-IE: On-line medical entity identification and ADR event extraction in Spanish," in *2016 IEEE International Conference on Bioinformatics and Biomedicine*, 2016, pp. 846–849.

[30] S. Santiso, A. Casillas, and A. Pérez, "The class imbalance problem detecting adverse drug reactions in electronic health records," *Health Informatics Journal*, pp. 1–11, 2018.

[31] S. Santiso, A. Casillas, A. Pérez, and M. Oronoz, "Medical entity recognition and negation extraction: Assessment of NegEx on health records in Spanish," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, 2017, pp. 177–188.

[32] ——, "Word embeddings for negation detection in health records written in Spanish," *Soft Computing*, pp. 1–7, 2018.

[33] S. Santiso, A. Pérez, and A. Casillas, "Smoothing dense spaces for improved relation extraction between drugs and adverse reactions," *International journal of medical informatics*, vol. 128, pp. 39–45, 2019.

[34] ——, "Exploring joint ab-lstm with embedded lemmas for adverse drug reaction discovery," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–8, 2018.

[35] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, vol. 1, 2001, pp. 282–289.