ISSN 2007-9737

Clinical Text Mining in Spanish Enhanced by Negation Detection and Named Entity Recognition

Antonio Jesús Tamayo Herrera[1], Diego A. Burgos[2], Alexander Gelbukh[1]

[1] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[2] Wake Forest University,
Department of Spanish, Winston-Salem,
USA

{atamayoh2019, gelbukh}@cic.ipn.mx, burgosda@wfu.edu

Abstract. Automatic identification of negation, uncertainty, and named entities are tasks of vital importance for clinical text mining. While several works have been published in English, only in recent years Spanish cases have been considered. In this work, we present a transfer learning framework based on a RoBERTa model pre-trained with biomedical documents and on multilingual BERT to identify diseases and organisms mentions as well as negations and uncertainty cues and scopes as a sequence labeling problem, utilizing the fact clinical datasets in Spanish for these four tasks. Our approach achieves results comparable to the state-of-the-art organism mentions identification and negation identification, competitive results in identifying diseases, and establishing state-of-the-art for uncertainty identification. Additionally, to remedy the lack of a unified dataset for the four tasks addressed, models to tackle them have been integrated into a web application that we built to allow effective clinical text mining in Spanish. The source code of this work is publicly available as well as the web application.

Keywords. Clinical text mining, negation scope detection, uncertainty scope detection, diseases, organisms mentions identification.

1 Introduction

The number of clinical texts in digital format is no stranger to the exponential growth curve of information currently available in digital format. However, this clinical information, which can be found in both structured and unstructured form, lacks systems that would allow making the most of it for the benefit of patients and health care providers. It is estimated that approximately 40% of the information collected in health systems contains text written in natural language. In these texts, there is vitally important information about diseases, organisms, symptoms, diagnoses, treatments, medication use, and adverse drug reactions for patients. This information can be used to improve the health and care of new patients [15].

Despite Spanish being the second most important language around the world for the number of speakers, according to the Cervantes Institute, only in recent years have emerged relevant works to tackle tasks in the clinical domain in this language, in contrast to English.

Clinical text mining has many aspects, but without a doubt, some of the most relevant tasks are the identification of diseases, and organisms [37, 38, 46].

Likewise, negation and uncertainty are highly relevant in the clinical domain [10]. The high frequency of negations in clinical documents makes many of the extractions of named entities such as diseases or organisms, should not be taken as present, but on the contrary.

To illustrate this phenomenon, see the following example. *El paciente/***Specie** hipertenso/**Disease** *[no presenta fiebre ni infección/***Disease***].*

Hypertensive/**Disease** patient/**Specie** [does not present fever or infection/**Disease**].

In this example, one of the three entities identified (*infección*) is inside the negation scope marked with square brackets, therefore it should not be considered as present for the patient.

Since, currently, there is no unified database labeled with diseases, organisms, negations, uncertainties, and their scopes, in this work we propose the transfer learning strategy using a language model pre-trained in the clinical domain [9] and multilingual BERT [17] to tackle the four tasks using the same hyperparameters, and under the BIO representation scheme [43] with the same preprocessing.

The remaining sections of this article are structured as follows: in section 2, we present the most relevant related works in the four tasks addressed. Section 3 deals with the methodological aspects of the datasets, the experiments we carried out, and the web application description. Analysis and discussion are presented in section 4, and we close with conclusions in section 5.

2 Related Works

This section is presented in two parts, the first dedicated to computational work on the phenomenon of negation and uncertainty, and the second to those focused on the identification of diseases and organisms in clinical texts.

2.1 Negation and Uncertainty

The phenomenon of negation, uncertainty, and their scopes in a written text has been well explored in the English language, however, there are very few works in Spanish, specifically in the clinical domain, despite being the second mother tongue with the largest number of speakers in the world. Here, we tried to group the main works carried out to solve this task, mainly focused on works in the clinical domain with corpus in Spanish, from the three following approaches reported in the literature: application of rules, using supervised machine learning with classical models, and using deep learning.

2.1.1 Rule-Based

One of the most important works under this approach is the one presented in [36] since it presents the definition and analysis of the different negation patterns and their typology in Spanish. On the same track, there is a work done around negation and its scopes, where different algorithms are implemented to detect whether a term of interest is within the scope of negation in the domain of radiology [14], for which authors used syntactic features such as grammatical categories, syntactic and dependency trees and an adaptation of the rule-based NegEx system proposed in [10]. Likewise, in [49], authors applied rules to identify speculations and enhance the lung cancer diagnosis extraction from clinical documents.

2.1.2 Machine Learning Systems

Negation detection in Spanish has also been addressed using supervised machine learning with classical models. We found that most of the works have been carried out using the SFU ReviewSP NEG corpus [29]. In these works, the predominant machine learning model has been Conditional Random Fields using syntactic features [4, 5, 18, 35, 50]. Likewise, Random Forest, SVM, XGBoost were used in [18]. In addition to SFU ReviewSP NEG corpus, a novel dataset for negation detection was built in [4].

2.1.3 Deep Learning Systems

With the era of deep learning and its accelerated growth in recent years, advances in natural language processing have been notorious and, as in other tasks, these advances have also been applied to negation detection and its scopes. As we mentioned above, most of the studies have been carried out in English.

Among the most relevant works under the deep learning approach for negation detection is one where authors tested multilayer perceptron and bidirectional LSTM architectures [23]. In [42], appears an approach based on convolutional neural networks, in [33], the authors propose a method based on recurrent neural networks.

For their part, in [24], a sophisticated architecture combining recursive neural networks with CRF prediction layer was implemented, while in [19], authors worked with an architecture combining convolutional networks and LSTM, and finally, the most recent works based on the transformer architecture presented in [8, 30, 31, 48].

In Spanish, we found relevant works using modern deep learning techniques such as recurrent and convolutional neural networks in [19, 20, 21, 22, 26, 27, 34, 41]. Regarding transformer-based approaches, both negation and uncertainty have also been faced [28, 41, 49].

2.2 Named Entity Recognition

Like many problems in natural language processing, including that of negation and uncertainty, which was discussed in section 2.1, named entity recognition (NER) in the medical domain has been mostly addressed in English. However, in [2], we found a published corpus with nested entities and addressed the NER problem using a biLSTM-CRF architecture and dense vectors as word embeddings trained with texts from the clinical domain and Spanish Wikipedia.

In [54], authors proposed a novel joint deep learning model to address NER and normalization on a Spanish cancer corpus achieving an F1 of 0.87. Recently, some shared tasks have released datasets that have allowed the evolution of clinical text mining in Spanish. In this way, we found works carried out utilizing the same datasets we used in this work to identify diseases and organisms.

For diseases identification, most of the works, including state-of-the-art models, tackle the problem using transformer-based approaches [7, 11, 39, 44, 51], while in [6], a vector database and a question-answering approach are used to identify diseases, and a dictionary-based approach is reported in [40].

Similarly, for organisms' identification, most of the works employ the transfer learning technique with BERT-based models [1, 3, 12, 13, 16, 25, 32, 47, 52, 53, 55, 56], while in [32], recurrent and convolutional neural networks are used, and a dictionary-based approach is reported in [13]. Additionally, an interesting approach based on FLERT is presented in [45].

3 Methodology

In this section, we describe the datasets used to train the models to tackle the following four tasks: disease identification, organism identification, negation detection, and uncertainty detection. Likewise, we present the experiments carried out to tackle the four tasks, as well as the description of the web application that we have developed to remedy the lack of a unified dataset labeled for the four tasks addressed.

3.1 Datasets

Datasets for NER (diseases and organisms) were taken from DisTEMIST [38] and LivingNER [37] shared tasks, respectively, while for negation and uncertainty detection, we used the largest publicly available dataset, NUBES [34].

The datasets for NER tasks have the same structure, with the following six fields in tab-separated values (TSV) format: document name, mention identifier (id), mention type (Disease or Human/Species), starting and ending position of the mention in the document, and text span.

On the other hand, the NUBES dataset was annotated using brat format. For this work, we used its negation and uncertainty cues and scope annotations. Each dataset is described below.

3.1.1 Diseases

Dataset with disease annotations consists of 1000 clinical cases partitioned into two subsets (750 for training and 250 for testing). Clinical cases were collected from different specialties (pediatrics, radiology, urology, oncology, and cardiology, among others).

This dataset was annotated by two specialists in the clinical domain with the supervision of a physician. An inter-annotator agreement of 82.3% between the two specialists was achieved. It includes 10,666 disease mentions, out of which 10,318 are unique.

3.1.2 Organisms

This dataset consists of 1,985 clinical cases. The clinical case reports in this dataset encompass a wide range of medical disciplines, including infectious diseases (including Covid-19 cases), cardiology, neurology, oncology, otolaryngology, dentistry, pediatrics, endocrinology, primary care, allergology, radiology, psychiatry, ophthalmology, psychiatry, urology, internal medicine, emergency and intensive care medicine, radiology, tropical medicine, and dermatology.

These reports involved the manual annotation of entities related to [SPECIES] and [HUMAN]. This dataset exhibits a diverse range of content, with annotations encompassing animals, plants, and microorganisms, such as bacteria, fungi, viruses, and parasites.

Both scientific names and common names were taken into consideration during the annotation process. It was annotated for different specialists in the clinical domain. An inter-annotator agreement of 94% for [SPECIES] and 89% for [HUMAN] labels was achieved.

The dataset includes 30,866 species mentions (43.9% humans), out of which 4,580 are unique. The dataset was partitioned into three subsets (training = 1000; validation = 500, and testing = 485 clinical cases).

3.1.3 Negation and Uncertainty

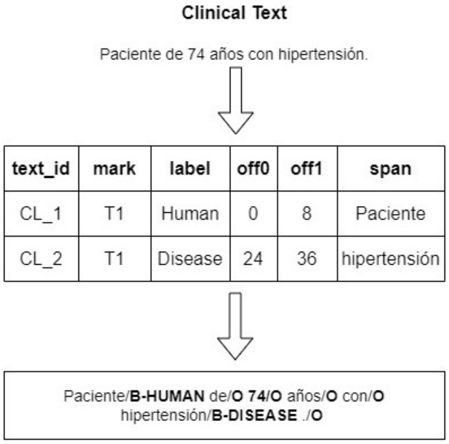
The NUBES dataset is the largest publicly available dataset for negation and uncertainty in Spanish. It includes annotations for negation and uncertainty cues as well as scopes for each phenomenon. Besides, it includes other annotations, but for the purpose of this work, we have not considered them.

It consists of 29,682 sentences extracted from anonymized health records, but for experiments, a portion of these was taken to build a balanced dataset between sentences affected for negation or uncertainty and sentences not affected.

The dataset for experiments counts 18,401 sentences, where nearly half have negations or uncertainties. A portion of the dataset was annotated by two people plus a referee. The rest was done for only one person.

3.2 Experiments

Since there is no dataset labeled with the two types of entities that in this work we intend to identify automatically and at the same time labeled with negations, uncertainty, and their scopes, our work consisted of fine-tuning two of the best large language models for the clinical domain in Spanish for each of the tasks independently, namely: identification of diseases, identification of organisms, and identification of negations/uncertainties and their scopes.



Clinical Text

Paciente de 74 años con hipertensión.

text_id	mark	label	off0	off1	span
CL_1	T1	Human	0	8	Paciente
CL_2	T1	Disease	24	36	hipertensión

Paciente/**B-HUMAN** de/**O** **74/O** años/**O** con/**O** hipertensión/**B-DISEASE** ./**O**

Fig. 1. Preprocessing pipeline

Table 1. Number of sentences per dataset partition

Task	Training	Development	Testing
Diseases	12,028	N/A	4,017
Organisms	27,810	12,480	12,955
Negation/ Uncertainty	13,801	1,839	2,761

Four tasks were faced with the same three steps we describe below.

3.2.1 Preprocessing

As we describe in section 3.1, datasets for named entity recognition used in this work were built in a similar format which is not compatible with the input required for the models used to address the tasks. To process data from the original format to the BIO scheme we used a custom script in Python.

This process takes each entity in the training datasets together with the character where they start and search in order of appearance in the text to assign the corresponding label in each case (disease and organism).

Additionally, we implemented a simple sentence tokenization with the period (".") token. Likewise, the dataset for negation and uncertainty was transformed from Brat to BIO format. This process is illustrated in Figure 1.

Table 1 shows details for the three datasets after the preprocessing, which were used to train the models as we explain below.

In the case of diseases, due to the lack of a development partition, we used a train-test split using the training dataset to have a custom development one to train the model (75% for training, 9,021 sentences; 25% for development, 3,007 sentences).

3.2.2 Fine-Tuning

The fine-tuning process for the four tasks addressed was carried out in two steps, as follows: 1) using the training and development partitions, and 2) using the training and testing partitions. We used two of the best large language models for Spanish in the clinical domain [9, 17].

Taking advantage of the Hugging Face environment to fine-tune transformer-based models, we carried out the fine-tuning process with the RoBERTa[1] biomedical model and multilingual BERT[2], both models are available at the Hugging Face Hub. For all the tasks addressed, we used the same hyperparameter configuration (epochs=7; learning rate=2e-5; weight decay=0.1).

The rest of the parameters were kept by default. A virtual machine with a GPU Tesla T4 with 27.3 gigabytes of available RAM was used to run all the experiments in Google Colab Pro. Source code for preprocessing, fine-tuning, and postprocessing, as well as datasets used in this work, are publicly available in a Github repository[3].

3.2.3 Postprocessing

We implemented post-processing to delete double spaces, non-content words, and punctuation marks at the beginning and end of the entity identified by the model. Likewise, we treated the subword tokenization inherent to the models used.

Additionally, when the models extract two or more entities separately, but they are contiguous, we group them into a unique entity. Figure 2 shows the experiments pipeline we carried out to tackle the four tasks.

3.2.4 Development of the Web Application for Clinical Text Mining

To allow the models trained in this study to work together, the Python programming language and the Django framework were used to develop a modern and secure web application under the Model, View, Template (MVT) architectural pattern.

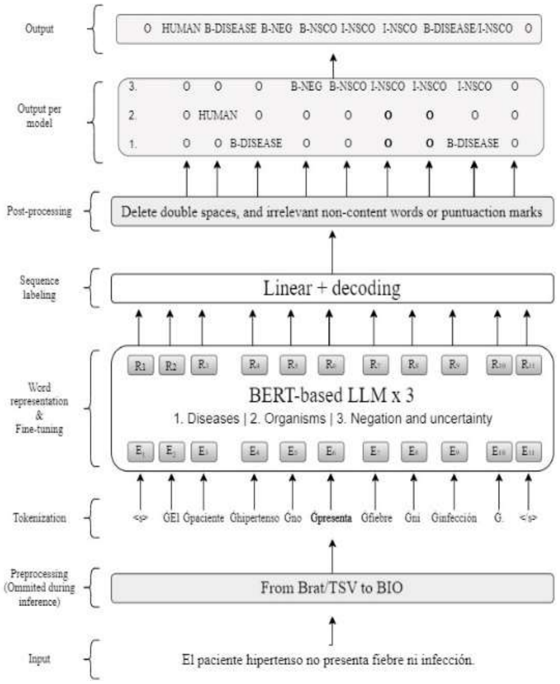


Fig. 2. Experiments' pipeline

Here we worked with a frontend layer managed through Django templates, using HTML, Javascript and CSS to consolidate a friendly web application for the end user. This frontend consists of a form with a text area field, which allows entering a clinical document written in natural language in Spanish, and a button to process the text.

The backend of the application, controlled by the views and models in the MVT architectural pattern, receives the clinical text, call the models trained to extract diseases, organisms, and negations/uncertainties and their scopes, and integrates the predictions of the four LLMs in a unique output to present clinical text mining in one step to the client. Figure 3 illustrates how the web application works.

---

[1] https://huggingface.co/PlanTL-GOB-ES/roberta-base-biomedical-clinical-es

[2] https://huggingface.co/bert-base-multilingual-cased

[3] https://github.com/ajtamayoh/ClinicalTextMining

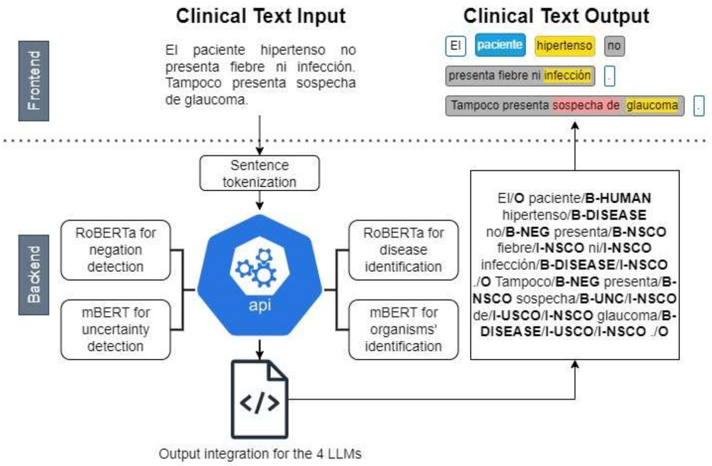


Fig. 3. Web application architecture

Table 2. Results for NER tasks in the development dataset

Task	Model	Pr.	Rec.	F1
Diseases	RoBERTa	0.70	0.76	**0.72**
	mBERT	0.59	0.66	0.62
Organisms	RoBERTa	0.95	0.96	**0.95**
	mBERT	0.95	0.94	**0.95**

Table 3. Results for NER tasks in the testing dataset

Task	Model	Pr.	Rec.	F1
Diseases	RoBERTa	0.70	0.74	**0.72**
	mBERT	0.61	0.66	0.63
Organisms	RoBERTa	0.94	0.95	**0.95**
	mBERT	0.95	0.94	0.94

4 Results and Discussion

In this section, we present the results achieved for our approach to identifying diseases, organisms, negations, and uncertainties, as independent tasks.

Results for all the experiments we carried out using the BIO scheme are presented with average precision (Pr.), average recall (Rec.), and average F1 (F1), to show the capacity of the models to make predictions per token. All the metrics are averaged for the 7 epochs of training.

Due to the results of DisTEMIST and LivingNER shared tasks were presented with micro precision (mP), micro recall (mR), and micro F1 (mF1) for an exact match approach, we also present our results in the testing datasets for NER tasks with the same measures obtained from the evaluation libraries provided by the organizers of the competitions, to allow comparisons with the most relevant works in the tasks addressed.

Tables 2 and 3 show the results for the named entity recognition tasks in the development and the testing datasets, respectively. These results

Table 4. Results for negation identification in the development dataset

Model	**Neg. cue**			**Neg. scope**		
	Pr.	**Rec.**	**F1**	**Pr.**	**Rec.**	**F1**
RoBERTa	0.93	0.95	**0.94**	0.85	0.89	0.87
mBERT	0.92	0.92	0.92	0.89	0.92	**0.91**

Table 5. Results for negation identification in the testing dataset

Model	**Neg. cue**			**Neg. scope**		
	Pr.	**Rec.**	**F1**	**Pr.**	**Rec.**	**F1**
RoBERTa	0.93	0.95	**0.94**	0.87	0.89	0.88
mBERT	0.90	0.92	0.91	0.89	0.90	**0.89**

illustrate that diseases are harder to be identified than organisms and establish that RoBERTa is more powerful than mBERT for this task.

It can be expected since RoBERTa was pre-trained using a large collection of documents in the biomedical field, where disease mentions appear frequently, whereas mBERT can be considered a general language model.

On the other hand, it can be observed that both models have similar accuracy in identifying organism mentions. From our point of view, it occurs due to the high frequency of organisms, specifically HUMAN mentions, which mBERT can identify clearly since they are present in the general language.

Tables 4 and 5 show the results of the negation identification task in the development and testing datasets, respectively.

Despite results in Tables 4 and 5 showing that mBERT slightly outperforms RoBERTa in identifying negation scopes, considering that RoBERTa has a better performance for the negation cue detection than mBERT, it should be considered the best option to identify the negation phenomenon completely since cues and scopes are coupled.

It is worthwhile to negation cues detection is a domain-dependent task, for this reason, RoBERTa can outperform mBERT in this task since it has been trained with more examples of negation cues in the clinical domain than mBERT (For example, discard, absence of …, lack of….or morphological negations as asymptomatic, among others).

Tables 6 and 7 show the results of the uncertainty identification task in the development and testing datasets, respectively.

According to Tables 6 and 7, mBERT outperforms RoBERTa in the identification of uncertainty, both for cues and scopes. After our analysis, it can be expected since uncertainty cues and scopes do not appear to be domain-dependent, and probably mBERT has been exposed to many more examples of uncertainty in general language. (For example, suspected of …, probably …, susceptible of …, among others).

On the other hand, Table 8 shows the results for the NER tasks considering an exact match constraint. That is, all the tokens of an entity must be identified exactly, otherwise the prediction is marked as incorrect.

These results are compared to the best results of DisTEMIST and LivingNER, which are state-of-the-art of disease and organism identification, respectively, considering exact match predictions. For this purpose, all our results were computed using the evaluation libraries provided by the shared tasks' organizers.

The models used to obtain these results were the best for each task during the training process, namely: RoBERTa for diseases, mBERT for organisms, RoBERTa for negation, and mBERT for uncertainty.

For the inference phase, we used the same sentence tokenization approach employed during the training, and the post-processing described in section 3.2.3.

Table 6. Results for uncertainty identification in the development dataset

Model	Unc. Cue			Unc. scope		
	Pr.	**Rec.**	**F1**	**Pr.**	**Rec.**	**F1**
RoBERTa	0.83	0.88	0.85	0.75	0.81	0.78
mBERT	0.87	0.91	**0.89**	0.84	0.84	**0.84**

Table 7. Results for uncertainty identification in the testing dataset

Model	Unc. Cue			Unc. Scope		
	Pr.	**Rec.**	**F1**	**Pr.**	**Rec.**	**F1**
RoBERTa	0.81	0.87	0.84	0.75	0.83	0.78
mBERT	0.85	0.87	**0.86**	0.81	0.84	**0.82**

Table 8. Results comparison for NER tasks in the testing dataset under the exact match approach

Task	Model	Pr.	Rec.	F1
Diseases	SOTA [39]	0.79	0.76	**0.78**
	Our approach	0.62	0.75	0.68
Organisms	SOTA [56]	0.96	0.94	**0.95**
	Our approach	0.85	0.90	0.88

Although the results in Table 8 show that the models presented in [39, 56] outperform our approach for disease and organism identification, respectively, if we relax the exact match constraint, our approach is useful and applicable from two usability perspectives of information extraction in the clinical domain, namely: (1) helping medical staff to streamline their work by allowing them to focus on the most relevant aspects of their patient’s situation analysis, and (2) linking entities are their respective codes in standardized resources such as SNOMED-CT. From the first perspective, our approach can identify disease and organism mentions and highlight them to allow quick scanning of the clinical document.

Despite some tokens could not be highlighted, the information presented in context is easy to read and completely useful to speed up the work of the medical staff, since errors due to nested entities and expanded extractions do not tarnish the results, as shown in the error analysis carried out in [51, 52].

From the second perspective, our approach is also competitive since the state-of-the-art methods for linking entities with standardized codes are based on semantic similarity, and in most cases, the models we propose do not extract noisy but relevant words to identify diseases or organisms.

In this way, our approach can extract organisms with an F1=0.95 and diseases with an F1=0.72, which is comparable with the state-of-the-art for organisms and highly competitive for disease identification.

Finally, Tables 9 and 10 show the comparison of the results for negation and uncertainty identification between state-of-the-art (SOTA) and our approach. Results in Tables 9 and 10 show our approach is very competitive to detect negations and establish state-of-the-art for uncertainty identification.

4.1 Clinical Text Mining Via Web

Having shown the power of our best models to identify diseases and organisms, as well as to detect negations, uncertainties, and their scopes in clinical documents, we deployed them using the hugging face environment via API to be consumed

Table 9. Results comparison for negation identification in the testing dataset

Model	Neg. Cue			Neg. Scope		
	Pr.	Rec.	F1	Pr.	Rec.	F1
SOTA [41]	0.96	0.95	**0.95**	0.92	0.93	**0.92**
Our approach	0.93	0.95	0.94	0.87	0.89	0.88

Table 10. Results comparison for uncertainty identification in the testing dataset

Model	Unc. Cue			Unc. Scope		
	Pr.	Rec.	F1	Pr.	Rec.	F1
SOTA [41]	0.86	0.83	0.84	0.82	0.79	0.80
Our approach	0.85	0.87	**0.86**	0.81	0.84	**0.82**

by the web application described in section 3.2.4, which we called SIMPLE (*Sistema Informático Médico con Procesamiento de Lenguaje*). SIMPLE is publicly available[4].

5 Conclusions

In this study, we have presented a transfer learning approach to address clinical text mining in Spanish through disease and organism mention identification, as well as negation and uncertainty detection. Models we trained can extract organisms with an F1 of 0.95 and diseases with an F1 of 0.72, which is comparable with the state-of-the-art for organisms and highly competitive for disease identification.

Likewise, we trained models to detect negation cues (F1=0.94), negation scopes (0.88), uncertainty cues (F1=0.86), and uncertainty scopes (F1=0.82). Our approach established state-of-the-art uncertainty detection and shows competitive results in negation identification.

Our study allows us to identify that RoBERTa, pretrained in the biomedical field with documents written in Spanish, exhibits better capabilities to extract diseases and negations, while multilingual BERT does better in identifying organisms and uncertainty.

To address the lack of a dataset labeled with diseases, organisms, negations, and uncertainty at the same time, we developed a publicly available web application called SIMPLE, which integrates four LLMs to tackle the four tasks at the same time et al. for better clinical text mining in Spanish since it identifies the presence, absence or suspect of a disease or organism in a clinical document.

SIMPLE can help to improve the health care of new patients and facilitate the work of medical staff through its effective information extraction system and can be used as an intermediate step to link clinical entities with their standardized codes.

Finally, it should be noted that this work was developed for Spanish since there are multiple approaches in English and some systems developed for that language, but they are of no use in real life for the Spanish-speaking population, which is the third most representative worldwide, according to the Cervantes Institute.

Acknowledgments

This work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONACYT, Mexico, grants 20220852 and 20220859 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico, and

---

[4] www.clinicaltextmining.com

acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

References

1. **Avram, A. M., Mitrofan, M., Păis, V. (2022).** Species entity recognition using a neural inhibitory mechanism. CEUR Workshop Proceedings.
2. **Báez, P., Villena, F., Rojas, M., Durán, M., Dunstan, J. (2020).** The chilean waiting list corpus: A new resource for clinical named entity recognition in Spanish. Proceedings of the 3rd Clinical Natural Language Processing Workshop, pp. 291–300. DOI: 10.18653/v1/2020.clinicalnlp-1.32.
3. **Bagherzadeh, P., Verma, H., Bergler, S. (2022).** Multi-Input RIM for named-entity recognition in spanish clinical reports. CEUR Workshop Proceedings.
4. **Bel-Enguix, G., Gómez-Adorno, H., Pimentel, A., Ojeda-Trueba, S. L., Aguilar-Vizuet, B. (2021).** Negation detection on mexican spanish tweets: The T-MexNeg corpus. Applied Sciences, Vol. 11, No. 9, pp. 3880. DOI: 10.3390/app11093880.
5. **Beltrán, J., González, M. (2019).** Detection of negation cues in spanish: The CLiC-Neg system. Proceedings of the Iberian Languages Evaluation Forum IberLEF'19, pp. 352–360.
6. **Bernika, M., Tovornika, R., Fabjana, B., Marco-Ruizb, L. (2022).** DIAGÑOZA: A natural language processing tool for automatic annotation of clinical free text with SNOMED-CT. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum. CEUR Workshop Proceedings, Vol. 3180.
7. **Borchert, F., Schapranow, M. P. (2022).** HPI-DHC@ BioASQ DisTEMIST: Spanish biomedical entity linking with pre-trained transformers and cross-lingual candidate retrieval. Working notes of conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings.
8. **Britto, B. K., Khandelwal, A. (2020).** Resolving the scope of speculation and negation using transformer-based architectures. DOI: 10.48550/arXiv.2001.02885.
9. **Carrino, C. P., Armengol-Estapé, J., Gutiérrez-Fandiño, A., Llop-Palao, J., Pàmies, M., Gonzalez-Agirre, A., Villegas, M. (2021).** Biomedical and clinical language models for spanish: On the benefits of domain-specific pretraining in a mid-resource scenario. DOI: 10.48550/arXiv.2109.03570.
10. **Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., Buchanan, B. G. (2001).** Evaluation of negation phrases in narrative clinical reports. Proceedings of the AMIA Symposium American Medical Informatics Association, pp.105–109.
11. **Chizhikova, M., Collado-Montañez, J., López-Úbeda, P., Díaz-Galiano, M. C., Ureña-López, L. A., Martín-Valdivia, M. T. (2022).** SINAI at CLEF 2022: Leveraging biomedical transformers to detect and normalize disease mentions. In Working Notes of Conference and Labs of the Evaluation (CLEF) Forum. CEUR Workshop Proceedings.
12. **Chizhikova, M., Collado-Montañez, J., López-Úbeda, P., Díaz-Galiano, M. C., Ureña-López, L. A., Martín-Valdivia, M. T. (2022).** SINAI at LivingNER shared task 2022: Species mention recognition and normalization using transfer learning and string-matching techniques. CEUR Workshop Proceedings CEUR-WS, Vol. 3202.
13. **Cossin, S., Diallo, G., Jouhet, V. (2022).** IAM at IberLEF 2022: NER of Species Mentions. CEUR Workshop Proceedings CEUR-WS, Vol. 3202.
14. **Cotik, V., Stricker, V., Vivaldi, J., Rodríguez-Hontoria, H. (2016).** Syntactic methods for negation detection in radiology reports in Spanish. Proceedings of the 15th Workshop on Biomedical Natural Language Processing, BioNLP'16 Association for Computational Linguistics, Berlin, pp. 156–165. DOI: 10.18653/v1/W16-2921.
15. **Dalianis, H., Hassel, M., Velupillai, S. (2009).** The Stockholm EPR corpus-characteristics and some initial findings. Proceedings of ISHIMR 2009, Evaluation and Implementation of e-Health and Health Information Initiatives: International Perspectives, 14th International

Symposium for Health Information Management Research, pp. 243–249.

16. **del Moral, R., Reyes-Aguillón, J., Ramos-Flores, O., Gómez-Adorno, H., Bel-Enguix, G. (2022).** Species mention entity recognition, linking and classification using RoBERTa in combination with spanish medical embeddings. CEUR Workshop Proceedings CEUR-WS, Vol. 3202.
17. **Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018).** Bert: Pre-training of deep bidirectional transformers for language understanding. DOI: 10.48550/arXiv.18 10.04805.
18. **Domınguez-Mas, L., Ronzano, F., Furlong, L. I. (2019).** Supervised learning approaches to detect negation cues in spanish reviews. Proceedings of the Iberian Languages Evaluation Forum IberLEF'19, CEUR Workshop Proceedings, Vol. 2421.
19. **Fabregat, H., Araujo-Serna, L., Martínez Romo, J. (2019).** Deep learning approach for negation trigger and scope recognition. Procesamiento de Lenguaje Natural, Vol. 62, pp. 37–44. DOI: 10.26342/2019-62-4.
20. **Fabregat, H., Duque, A., Araujo, L., Martinez-Romo, J. (2021).** LSI_UNED at CLEF eHealth2021: Exploring the effects of transfer learning in negation detection and entity recognition in clinical texts. CLEF Working Notes, Vol. 2936, pp. 780–793.
21. **Fabregat, H., Duque, A., Martínez-Romo, J., Araujo, L. (2019)**. Extending a deep learning approach for negation cues detection in spanish. Proceedings of the Iberian Languages Evaluation Forum CEUR Workshop Proceedings, CEUR-WS, Vol. 2421.
22. **Fabregat, H., Martínez-Romo, J., Araujo, L. (2018).** Deep learning approach for negation cues detection in spanish at NEGES 2018. Proceedings of NEGES 2018: Workshop on Negation in Spanish, Vol. 2174, pp. 43–48.
23. **Fancellu, F., Lopez, A., Webber, B. (2016).** Neural networks for negation scope detection. Proceedings of the 54th annual meeting of the Association for Computational Linguistics, Vol. 1, pp. 495–504.
24. **Fei, H., Ren, Y., Ji, D. (2020).** Negation and speculation scope detection using recursive neural conditional random fields. Neurocomputing, Vol. 374, pp. 22–29. DOI: 10.1016/j.neucom.2019.09.058.
25. **Francis, S., Moens, M. F. (2022).** Task-aware contrastive pre-training for spanish named entity recognition in LivingNER challenge. CEUR Workshop Proceedings, CEUR-WS, Vol. 3202.
26. **García-Largo, M. Á. M. C., Segura-Bedmar, I. (2021).** Extracting information from radiology reports by natural language processing and deep learning. CLEF, Working Notes, Vol. 2936, pp. 804–817.
27. **Giudice, V. (2019).** Aspie96 at NEGES (IberLEF 2019): Negation cues detection in spanish with character-level convolutional RNN and tokenization. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019), CEUR Workshop Proceedings, Vol. 2421.
28. **Hartmann, M., Søgaard, A. (2021).** Multilingual negation scope resolution for clinical text. Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis, pp. 7–18.
29. **Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Urena-López, L. A., Martí, M. A. (2018).** SFU Review SP-NEG: a spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns, Language Resources and Evaluation, Vol. 52, pp. 533–569. DOI: 10.1007/s10579-017-9391- x.
30. **Khandelwal, A., Britto, B. K. (2020).** Multitask learning of negation and speculation using transformers. Proceedings of the 11th international workshop on health text mining and information analysis, pp. 79–87. DOI: 10.18653/v1/2020.louhi-1.9.
31. **Khandelwal, A., Sawant, S. (2019).** NegBERT: a transfer learning approach for negation detection and scope resolution. DOI: 10.48550/arXiv.1911.04211.
32. **Kocaman, V., Pirge, G., Polat, B., Talby, D. (2022).** Biomedical named entity recognition in

eight languages with zero code changes. CEUR Workshop Proceedings, Vol. 3202.

33. **Lazib, L., Zhao, Y., Qin, B., Liu, T. (2016).** Negation scope detection with recurrent neural networks models in review texts. International Journal of High-Performance Computing and Networking, Vol. 13, No. 2, pp.211–221. DOI: 10.1504/IJHPCN.2019.097501.
34. **Lima, S., Pérez, N., Cuadros, M., Rigau, G. (2020).** NUBes: A corpus of negation and uncertainty in spanish clinical texts. Proceedings of the 12th LREC, pp. 11–16.
35. **Loharja, H., Padró, L., Turmo, J. (2018).** Negation cues detection using CRF on Spanish product review texts. NEGES 2018: Workshop on Negation in Spanish, pp. 19-21.
36. **Martí, M. A., Taulé, M., Nofre, M., Marsó, L., Martín-Valdivia, M. T., Jiménez-Zafra, S. M. (2016).** La negación en español: análisis y tipología de patrones de negación. Procesamiento del Lenguaje Natural, Vol. 57, pp. 41–48.
37. **Miranda-Escalada, A., Farré-Maduell, E., Lima-López, S., Estrada, D., Gascó, L., Krallinger, M. (2022).** Mention detection, normalization & classification of species, pathogens, humans and food in clinical documents: Overview of the LivingNER shared task and resources. Procesamiento del Lenguaje Natural, Vol. 69, pp. 241–253.
38. **Miranda-Escalada, A., Gascó, L., Lima-López, S., Farré-Maduell, E., Estrada, D., Nentidis, A., Krithara, A., Katsimpras, G., Paliouras, G., Krallinger, M.** (2022). Overview of DisTEMIST at BioASQ: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings.
39. **Moscato, V., Postiglione, M., Sperlí, G. (2022).** Biomedical spanish language models for entity recognition and linking at BioASQ DisTEMIST. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings.
40. **Neves, A. (2022).** Unicage at DISTEMIST-named entity recognition system using only bash and unicage tools. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings.
41. **Pabón, O. S., Montenegro, O., Torrente, M., González, A. R., Provencio, M., Menasalvas, E. (2022).** Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach. PeerJ Computer Science, Vol. 8, No. e913. DOI: 10.7717/peerj-cs.913.
42. **Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., Luo, W. (2016).** Speculation and negation scope detection via convolutional neural networks. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 815-825.
43. **Ramshaw, L. A., Marcus, M. P. (1999).** Text chunking using transformation-based learning. In: Armstrong, S., Church, K., Isabelle, P., Manzi, S., Tzoukermann, E., Yarowsky, D. (eds) Natural Language Processing Using Very Large Corpora, Text, Speech and Language Technology, Vol. 11, pp. 157–176. DOI: 10.1007/978-94-017-2390-9_10.
44. **Reyes-Aguillón, J., del Moral, R., Ramos-Flores, O., Gómez-Adorno, H., Bel-Enguix, G. (2022).** Clinical named entity recognition and linking using BERT in combination with spanish medical embeddings. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings, Vol. 3180.
45. **Rojas, M., Barros, J., Araneda, M., Dunstan, J. (2022).** FLERT-Matcher: A two-step approach for clinical named entity recognition and normalization. CEUR Workshop Proceedings, Vol. 3202.
46. **Sánchez, L. G., Zavala, D. E., Farré-Maduell, E., Lima-López, S., Miranda-Escalada, A., Krallinger, M. (2022).** The SocialDisNER shared task on detection of disease mentions in health-relevant content from social media: methods, evaluation, guidelines and corpora. Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, pp. 182–189.
47. **Schwarz, M., Chapman, K., Häussler, B. (2022).** Multilingual medical entity recognition and cross-lingual zero-shot linking with

facebook AI similarity search. CEUR Workshop Proceedings, Vol. 3202.

48. **Shaitarova, A., Furrer, L., Rinaldi, F. (2020).** Cross-lingual transfer-learning approach to negation scope resolution. CEUR Workshop proceedings, CEUR-WS.
49. **Solarte-Pabón, O., Torrente, M., Provencio, M., Rodríguez-Gonzalez, A., Menasalvas, E. (2021).** Integrating speculation detection and deep learning to extract lung cancer diagnosis from clinical notes. Applied Sciences, Vol. 11, No. 2, pp. 865. DOI: 10.3390/app11020865.
50. **Tamayo, A., Angel, J., Gelbukh, A. (2022).** Detección del alcance de las negaciones en español usando conditional random fields. Research in Computing Science, Vol. 151, No. 5, pp.199–212.
51. **Tamayo, A., Burgos, D. A., Gelbukh, A. (2022).** mBERT and simple post-processing: A baseline for disease mention detection in spanish. Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings.
52. **Tamayo, A., Burgos, D. A., Gelbukh, A. (2022).** ParTNER: paragraph tuning for named entity recognition on clinical cases in spanish using mBERT+ Rules. *CEUR Workshop Proceedings*, Vol. 3202.
53. **Vaca-Serrano, A. (2022).** Named entity recognition for humans and species with domain-specific and domain-adapted transformer models. CEUR Workshop Proceedings, Vol. 3202.
54. **Xiong, Y., Huang, Y., Chen, Q., Wang, X., Nic, Y., Tang, B. (2020).** A joint model for medical named entity recognition and normalization. Vol. 2664, http://ceur-ws.
55. **Yepes, A. J., Verspoor, K. (2022).** The read-biomed team in livingner task 1 (2022): Adaptation of an english annotation system to spanish. LiverNER challenge, IberLEF, CEUR Workshop Proceedings, Vol. 3202.
56. **Zotova, E., García-Pablos, A., Perez, N., Turón, P., Cuadros, M. (2022).** Vicomtech at livingner 2022. CEUR Workshop Proceedings, Vol. 3202.

Article received on 25/04/2023; accepted on 10/06/2023.
Corresponding author is Alexander Gelbukh.