

# FRE at SocialDisNER: Joint Learning of Language Models for Named Entity Recognition

Kendrick Cetina and Nuria García-Santa

Fujitsu Research of Europe (FRE)

Camino Cerro de los Gamos 1, 28224. Pozuelo de Alarcón (Madrid), Spain.

{kendrick.cetina, nuria.garcia.uk} @fujitsu.com

## Abstract

This paper describes our followed methodology for the automatic extraction of disease mentions from tweets in Spanish as part of the SocialDisNER challenge within the 2022 Social Media Mining for Health Applications (SMM4H) Shared Task. We followed a Joint Learning ensemble architecture for the fine-tuning of top performing pre-trained language models in biomedical domain for Named Entity Recognition tasks. We used text generation techniques to augment training data. During practice phase of the challenge our approach showed results of 0.87 F1-Score.

## 1 Introduction

Twitter is one of the main channels for communication at present. The high number of users and the daily activity worldwide on this platform, in addition to the nature of tweets, i.e., short informal text messages, make it as an ideal source of information for Natural Language Processing (NLP) tasks (Yang et al., 2020). The combination of such amount of information with pre-trained transformer language models has changed the way we do NLP nowadays (Qiu et al., 2020). These pre-trained language models have brought significant breakthroughs in deep learning, creating a revolutionary ecosystem for NLP techniques, such as Named Entity Recognition, Text Classification, etc. In the 2022 SocialDisNER challenge, we focused on fine-tuning several pre-trained language models architectures to detect disease mentions in tweets.

### 1.1 Task Description

Social Media Mining for Health Applications (SMM4H) workshop offers the opportunity to develop NLP systems for automatic extraction of relevant knowledge from social media data. On this regard, the purpose of the 2022 subtask, Mining Social Media Content for Disease Mentions (SocialDisNER) (Gasco et al., 2022), is to recognise

disease mentions in tweets written in Spanish, including informal and professional language style. The objective is to extract the beginning and end positions of disease mentions<sup>1</sup>.

## 2 System Description

The summary of our system is the following: first, we search for the two best performing pre-trained language models for our task. Then, we used data augmentation techniques to extend the initial training data and fine-tune the selected language models with the Joint Learning ensemble approach described in (García-Santa and Cetina, 2021). Finally, by qualitative analysis of results, we cleaned the model predictions with post-processing techniques.

### 2.1 Model Selection

We used Hugging Face<sup>2</sup> models hub to search for Fill-Mask and Token Classification models suitable for Spanish Named Entity Recognition of Diseases. Models we tested include BioBERT, several BERT-Based models, RoBERTa, Multi-lingual BERT and DistilBERT. We tested more than 15 different language models fine-tuning with training and testing data provided by the challenge organizers. After performance analysis of all pre-trained language models tested we selected the two highest performing models to be used in our Joint Learning approach: SINAI: a RoBERTa-based model<sup>3</sup> from SINAI team (Chizhikova et al.) at DISease TEXT Mining Shared Task 2022 (Miranda-Escalada et al., 2022) and bsc-bio-es model<sup>4</sup> from Barcelona Supercomputing Center that is a RoBERTa-based model trained on a biomedical corpus in Spanish collected from several sources (Carrino et al., 2022).

<sup>1</sup><https://temu.bsc.es/socialdisner/>

<sup>2</sup><https://huggingface.co/models>

<sup>3</sup>[https://huggingface.co/chizhikchi/Spanish\\_disease\\_finder](https://huggingface.co/chizhikchi/Spanish_disease_finder)

<sup>4</sup><https://huggingface.co/PlanTL-GOB-ES/bsc-bio-es>

## 2.2 Data Augmentation

We used GPT-2 (Radford et al., 2019) for text generation (Li et al., 2022). GPT-2 is a transformers model pre-trained on a very large corpus of English data in a self-supervised fashion trained with the goal of generating texts from a prompt.

For our system we used the challenge training set as input to fine-tune GPT-2<sup>5</sup> for 10 epochs. Then, we used 1000 random noun-chunks from the provided twitter data to generate sentences of 150 words with fine-tuned GPT-2. Next, we used pattern matching with known entities to tag the diseases in generated text.

Figure 1 shows the F1 score in each training epoch of the selected highest performance models: SINAI and bsc-bio-es. In green we can see the performance of augmenting the training set with GPT-2 and fine-tuning bsc-bio-es.

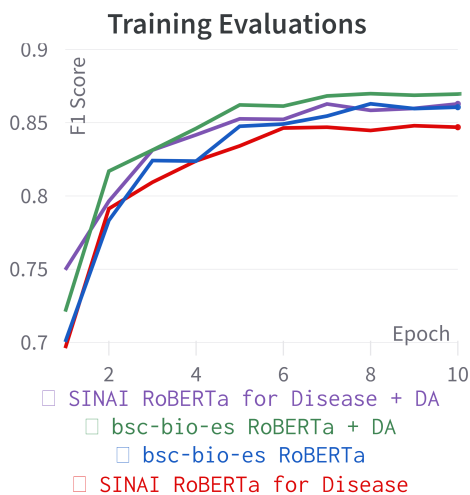


Figure 1: Training performance

## 2.3 Learning & Post-processing

We used the approach described in (García-Santa and Cetina, 2021) for the joint learning of our selected best-performance models (SINAI & bsc-bio-es). The workflow of our joint learning architecture comprises language models as sub-networks concatenated to form a larger network. The sub-networks learn together the best combined classification, retrieving an ensemble model. This architecture allows weight update of language model sub-networks jointly.

Finally, a qualitative assessment of model prediction lead us to implement a post-processing com-

mon to all predictions made. This post-processing includes cleaning of predicted entities, removal of emojis and hashtags and concatenation of consecutive entities predicted as two different entities.

## 3 Evaluation

During the practice phase of the challenge, participants were given the opportunity to test their approach against the evaluation system from the organizers to assess the quality in an intermediate stage. During this phase we obtained strict F1-Scores of 0.84 with our joint-embedding model fine-tuned with the augmented data and 0.87 with our post-processing techniques.

Two sets of predicted data were submitted for final evaluation. Submission 1 consisted of the Joint Learning Model predictions with our post-processing cleaning and Submission 2 consisted of the Joint Learning Model predictions without any kind of post-processing.

Performances of each submission are presented in Table 1. We can see that model predictions with post-processing resulted in the highest performance.

	Strict P	Strict R	Strict F
Submission 1	0.680	0.805	0.738
Submission 2	0.667	0.805	0.729

Table 1: Final evaluation results of our submitted predictions.

## 4 Conclusion

In this paper we presented our system focused on fine-tuning two RoBERTa-based language models together with our Joint Learning approach for disease mentions detection task in Spanish tweets. We extended the training data by generating text with a GPT-based model fine-tuned with biomedical and twitter data. Combining the Joint Learning and augmented data, our system achieved 0.87 F1-Score in the practice phase of the challenge. Our final results in evaluation phase of the challenge achieved 0.73 and 0.74 F1-Scores. In the future, we plan to apply transfer learning techniques to extend our system to the analysis of social media messages in other languages.

<sup>5</sup><https://huggingface.co/gpt2>

## References

- Casimiro Pio Carrino, Joan Llop, Marc Pàmies, Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Joaquín Silveira-Ocampo, Alfonso Valencia, Aitor Gonzalez-Agirre, and Marta Villegas. 2022. [Pre-trained biomedical language models for clinical NLP in Spanish](#). In *Proceedings of the 21st Workshop on Biomedical Language Processing*, pages 193–199, Dublin, Ireland. Association for Computational Linguistics.
- Mariia Chizhikova, Jaime Collado-Montañéz, Pilar López-Úbeda, Manuel C. Díaz-Galiano, L. Alfonso Ureña-López, and M. Teresa Martín-Valdivia. [Sinai at clef 2022: Leveraging biomedical transformers to detect and normalize disease mentions](#). pages 265–273.
- Nuria García-Santa and Kendrick Cetina. 2021. [Extracting multilingual relations with joint learning of language models](#). In *Joint European Conference on Machine Learning and Practice Knowledge Discovery in Databases*, pages 401–407. Springer.
- Luis Gasco, Darryl Estrada-Zavala, Eulàlia Farré-Maduell, Salvador Lima-López, Antonio Miranda-Escalada, and Martin Krallinger. 2022. [Overview of the SocialDisNER shared task on detection of diseases mentions from healthcare related and patient generated social media content: methods, evaluation and corpora](#). In *Proceedings of the Seventh Social Media Mining for Health (#SMM4H) Workshop and Shared Task*.
- Bohan Li, Yutai Hou, and Wanxiang Che. 2022. [Data augmentation approaches in natural language processing: A survey](#). *AI Open*.
- Antonio Miranda-Escalada, Luis Gascó, Salvador Lima-López, Eulàlia Farré-Maduell, Darryl Estrada, Anastasios Nentidis, Anastasia Krithara, Georgios Katsimpras, Georgios Paliouras, and Martin Krallinger. 2022. [Overview of distemist at biosq: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources](#).
- Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. [Pre-trained models for natural language processing: A survey](#). *Science China Technological Sciences*, 63(10):1872–1897.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. [Language models are unsupervised multitask learners](#).
- Sihui Yang, Zifan Ning, and Yaping Wu. 2020. [Nlp based on twitter information: A survey report](#). In *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pages 620–625. IEEE.