Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach

Oswaldo Solarte Pabón Corresp., [1, 2], **Orlando Montenegro** Corresp., [3], **Maria Torrente** [4], **Alejandro Rodriguez-Gonzalez** [1], **Mariano Provencio** [5], **Ernestina Menasalvas** [1]

[1] Universidad Politécnica de Madrid, Madrid, Madrid, Spain

[2] Universidad del Valle, Cali Colombia, Cali, Valle, Colombia

[3] Universidad del Valle, Cali Colombia, Cali, Colombia

[4] Hospital Universitario Puerta de Hierro, Madrid, Madrid, Spain

[5] Hospital Universitatio Puerta de Hierro, Majada Honda, Madrid, Spain

Corresponding Authors: Oswaldo Solarte Pabón, Orlando Montenegro
Email address: oswaldo.solartep@alumnos.upm.es, orlando.montenegro@correounivalle.edu.co

Detecting negation and uncertainty is crucial for medical text mining applications; otherwise, extracted information can be incorrectly identified as real or factual events. Although several approaches have been proposed to detect negation and uncertainty in clinical texts, most efforts have focused on the English language. Most proposals developed for Spanish have focused mainly on negation detection and do not deal with uncertainty. In this paper, we propose a deep learning-based approach for both negation and uncertainty detection in clinical texts written in Spanish. The proposed approach explores two deep learning methods to achieve this goal: i) Bidirectional Long-Short Term Memory with a Conditional Random Field layer (BiLSTM-CRF) and ii) Bidirectional Encoder Representation for Transformers (BERT). The approach was evaluated using NUBES and IULA, two public corpora for the Spanish language. The results obtained showed an F-score of 92% and 80% in the scope recognition task for negation and uncertainty, respectively. We also present the results of a validation process conducted using a real-life annotated dataset from clinical notes belonging to cancer patients. The proposed approach shows the feasibility of deep learning-based methods to detect negation and uncertainty in Spanish clinical texts. Experiments also highlighted that this approach improves performance in the scope recognition task compared to other proposals in the biomedical domain.

Negation and Uncertainty detection in clinical texts written in Spanish: A deep learning-based approach

Oswaldo Solarte-Pabón[1,2], Orlando Montenegro[2], Maria Torrente[3], Alejandro Rodriguez-Gonzalez[1], Mariano Provencio[3], and Ernestina Menasalvas[1]

[1]**Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Madrid, Spain**
[2]**Universidad del Valle, Cali, Colombia**
[3]**Hospital Universitario Puerta de Hierro de Madrid, Spain**

Corresponding author:
Oswaldo Solarte-Pabón[1]

Email address: oswaldo.solartep@alumnos.upm.es

ABSTRACT

Detecting negation and uncertainty is crucial for medical text mining applications; otherwise, extracted information can be incorrectly identified as real or factual events. Although several approaches have been proposed to detect negation and uncertainty in clinical texts, most efforts have focused on the English language. Most proposals developed for Spanish have focused mainly on negation detection and do not deal with uncertainty. In this paper, we propose a deep learning-based approach for both negation and uncertainty detection in clinical texts written in Spanish. The proposed approach explores two deep learning methods to achieve this goal: i) Bidirectional Long-Short Term Memory with a Conditional Random Field layer (BiLSTM-CRF) and ii) Bidirectional Encoder Representation for Transformers (BERT). The approach was evaluated using NUBES and IULA, two public corpora for the Spanish language. The results obtained showed an F-score of 92% and 80% in the scope recognition task for negation and uncertainty, respectively. We also present the results of a validation process conducted using a real-life annotated dataset from clinical notes belonging to cancer patients. The proposed approach shows the feasibility of deep learning-based methods to detect negation and uncertainty in Spanish clinical texts. Experiments also highlighted that this approach improves performance in the scope recognition task compared to other proposals in the biomedical domain.

1 INTRODUCTION

Narrative medical records can provide valuable information to support clinical research, but frequently this information contains uncertain and negated findings (Vincze et al., 2008). Detecting negation and uncertainty is important for medical text mining applications because extracted findings can be incorrectly identified as real or factual events. However, due to the complexity of natural language, automatic identification of negated and uncertain events in clinical texts is not an easy task (Agarwal and Yu, 2010a,b). Moreover, clinical texts are written by highly skilled physicians and nurses using domain-specific terms, under time pressure, with a rich and complex jargon, which makes these texts differ from those of other domains (Dalianis, 2018).

Negation changes the meaning of an affirmative sentence, phrase, or word in a negative way. While uncertainty is used to describe ambiguous or suspected events where their truth value cannot be determined due to a lack of information (Jean et al., 2016; Szarvas et al., 2012b). In the medical field, it must be known whether the patient definitely suffers, probably suffers, or does not suffer from an illness (Vincze, 2014). In the sentence “A 74-year-old patient with suspected lung carcinoma.”, the truth value of the clinical finding *“lung carcinoma”* cannot be confirmed, as this finding is uncertain, suspicious, or speculative. Uncertainty detection has also been studied in terms of modality, and it involves related concepts such as subjectivity, hedging, and speculation (Morante and Sporleder, 2012; Cruz Díaz et al., 2012; Solarte

Pabón et al., 2021). Furthermore, uncertainty is inherent in many medical decisions, as physicians face uncertain results when they are diagnosing or treating patients (Nikfarjam et al., 2014). The breadth and complexity of possible diagnoses in medical practice make uncertainty very common in clinical narratives (Alam et al., 2017; Bhise et al., 2018). Consequently, both negation and uncertainty detection are crucial tasks for information extraction in the medical domain.

Negation and uncertainty detection is commonly divided into two sub-tasks: i) *cue identification* and ii) *scope recognition*. Cues are words or terms that express negation (e.g., not, without, denies) or uncertainty (e.g., possible, probable, suggest) (Cruz Díaz and Maña López, 2019). The scope is the text fragment affected by the corresponding cue in a sentence (De Albornoz et al., 2012). In the sentence: *"**Probable** lung carcinoma with high fever since yesterday, biopsy test will be taken on 25-07-2018.*", the cue is shown in bold and the scope is underlined.

The natural language processing (NLP) community has paid considerable attention to uncertainty and negation detection (Farkas et al., 2010a; Morante and Blanco, 2012). Moreover, several corpora annotated for negation and uncertainty have been proposed in the biomedical domain (Vincze et al., 2008; Vincze, 2010; Uzuner et al., 2009). However, most of these proposals have focused on the English language, while information extraction in the medical domain represents its own challenges in languages other than English (Névéol et al., 2018). Most proposals developed for medical texts written in Spanish (Santiso et al., 2018, 2020; Cotik et al., 2016; Costumero et al., 2014), have been focused only on negation detection. Uncertainty detection for Spanish medical texts has not yet been sufficiently addressed and can be improved.

Negation and uncertainty detection have been widely addressed using rule-based approaches (Chapman et al., 2001; Harkema et al., 2009; Kesterson et al., 2015), and classical machine learning-based approaches (Cruz Díaz et al., 2012; Morante and Daelemans, 2009b; Jiménez-Zafra et al., 2021). Rule-based methods can suffer from a lack of flexibility and universality (Zhou et al., 2018). While classical machine learning methods depend on hand-crafted features, and they often require a complex and time-consuming feature engineering process and analysis to obtain a good performance (Minaee et al., 2020). Recently, deep learning approaches have been shown to improve performance at processing natural language texts in several tasks such as named entity recognition (NER) (Lample et al., 2016), question answering (Bordes et al., 2014), and language translation (Sutskever et al., 2014). One of the advantages of deep learning approaches is they can automatically learn features from data, instead of adopting hand-crafted features. Embedding models such as Word2Vec (Mikolov et al., 2013), Glove (Pennington et al., 2014), and FasText(Bojanowski et al., 2017) have been popularly used in text processing applications. These models also have been applied in the biomedical field (Wang et al., 2018; Soares et al., 2019). Moreover, the development of contextual embeddings (Peters et al., 2018) and transformer-based models (Devlin et al., 2019) have shown that such representations are able to improve performance on a wide range of natural language processing tasks (Liu et al., 2020; Gu et al., 2020; Pires et al., 2019).

Motivated by improvements in deep learning methods to process natural language texts, in this paper we propose an approach for negation and uncertainty detection in clinical texts written in Spanish. This approach explores two deep learning methods to perform negation and uncertainty detection: i) Bidirectional Long-Short Term Memory with a Conditional Random Field layer (BiLSTM-CRF) and ii) Bidirectional Encoder Representation for Transformers (BERT). The proposed approach takes advantage of transfer learning techniques to perform uncertainty and negation detection in clinical texts. Transfer learning aims to transfer knowledge from pre-trained resources to improve the performance on a new target task (Liu et al., 2019a; Peng et al., 2020; Ortiz Suarez et al., 2012; Panigrahi et al., 2021). In this approach we exploit pre-trained resources such as word embeddings (Soares et al., 2019; Mikolov et al., 2013) and contextualized embeddings (Devlin et al., 2019) to perform negation and uncertainty detection as a sequence labeling task. The most significant contributions of this paper are:

- A deep learning-based approach for negation and uncertainty detection in clinical texts written in Spanish. The main advantage of this approach is the use of word embeddings and contextual embeddings to automatically represent text features, avoiding the time-consuming feature engineering process. In the Spanish language, most of the previous studies have focused only on negation detection. Meanwhile, our approach goes further and in addition to negation, it also performs uncertainty detection. Code developed in this approach is public accessible from GitHub[1].

[1] https://github.com/solarte7/NegationAndUncertainty

- Exploiting transfer learning techniques to perform negation and uncertainty detection in Spanish clinical texts. In particular, transfer learning exploitation is applied in two ways: i) Creating pre-trained clinical embeddings and ii) Fine-tuning the BERT model. The generated clinical embeddings improve the performance of the BiLSMT-CRF neural model for detecting uncertainty and negation. Furthermore, the BERT model is fine-tuned with a classification layer on top. To the best of our knowledge, this is the first approach that uses a transformer-based method to perform both uncertainty and negation detection in clinical text written in Spanish.
- Improvement of performance in the scope recognition task, compared to other proposals in the biomedical domain for the Spanish language. Results obtained with both BiLSTM-CRF and BERT models have shown an improvement over other studies. Performed tests were evaluated using NUBES (Lima Lopez et al., 2020) and IULA (Marimon et al., 2017), two public corpora for the Spanish language. Obtained results in the scope recognition task have shown an F-score of 92% and 80% for negation and uncertainty detection, respectively.
- Validation of the proposed approach with a real-life dataset which contains annotations of patients diagnosed either with lung or breast cancer. To perform this validation, we used trained models on the NUBES corpus (Lima Lopez et al., 2020) to predict negation and uncertainty in this new dataset. The validation process shows the ability of deep learning-based models to predict negation and uncertainty in a different dataset to the one they were trained on.

The remainder of this paper is organized as follows: Section 2 shows previous studies about uncertainty and negation detection in the biomedical domain. In Section 3 datasets and proposed methods to perform negation and uncertainty detection are described. Section 4 explains the experiments carried out to validate our approach, and Section 5, provides a discussion of the results obtained. Finally, Section 6 includes conclusions and future work.

2 RELATED WORKS

The high percentage of uncertain and negated sentences within clinical texts has motivated more research on this field. In particular, 12% of the sentences contained in Medline abstracts are uncertain, and 20% are negated (Vincze, 2014). Several annotated corpora have been proposed in the biomedical domain (Vincze et al., 2008; Vincze, 2010; Uzuner et al., 2009). In the Bioscope corpus (Vincze et al., 2008), 18% of the sentences contain uncertain findings (Szarvas et al., 2012a). Uncertainty and negation detection has been commonly addressed by three approaches that we will review as follows: i) rule-based, ii) machine learning-based, and iii) deep learning-based.

2.1 Rule-based approaches

Rule-based approaches use declarative methods for creating manually crafted rules that extract uncertain and negated findings. One of the most widely used rule-based algorithm to detect negation in medical records is NegEx (Chapman et al., 2001). This algorithm has been recognized as one of the most useful approaches for the detection of negated medical concepts. However, several studies have been proposed to improve NegEx's performance (Elazhary, 2017; Harkema et al., 2009; Kesterson et al., 2015). In particular, the ConText algorithm (Harkema et al., 2009) extends NegEx for determining whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient. The proposal described in (Wu et al., 2011) also extends the NegEx algorithm to detect uncertainty by adding a separate category of uncertainty terms. In Velupillai et al. (2014), the authors proposed ConTextSwe, an adaptation of the ConText algorithm to Swedish. Several proposals have also been presented to adapt the NegEx algorithm to Spanish (Costumero et al., 2014; Stricke et al., 2015; Santamaria, 2019), but these proposals have only focused on negation detection.

The above-mentioned proposals use a similar approach to recognize the scope; that is the search for a termination term in a lexicon which indicates the end of the scope. However, one disadvantage of this approach occurs when the sentence does not contain any termination term. In these cases, the scope recognition fails because all tokens in a sentence can be taken as the scope. In these cases, the scope recognition fails because all tokens in a sentence can be taken as the scope. To deal with this problem, other studies have included the use of syntactic properties of the sentence to extract the scope (Cotik et al., 2016; Zhou et al., 2015; Peng et al., 2018). Although rule-based approaches have been widely used in the biomedical domain, their main disadvantages are the large amount of time it takes to create rules manually, and the lack of flexibility and universality (Zhou et al., 2018).

2.2 Machine learning-based approaches

In machine learning-based approaches, negation and uncertainty detection is formulated as a classification problem where both cues and scope detection are considered as sequence labeling tasks. These approaches commonly follow two steps: in the first step, hand-crafted features are extracted from documents and, in the second step those features are trained into a classifier to perform predictions. Early machine learning-based proposals deal only with recognizing negation and uncertainty at the sentence level (Clausen, 2010; Skeppstedt et al., 2016; Velupillai et al., 2011; Shaodian et al., 2016). In these cases, the complete sentence is considered an uncertain or negated fact. However, it is necessary to identify what tokens in the sentence are affected and which are not. One of the firsts machine learning proposals which deal with scope recognition is described in Morante and Daelemans (2009b). This study uses four classifiers for negation detection using approaches such as Support Vector Machines (SVM) and Conditional Random Fields (CRF) (Lafferty et al., 2001). The approach was evaluated using the Bioscope corpus (Vincze et al., 2008) and obtained an F1-score of 80.4% in the scope recognition task. In Morante and Daelemans (2009a), the above proposal was extended to deal with uncertainty detection.

Another proposal for negation detection is described in Agarwal and Yu (2010a). In this work, negation cues and their scope are detected in clinical notes and biological literature using CRF as a machine learning algorithm. With the goal of detecting both negation and uncertainty, Cruz Díaz et al. (2012) proposed a two phases machine learning model. The first phase classifies the cues, and the second predicts the scope. Reported results showed an F1-score of 91% and 72% for negation and uncertainty scope, respectively. All the machine learning-based proposals mentioned above use the BioScope corpus (Vincze et al., 2008), which is focused on the English language. In the case of the Spanish language, Santiso et al. (2018) proposed a CRF-based classification model for negation detection in clinical records. This study was evaluated using IULA (Marimon et al., 2017), a corpus annotated with negation in clinical text written in Spanish, and obtains an F1-score of 81% in scope recognition.

Although classical machine learning-based approaches have addressed limitations of rule-based methods, one disadvantage of these approaches is the reliance on the hand-crafted features that require tedious, time-consuming feature engineering along with analysis to obtain good performance (Minaee et al., 2020). This fact can be seen in studies such as the ones reported in Jiménez-Zafra et al. (2020), where a CRF-based system is trained with a considerable set of hand-crafted features to perform negation detection.

2.3 Deep learning-based approaches

The core component of these approaches is the use of word embedding models that map a set of texts into a low-dimensional continuous space (Mikolov et al., 2013; Pennington et al., 2014; Bojanowski et al., 2017). Contextualized embeddings such as ELMO (Peters et al., 2018) and BERT (Devlin et al., 2019) have also shown that such representations are able to improve performance on sequence labeling tasks. Recurrent neural networks (RNN) (Goldberg and Hirst, 2017; Hochreiter and Schmidhuber, 1997), and Convolution Neural Networks (CNN) (Lopez and Kalita, 2017) have also been used to process text in the biomedical domain.

In Qian et al. (2016), the authors proposed a Convolutional Neural Network-based model to extract negation and uncertainty scope from biomedical texts written in English. This model first extracts path and position features from syntactic trees with a convolutional layer that's features are concatenated into one feature vector. This vector is finally fed into the softmax layer to obtain the output vector. Tests and validation are conducted using the Bioscope corpus, showing the ability of the deep neural approaches to deal with negation and uncertainty detection. In Fancellu et al. (2017) is proposed a Bidirectional Long-Short Term Memory (BiLSTM) model to extract negation scope from English and Chinese texts. The performance shows an F-score of 89% for English and 79% for Chinese using the CNeSp corpus (Zou et al., 2016). The conclusion of that research is a suggestion to use more training data to make progress on negation detection. In Taylor and Harabagiu (2018), the authors proposed a BiLSTM neural model for negation detection from electroencephalography reports written in English. Reported results show an F-score of 88% for the scope recognition task using the Bioscope corpus. In Bhatia et al. (2018), an encoder-decoder neural architecture that combines a shared encoder and different decoding schemes to jointly extract entities and negations is proposed. Reported results show 90% in F-score for negation detection using data from the 2010 i2b2/VA challenge task (Uzuner et al., 2011).

An attention mechanism to perform uncertainty cue identification using data from CoNLL2010 shared task (Farkas et al., 2010b) is developed in Adel and Schütze (2017). Reported results have shown that combining attention layers with RNN and CCN models increases the performance for uncertainty detection in the English language, showing 85% in F1-score. The attention layer helps the model to recognize which part of the input data is important during the training, allowing the networks to focus on specific information by generating a weight vector. Another proposal that combines an attention layer with an RNN for detecting negated, possible and hypothetical medical findings from clinical notes is described in Chen (2019). In Khandelwal and Sawant (2020), the authors proposed NegBert, a model for negation detection using BERT contextual embeddings. NegBert has been trained for the English language using three corpora from different domains, including the BioScope corpus for the biomedical domain (Vincze et al., 2008). Reported results have shown an improvement in the scope resolution task with 93% in F1-score and comparable results in the cue identification task. In Shaitarova et al. (2020), the authors extended the Khandelwal and Sawant (2020) proposal to deal with both uncertainty and negation detection using transformer-based architectures such as BERT (Devlin et al., 2019), XLNet (Yang et al., 2019) and Roberta (Liu et al., 2019b). This study was also focused on the English language using the Bioscope corpus and the SFU Review corpus (Konstantinova et al., 2012).

In recent years, the interest in processing negation and uncertainty has also grown in languages other than English. In Dalloux et al. (2019), a BiLSTM-CRF model focused on the French language is presented. The approach was validated using a dataset from the National Cancer Institute in France (NCI[2]) and obtained an F-score of 90% and 86% for negation and uncertainty scope detection, respectively. (Al-khawaldeh, 2019) proposed an Attention-based BiLTSM model to perform speculation detection in Arabic medical texts. This proposal obtained an F-score of 73.5% in the scope recognition task using the BioArabic corpus (Al-khawaldeh, 2016).

In Santiso et al. (2020) is proposed a BiLSTM-based model for negation detection in the Spanish language. This proposal uses the IxaMed-GS corpus (Oronoz et al., 2015), a private dataset which consists of 75 clinical notes written in Spanish. Reported results showed 83% in F-score for the scope recognition task. In Zavala and Martinez (2020), the authors compare the performance of different deep learning-based models to perform negation and speculation detection using several corpora from English and Spanish. In the case of the Spanish language, performed tests were conducted using IULA (Marimon et al., 2017) and SFU ReviewSP-NEG (Jiménez-Zafra et al., 2018), two corpora focused only on negation annotation, but not on uncertainty. Finally, in Lima Lopez et al. (2020) is proposed NUBES, a corpus with both negation and uncertainty annotations from clinical notes written in Spanish. Moreover, the authors provide a BiLSTM-based model for testing with this corpus. Reported results showed an 90% and 78% in F1-score for negation and uncertainty scope detection, respectively. Although this study showed promising results, the main disadvantage is the reliance on several hand-crafted features used to feed the system, which makes the feature engineering process time-consuming.

Table 1 shows a summary of the most relevant deep learning-based approaches to perform negation and uncertainty detection. From this table, it is important to highlight the following facts: i) Most of the proposals have focused on the English language, as a result of the availability corpora (Vincze et al., 2008; Uzuner et al., 2011; Farkas et al., 2010b); ii) Most of the existing proposals developed for the Spanish language concentrate only on negation detection, iii) Although the proposal described in (Lima Lopez et al., 2020) aims to detect uncertainty and negation in Spanish, its main weakness is the dependence on hand-crafted features. This suggests that detecting both uncertainty and negation in Spanish clinical texts can be improved.

3 MATERIALS AND METHODS

In this section, we first show the datasets used for training, testing, and validating the proposed approach. Next, we will describe the deep learning-based methods for negation and uncertainty detection.

[2]https://en.e-cancer.fr/

Table 1. Summary of Deep learning-based approaches

Proposal	Approach	Language	Corpus	Negation	Uncertainty
Qian et al. (2016)	CCN	English	Bioscope	yes	yes
Fancellu et al. (2017)	BiLSTM	English & Chinese	Bioscope & CNeSp	yes	no
Taylor and Harabagiu (2018)	BiLSTM	English	BioScope	yes	no
Uzuner et al. (2011)	Encoder-Decoder	English	i2b2/VA challenge	no	yes
Adel and Schütze (2017)	RNN + Attention	English	Bioscope	yes	yes
Chen (2019)	RNN + Attention	English	i2b2/VA NLP challenge	yes	no
Khandelwal and Sawant (2020)	BERT	English	Bioscope	yes	no
Dalloux et al. (2019)	Bi-LSTM	French	NCI - France	yes	yes
Al-khawaldeh (2019)	Attention + BiLSTM	Arabic	Bio Arabic	no	yes
Santiso et al. (2018)	Embeddings + CRF	Spanish	IULA	yes	no
Santiso et al. (2020)	BiLSTM	Spanish	IxaMed-GS	yes	no
Zavala and Martinez (2020)	BiLSTM, BERT	Spanish	IULA	yes	no
Lima Lopez et al. (2020)	BiLSTM	Spanish	NUBES	yes	yes

3.1 Datasets

NUBES (Lima Lopez et al., 2020) and IULA (Marimon et al., 2017) are two public corpora available for the Spanish language that will be used to train models. Additionally, an in-house annotated dataset with real-life data of cancer patients was manually annotated and will be used for validation purposes. Details about each dataset are given as follows:

- **NUBES**[3]: a public corpus which consists of 29,682 sentences obtained from anonymized health records annotated with negation and uncertainty (Lima Lopez et al., 2020). NUBES is the largest publicly available corpus for negation in Spanish clinical records, and the first corpus that also incorporates the annotation of uncertainty. This corpus contains annotations for syntactic, lexical, and morphological negation cues.
  - **Syntactic negation**: are cues represented by function words or adverbs, for instance: not, without, never *(“no, sin, nunca”)*
  - **Lexical negation**: are words or multi-word expressions which indicate negation depending on the context. They include verbs, adjectives or noun phrases, for example: negative, denies, withhold *(“negativo, niega, suspender”).*
  - **Morphological Negation:** are words which refer to negation by means of affixes, for instance: afebrile, asymptomatic *(“Afebril, Asintomático”).*

  In the case of uncertainty, the NUBES corpus contains annotations for lexical and syntactic cues. Lexical cues are words that express uncertainty depending on the context. Lexical cues include words such as “probable”, “possible”, and “compatible with” *(“Probable, posible, compatible con”).* Syntactic cues include only the disjunctions words “Versus”, “Vs”, “Or” *(“Versus, vs, o”).* These cues were only annotated when they appeared by themselves in a context of uncertainty.
- **IULA**[4]: a public corpus which contains 3,194 sentences extracted from anonymized clinical records and manually annotated with negation cues and their scopes (Marimon et al., 2017). This corpus

[3] https://github.com/Vicomtech/NUBes-negation-uncertainty-biomedical-corpus
[4] http://eines.iula.upf.edu/brat/#/NegationOnCR_IULA/

was extracted from clinical notes from one hospital in Barcelona (Spain) and contains annotations only with negation. Syntactic negation, and lexical negation cues have been annotated in this corpus, but not morphological negation.

- **Cancer dataset**: an in-house manually annotated dataset with data from patients either suffering with lung or breast cancer. This dataset was extracted from real-life clinical notes belonging to cancer patients from ”Hospital Universitario Puerta de Hierro, Madrid Spain”. The dataset contains 2,700 sentences annotated with both negation and uncertainty. We have received informed consent from participants of our study before annotating this dataset. The Cancer dataset was annotated with syntactic, lexical and morphological negation. In the case of uncertainty, this dataset contains syntactic and lexical cues with their respective scopes. In the Cancer dataset, negation was more frequently found in sentences that describe symptoms, medical tests results, and treatments. On the other hand, uncertainty was more frequently found in sentences that describe cancer diagnosis. Figure 1 shows a set of sentence examples extracted from the Cancer dataset. These sentences show negation and uncertainty cues and their scopes.

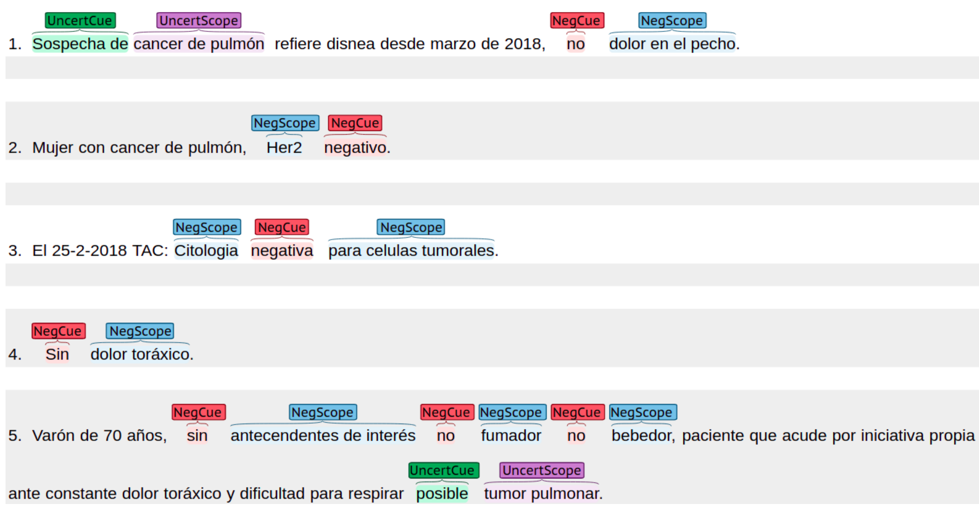

Figure 1. Sentences with negation and uncertainty cues and their scopes.

Table 2 shows a descriptive analysis of the datasets previously mentioned. This analysis aims to check how cues and scopes behave in Spanish clinical texts. The analysis has shown that negation and uncertainty have specific features as follows:

- **Sentence length:** is the number of tokens in sentences where uncertainty or negation appears. According to Table 2, the analyzed datasets have similar values in indicators such as median, first quartile, and third quartile. In the case of the NUBES corpus, the value of median is 14, which indicates that 50% of the sentences have 14 tokens or fewer. Similar behavior occurs in the IULA corpus and the Cancer dataset, where the median is 10 and 12, respectively. This fact suggests that in clinical texts written in Spanish negation and uncertainty can frequently occur in short sentences. However, there are also large and more complex sentences where several cues can appear as well as negation as uncertainty. In fact, negation and uncertainty can appear in the same sentence. According to Figure 1, in the fourth example there is a short sentence, while the fifth example shows a large and more complex sentence.

- **Cues:** are distributed as follows; in the NUBES corpus 85% of negation annotations contain syntactic negation, 6% lexical, and 9% morphological negation. Similar values can be found in the IULA and the Cancer datasets. In the case of uncertainty, 98% of annotations contain lexical uncertainty and only 2% syntactic uncertainty, for the case of the NUBES corpus. Similar values can be found in the Cancer dataset.

- **Scopes:** can behave in a continuous or discontinuous way. A continuous scope occurs when the tokens affected by a specific negation or uncertainty cue are continuous in the sentence. On the other hand, a discontinuous scope occurs when the sequence of tokens affected by a cue are separated in different positions of the text sentence. In Figure 1, the first and the fourth sentences contain continuous scopes. In contrast, the third sentence contains a discontinuous scope. According to Table 2, most cases are continuous scopes. In the NUBES corpus, 95% of annotations correspond to continuous scopes and only 5% to discontinuous scopes. Similar behavior can be seen in the IULA and the Cancer datasets. Additionally, the scope can appear after the cue, before the cue, or on both sides. According to Figure 1, in the fourth sentence the scope appears to the right (after) of the cue *"Sin"*. In the second sentence, the scope appears to the left (before) of the cue *"negativo"*. Meanwhile, in the third example, the scope is in both sides of the cue.

Table 2. A summary of the datasets used in the proposed approach

Indicators	NUBES	IULA	Cancer Dataset
Number of sentences	29,682	3,194	2,700
Sentences with negation	25.5%	34%	27%
Sentences with uncertainty	7.5%	-	12%
Maximum number of tokens	210	159	181
Mean (Number of tokens)	18	14	15
Median (Number of tokens)	14	10	12
First quartile	9	6	7
Third quartile	23	19	18
Syntactic negation cues	85%	92%	83%
Lexical negation cues	6%	8%	10%
Morphological negation cues	9%	-	7%
Syntactic uncertainty cues	2%	-	1%
Lexical uncertainty cues	98%	-	99%
Continuous scopes	95%	96%	97%
Discontinuous scopes	5%	4%	3%

3.2 Deep learning-based methods for negation and uncertainty detection

The proposed approach addresses negation and uncertainty detection as a sequence-labeling task, where each token in a sentence is classified as being part of the cue or the scope. This approach recognizes cues and their scopes in a single stage for both negation and uncertainty. The BIO tagging format (short for Beginning, Inside, Outside) is used to represent predicted entities. For instance, the sentence: *"Paciente* ***sin*** *dolor torácico."* (Patient without chest pain.), it can be formatted as:

['Paciente : O', 'Sin : B-NegCue', 'dolor : B-NegScope', 'torácico : I-NegScope', '. : O '].

To perform negation and uncertainty detection from clinical text written in Spanish, we explore two deep learning-based models: BiLSTM-CRF and BERT.

3.2.1 BiLSTM-CRF

The first model is a Bidirectional Long-Short Term Memory with a CRF layer (BiLSTM-CRF) neural net. This model is based on neural architectures described in (Lample et al., 2016; Huang et al., 2015; Collobert et al., 2011) and consist of three layers: Embedding layer, BiLSTM layer, and CRF layer (see Figure 2).

- **Embedding layer:** This layer allows the approach to automatically represent text features using dense vector representations. In these vectors, words with a similar context in the text have a similar representation. The Embedding layer enables the approach to represent each word into a fixed length vector of defined size. We use two different word embeddings in this approach:

- *Biomedical embeddings*: These embeddings were trained using full-text medical papers written in Spanish (Soares et al., 2019). The papers were taken from Scielo (a scientific electronic library[5]), and a subset of Wikipedia articles related to Pharmacology, Medicine, and Biology.
- *Clinical embeddings*: We create in-house embeddings trained with more than 1 million clinical notes written in Spanish. These clinical documents were provided in a raw format by two public hospitals in Madrid (Spain) and Cali (Colombia). We use the FasText (Bojanowski et al., 2017) method for creating word embeddings using a vector size of 300 positions by default.

- **BiLSTM layer**: the Bidirectional LSTM (BiLSTM) layer captures both left and right contexts of words to produce a vector representation of text sequences. Given a sentence $(w_1, w_2, w_3, \ldots w_n)$ as input, where n is the number of words, this layer processes the sentence using two steps:
  - A Forward step processes the sentence from left to right, where each element f_i represents the value of the left context for the word w_i in the sentence.
  - A Backward step processes from right to left where each element b_i represents the value of the right context for the word w_i in the sentence.

  As an output, the BiLSTM layer generates a vector representation h_i for each word by concatenating the values f_i and b_i. The output vector h_i contains a sequence of probabilities for each label to be predicted. The BiLSTM layer computes the Forward and Backward steps separately. Therefore, the values f_i and b_i are calculated independently.

- **CRF layer:** this layer predicts the label sequence with the highest prediction score from all sequences generated by the BiLSTM layer. Although the BiLSTM layer generates probabilities for each label to be predicted, these probabilities are calculated independently. For sequence labeling tasks it is crucial to consider correlations and dependencies across output labels. Therefore, this layer uses an implementation of the CRF algorithm (Lafferty et al., 2001) to improve the predictions for each label. The CRF algorithm considers correlations between other labels and jointly decodes the best chain of labels for a given input text sentence.

3.2.2 Bidirectional Encoder Representation for Transformers (BERT)

BERT (Devlin et al., 2019) uses a transformer-based architecture to learn representation of texts in a bidirectional way by considering both the left and the right context of words. In the proposed approach, the BERT model is fine-tuned with a classification layer on top. We use multilingual BERT as contextualized embeddings to perform negation and uncertainty detection in clinical notes written in Spanish. Multilingual BERT has been pre-trained on data in 104 languages with the same training objectives as BERT: masked language modeling and next-sentence prediction.

Figure 3 shows the process to detect negation and uncertainty using multilingual BERT. This process consists of three steps: Tokenization, BERT Processing, and Classification & Post-processing.

- **Tokenization:** the goal in this step is to take as input a raw text sentence and tokenize it using a WordPiece Tokenization method (Wu et al., 2016). For each word in the sentence, this method decides to keep the whole word or to split it into a set of sub-words. According to Figure 3, after tokenizing the input sentence, the output tokens are: *"La bio ##psia no muestra células cancer ##igen ##as"*. The word *"biopsia"* has been divided into two sub-words: *"bio"* and *"psia"*. Furthermore, the word *"cancerígenas"* was divided into three sub-words: *"cáncer"*, *"igen"* and *"as"*. This tokenization method aims to improve handling of rare and unseen words in a dataset by providing a balance between the flexibility of character-delimited tokenizers and the efficiency of word-delimited tokenizers. Additionally, in this step two special tokens are added to the sentence: [CLS] and [SEP]. The [CLS] token always appears at the beginning of the text, and the [SEP] token is used to separate sentences.

[5] https://scielo.isciii.es/scielo.php

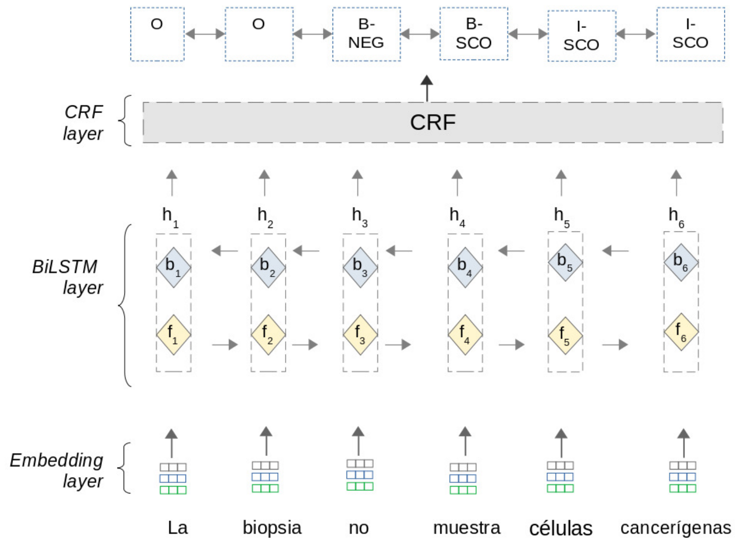

Figure 2. Negation and uncertainty detection using the BiLSTM-CRF model

- **BERT Processing:** In this step, the approach takes as input the tokenized sentence from the previous step and process it as follows:
  - First, the approach obtains an embedding representation (E_n) for each word in the sentence. This representation is created using three embeddings: token, segment, and position embeddings. The token embedding contains a vector representation for each word. The segment embedding is used to distinguish the vector representation for two sentences in a sentence pair. Finally, the position embedding is used to specify the position of words in the text sentence.
  - Next, the BERT Transformer Block takes the embedding representation as input (E_1, E_2, E_n), and produces a final representation (R_n) for each word in the processed sentence. This representation is a score calculated by BERT and represents a contextualized value for a specific word in relation to all other words.
- **Classification & Post-Processing:** In this step, the approach takes as input the predicted BERT representations (R_1, R_2, R_n) and feeds them into the softmax function. This function obtains a label for each token in the sentence. A post-processing step is needed to convert BERT predictions into BIO format labels. The probability P for each label is calculated as follows:

$$P\ (l|\mathbf{R}_i) = \mathbf{Softmax}(\mathbf{W}_0\mathbf{R}_i + \mathbf{b}_0) \tag{1}$$

where the label l belongs to the set of labels to be predicted, W_0 and b_0 are weight parameters. Finally, the special tokens '[CLS], [SEP], [PAD]' are removed to obtain the final BIO labels at the end of post-processing step.

4 EXPERIMENTATION AND RESULTS

In this section, we describe experiments carried out to evaluate the proposed approach for negation and uncertainty detection. We will first describe the evaluation methodology, then the experiments that were carried out, followed by the results that were obtained.

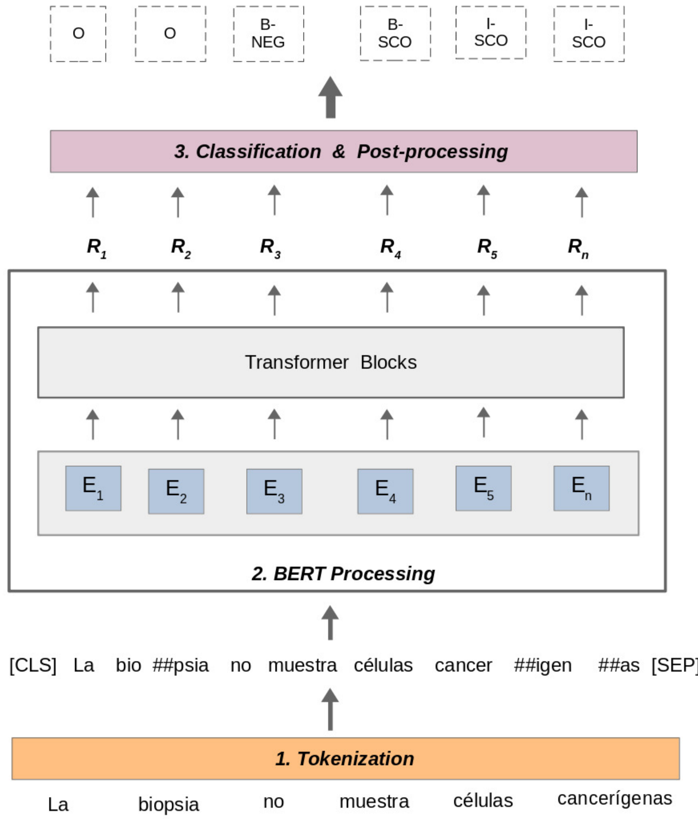

Figure 3. Negation and uncertainty detection using multilingual BERT.

4.1 Evaluation methodology

The evaluation methodology depends on the dataset used, as follows:

- NUBES corpus: this corpus was split by their authors into three subsets: training (75%), developing(10%), and testing (15%). Models trained with the NUBES corpus were executed just once. The testing subset was used to calculate the performance metrics.

- IULA corpus: this corpus does not provide an explicit division for training and testing subsets. Therefore, in this case we followed a cross-validation strategy with $k = 5$. The performance was calculated as the average of all five folds executed by the cross-validation strategy.

To evaluate the performance of the proposed approach, we used the following standard metrics: Precision (P), Recall (R), and F-score (F1). The F-score is calculated as a weighted average of the Precision and Recall measurements. A token is correctly classified when the predicted label is equal to the label indicated by the annotated corpus. The performance for the cue identification task and the scope recognition task is analyzed separately.

$$\textbf{Precision} = \frac{\text{Number of tokens correctly predicted}}{\text{Number of predicted tokens}} \tag{2}$$

$$\textbf{Recall} = \frac{\text{Number of tokens correctly predicted}}{\text{Number of tokens in the dataset}} \tag{3}$$

$$\textbf{F-score} = 2 * \frac{\text{Precision * Recall}}{\text{Precision + Recall}} \tag{4}$$

4.2 Experiments

The BiLSTM-CRF model proposed by Huang et al. (2015) was used as baseline system to perform negation and uncertainty detection. Next, we performed the following experiments:

1. Experiment 1: Spanish biomedical embeddings proposed by Soares et al. (2019) were added to the BiLSTM-CRF model. The goal of this experiment was to analyze the impact of adding biomedical embeddings to the BiLSTM-CRF model.
2. Experiment 2: the goal of this experiment was to test the impact of using embeddings trained on clinical notes written in Spanish. Consequently, in-house clinical embeddings (Section 3.2.1) were added to the BiLSTM-CRF model.
3. Experiment 3: the goal of this experiment was to analyze the impact of using multilingual BERT embeddings to perform negation and uncertainty detection in Spanish. To perform this experiment, we used the BERT model as described in Section 3.2.2.

4.3 Validation

The Cancer dataset described in section 3.1 is used for validating the performance of trained models on the NUBES corpus. We used these models for validation because the NUBES corpus contains annotations for both negation and uncertainty (the IULA corpus cannot be used as it does not contain uncertainty annotations). This validation aims to measure the performance of trained models on the NUBES corpus to predict negation and uncertainty in a new dataset. Results of the validation are shown in table 5.

4.4 Implementation and Hyperparameters setting

To perform the previously explained experiments Python 3.7, TensorFlow[6] and Keras[7] were used. For the BiLSTM-CRF model the following parameters were settled: learning rate as 0.001, dropout as 0.5, the number of epochs was set to 60, the BiLSTM hidden size was set to 300, and the batch size to 512. For the BERT model, the fine-tuning was performed with a sequence length of 256 tokens, a batch size of 64, and 5 epochs. These values were established after training the models different times, and checking the best performance for these parameters. Data and code of the proposed approach can be found in GitHub[8].

4.5 Results

In order to analyze obtained results, we first show the results of experiments for cue identification, then results for scope recognition, and finally, the validation results.

4.5.1 Cue identification

Table 3 shows the results obtained for the cue identification task in the experiments previously described. These results show the feasibility of both BiLSTM-CRF and BERT models to perform negation and uncertainty cue identification in clinical texts written in Spanish. As can be seen, the best performance was obtained in the third experiment in which the model was trained using multilingual BERT. Using the NUBES corpus, this model obtained an F-score of 95% and 84% for negation and uncertainty detection, respectively. While for the IULA corpus, the model obtained an F-score of 92% for negation detection.

In addition, Table 3 shows that when the BiLSTM-CRF model is combined with biomedical and clinical embeddings, it obtains competitive results in the cue identification task. In the first experiment, the BiLSTM-CRF model obtained an F-score of 93% and 83% for negation and uncertainty, respectively. In the second experiment an F-score of 92% and 82% were obtained. These results suggest that using biomedical and clinical embeddings is a useful approach to improve the performance of the BiLSTM-CRF model to detect negation and uncertainty in Spanish clinical texts. Moreover, using biomedical and clinical embeddings also improved the performance of the BiLSTM-CRF model in the IULA corpus.

According to Table 3, negation detection showed better performance than uncertainty detection. In the first experiment, the BiLSTM-CRF model obtained an F-score of 93% for negation detection and 83% for uncertainty detection. Meanwhile, models trained using multilingual BERT obtained an F-score of 95% for negation detection and 84% for uncertainty detection. Thus, these results highlight the fact that uncertainty detection is more difficult than negation detection in clinical texts written in Spanish.

[6]https://www.tensorflow.org/?hl=es-419
[7]https://keras.io/
[8]https://github.com/solarte7/NegationAndUncertainty

Table 3. Results for cue identification.

	NUBES corpus						IULA corpus		
	Negation			Uncertainty			Negation		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
BiLSTM-CRF	0.86	0.82	0.83	0.79	0.76	0.77	0.82	0.78	0.80
BiLSTM-CRF + Biomedical Embbedings	0.94	0.92	0.93	0.85	0.81	0.83	0.91	0.90	0.90
BiLSTM-CRF + Clinical Embbedings	0.93	0.91	0.92	0.84	0.80	0.82	0.90	0.88	0.89
Multilingual BERT	0.95	0.93	**0.95**	0.86	0.83	**0.84**	0.92	0.93	**0.92**

4.5.2 Scope Recognition

Table 4 describes the results obtained in the scope recognition task. These results show the feasibility of deep learning-based methods to perform the scope recognition task for both negation and uncertainty detection. The best performance was obtained by using the BERT model. A 92% F-score for negation and 80% F-score for uncertainty were obtained using the NUBES corpus. Using the IULA corpus, this model obtained an F-score of 89% for negation detection.

The BiLSTM-CRF model combined with biomedical and clinical embeddings also showed a competitive performance in the scope recognition task. In the first experiment, it obtained an F-score of 90% for negation and 79% for uncertainty detection using the NUBES corpus. In the second experiment, 89% and 78% were obtained for negation and uncertainty, respectively. Results obtained suggest that adding clinical and biomedical embeddings increases the ability of the BiLSTM-CRF model to perform the scope recognition task. Results obtained in the IULA corpus also show that using biomedical and clinical embeddings has a positive impact on the performance of the BiLSTM-CRF model.

Table 4 shows that negation detection performs better than uncertainty detection in experiments carried out with the NUBES corpus. This suggests that extracting the uncertainty scope is more difficult than extracting the negation scope. This behavior can be explained by the fact that negation also performs better than uncertainty in the cue identification task. Therefore, the difficulties in extracting uncertainty cues can also affect scope recognition, since cue identification and scope recognition are related tasks.

Table 4. Results for scope recognition.

	NUBES corpus						IULA corpus		
	Negation			Uncertainty			Negation		
	P	R	F1	P	R	F1	P	R	F1
BiLSTM-CRF	0.84	0.76	0.79	0.72	0.69	0.70	0.77	0.74	0.75
BiLSTM-CRF + Biomedical Embbedings	0.92	0.89	0.90	0.82	0.77	0.79	0.88	0.84	0.86
BiLSTM-CRF + Clinical Embbedings	0.92	0.87	0.89	0.81	0.75	0.78	0.87	0.83	0.85
Multilingual BERT	0.93	0.90	**0.92**	0.82	0.79	**0.80**	0.91	0.86	**0.88**

4.5.3 Validation results

Table 5 shows obtained results for the validation process with the Cancer dataset. These results show the performance of trained models with the NUBES corpus for predicting cues and scopes on the Cancer dataset. Table 5 shows that the best performance was obtained by using the BERT model. In the cue identification task, this model obtained an F-score of 90% and 82% for negation and uncertainty, respectively. In the scope recognition task, the BERT model obtained an F-score of 87% and 78% for negation and uncertainty, respectively. In addition, the performance of the BiLSTM-CRF model combined with biomedical and clinical embeddings also showed competitive results. For instance, the model that uses clinical embeddings obtained an F-score of 89% and 80% in cue identification for negation and uncertainty, respectively. In the scope recognition task, this model obtained an F-score of 86% and 77% for negation and uncertainty, respectively.

Although results from Table 5 show a lower performance compared to those described in Tables 3 and 4, results are still promising. The validation process showed the ability of the deep learning-based models trained with the NUBES corpus to predict negation and uncertainty in a different dataset. This fact suggests that models trained with the NUBES corpus can be used to detect negation and uncertainty in new clinical texts.

Table 5. Validation Results

	Cue detection						**Scope recognition**					
	Negation			**Uncertainty**			**Negation**			**Uncertainty**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BiLSTM-CRF + Biomedical Embbedings	0.89	0.87	0.88	0.75	0.80	0.78	0.86	0.83	0.84	0.79	0.74	0.76
BiLSTM-CRF + Clinical Embbedings	0.91	0.88	0.89	0.80	0.81	0.80	0.87	0.85	0.86	0.79	0.76	0.77
Multilingual BERT	0.91	0.89	**0.90**	0.84	0.80	**0.82**	0.89	0.86	**0.87**	0.79	0.78	**0.78**

Comparing the performance of the proposed approach with other studies in the literature, we found that in the cue identification task, our approach obtains competitive results compared to those reported by (Santiso et al., 2018, 2020; Costumero et al., 2014). However, those proposals only perform negation detection, and do not deal with uncertainty detection.

In the scope recognition task, the proposed approach outperforms previous studies both for negation and uncertainty detection in Spanish. Table 6 shows the performance of the proposed approach in comparison with other proposals for the scope detection task. If one further analyzed the results from the NUBES corpus, the approach presented in this paper outperforms the Lima Lopez et al. (2020) proposal in both tasks (improvement of 2% for the case of uncertainty and 2% for the case of negation). If one analyzes the IULA corpus, the approach presented in this paper outperforms the results reported in Zavala and Martinez (2020) and Santiso et al. (2018), improving by 3% and 5%, respectively. In addition, the approach proposed in this paper presents the following advantages over other studies:

- The proposed approach takes advantage of transfer learning techniques, word embeddings, and pre-trained models to automatically represent text features using dense vector representations. These representations are used for negation and uncertainty detection. In contrast, other proposals such as Lima Lopez et al. (2020) require a considerable set of hand-crafted features to obtain comparable results. The process of manual feature extraction can be time-consuming and costly.

- The proposed approach deals with negation and uncertainty detection in a single step, improving those approaches (Santiso et al., 2018, 2020; Costumero et al., 2014; Cotik et al., 2016) that only deal with negation detection in Spanish clinical texts.

- This approach improves performance in the scope recognition task in comparison with approaches Lima Lopez et al. (2020) dealing both with negation and uncertainty in Spanish, as one can see in Table 6.

5 DISCUSSION

The proposed approach has shown the feasibility of deep learning-based methods to perform negation and uncertainty detection from clinical texts written in Spanish. We found that both BiLSTM-CRF and BERT models obtained competitive results for both tasks: cue identification and scope recognition. The use of biomedical embeddings for Spanish and contextualized embeddings from multilingual BERT results in an

Table 6. Comparison with other proposals in the scope recognition task (F-score).

	NUBES corpus		IULA corpus
Proposal	**Negation**	**Uncertainty**	**Negation**
Santiso et al. (2018)	-	-	0.83
Zavala and Martinez (2020)	-	-	0.85
Lima Lopez et al. (2020)	0.90	0.78	-
Our approach (BiLSTM-based)	0.90	0.79	0.86
Our approach (BERT-based)	**0.92**	**0.80**	**0.88**

improvement of the negation and uncertainty detection process. The proposed approach automatically represents text features using word embeddings and contextual embeddings, and uses them to detect uncertainty and negation. This is an advantage over previous proposals that require complex hand-crafted rules (Costumero et al., 2014; Cotik et al., 2016; Solarte-Pabón et al., 2020) or a considerable set of hand-crafted features (Lima Lopez et al., 2020) to obtain comparable results.

Obtained results showed that scope recognition is a more complex task than cue identification. This could be because in most cases negation and uncertainty cues consist of a single token. However, the scope frequently contains a longer sequence of tokens which makes it more difficult to detect it properly. As Table 4 shows, the proposed approach in this paper obtained competitive results for the scope recognition task, outperforming previous studies available for this task (Santiso et al., 2018; Lima Lopez et al., 2020).

The proposed approach performs better negation detection than uncertainty detection. Specifically, in the cue identification task, the best performance showed an F-score of 95% and 84% for negation and uncertainty detection, respectively (See Table 3). This behavior can be explained by two facts:

- The number of annotations with negation is higher than with uncertainty in the NUBES corpus. In this corpus, there are more than 7,500 sentences annotated with negation and only 2,219 sentences annotated with uncertainty, which affects the training of the models.

- Negation cues have less variability than uncertainty cues. Only five negation cues are used in the NUBES corpus *("no, sin, negativo, negativos, niega")* to express 87% of all negation annotations. However, the set of words used to express uncertainty is much broader. The five more frequent uncertainty cues *("probable, posible, compatible con, sospecha de, parece")* are used only in less than 48% of all uncertainty annotations. This fact once again affects the training process, reducing the performance of the models to perform uncertainty detection.

Although both BiLSTM-CRF and BERT models have shown feasibility to perform negation and uncertainty detection in Spanish, BERT obtains better results than BiLSTM-CRF in some specific cases. In particular, we observed that the BERT model tends to learn the scope better in those cases in which it appears before the cue, such as in the sentence:

*"Mujer con cáncer de pulmón, <u>HER2</u> **negativo**"* (Woman with lung cancer, HER2 negative),

In this sentence, the scope is underlined and the negation cue is in bold. In this case, the BERT model is able to learn that the scope is before the cue. The BiLSTM-CRF model failed in most of these cases. The BERT model also showed better performance than the BiLSTM-CRF model in those cases where the scope appears in both directions of the cue, as in the following sentence:

*"<u>TAC cerebral</u> **negativo** <u>para células tumorales.</u>"* (Negative brain CT scan for tumor cells.)

In this sentence, we can see that the scope (underlined) is before and after the cue (showed in bold). In these cases, the BERT model recognizes the scope in both directions (after and before the cue). On the other hand, the BiLSTM-CRF model tends to recognize only the part of the scope that appears after the cue.

The BERT model also outperforms the BiLSTM-CRF model when predicting labels with fewer annotations in the annotated corpus. In particular, the NUBES corpus contains more than 6500 annotations for *syntactic negation*, and only 460 for *lexical negation*. In the syntactic negation, both BiLSTM-CRF and BERT models perform accurately. However, for lexical negation, the BiLSMT model fails to recognize some lexical cues with a low number of annotations in the corpus, while the BERT model recognizes them properly.

Despite the promising results, there are still some limitations that need to be addressed. In particular, we identify the following causes of error which can affect the performance of the proposed approach:

- Those cues that appear rarely are not detected. These cases mainly occur in uncertainty detection where some syntactic cues such as "Versus" and "O" (or), are annotated as uncertainty cues. However, the NUBES corpus has very few annotations using these words which means the approach does not recognize them as cues. As a consequence, the approach fails in sentences such as these examples: (Cue shown in bold):

  – *"Cáncer pulmonar **Versus** inflamación del lóbulo derecho"*. (Lung cancer Versus right lobe inflammation).

  – *"Carcinoma estadio IV **o** estadio IIIB"*. (Stage IV or stage IIIB carcinoma.)

- In the scope recognition task, most errors are caused by discontinuous scopes. This occurs when the sequence of tokens affected by a negation cue or an uncertainty cue are separated in different positions of the text sentence. In the next example, the scope (shown underlined) is discontinuous:

  *"Paciente parcialmente orientado (si en tiempo, **no** en espacio)* (Partially oriented patient (in time, not in space))

  Another case of discontinuous scopes occurs when a sentence contains a sequence of cues, but some tokens in the sentence are not affected by any cue. A sequence of cues is more frequent in negation detection, as one can see in the following examples:

  *i) "**No** hipertensión arterial, **no** no vómito, **no** sangrado."* (No high blood pressure, no vomiting, no bleeding.)

  *ii) "**No** dolor, fiebre alta desde la noche anterior, **no** sangrado, **no** vomito."* (No pain, high fever since last night, no bleeding, no vomiting.)

  Note that in the first sentence, each token is affected by a negation cue. In this case, the approach correctly extracts the scope for each cue. However, in the second example, the tokens *"fiebre alta desde la noche anterior"* are not negated, but the approach classifies them as negated. Discontinuous scopes are not frequent. In the NUBES corpus, about 5% of annotated scopes are discontinuous. However, these cases should be analyzed further in the future to improve performance.

6 CONCLUSIONS

In this paper, we proposed a deep learning-based approach for negation and uncertainty detection in clinical texts written in Spanish. Results obtained have shown the ability of deep learning methods to perform both cue identification and scope recognition tasks. The proposed approach takes advantage of transfer learning techniques, word embeddings, and pre-trained models to automatically represent text features, thus avoiding the time-consuming feature engineering process. This approach is useful for medical text mining applications because it can recognize negation and uncertainty in a single step. Moreover, the approach improves previous studies, as has been shown.

Both BiLSTM and BERT models have shown promising results for negation and uncertainty detection in clinical texts written in Spanish. The BERT model performed better than the BiLSTM model in some specific cases. In the cue identification task, the BERT model performed better at predicting labels that have fewer annotations in a corpus. In the scope recognition task, this model worked better at extracting the scope that appears before the cue, or when it appears on both sides of the cue. These facts increased the performance of the BERT model over the BiLSTM-CRF model.

In this paper, we also conducted a validation process using real-life clinical data which showed promising results. Deep learning models trained on the NUBES corpus have shown the ability to predict negation and uncertainty in datasets that are different from those on which they were trained. In future studies, we will explore more transformer-based architectures to perform negation and uncertainty detection.

ACKNOWLEDGMENTS

This paper is supported by European Union's Horizon 2020 research and innovation program under grant agreement No. 875160, project CLARIFY (Cancer Long Survivors Artificial Intelligence Follow Up).

DATA AVAILABILITY

Data and code are available in:
https://github.com/solarte7/NegationAndUncertainty

The Cancer dataset is available "upon request". This dataset can be accessible after an evaluation by the hospital's ethics committee. To request access to the anonymized data, please contact Dr. Maria Torrente at the following email: maria.torrente@salud.madrid.org

REFERENCES

Adel, H. and Schütze, H. (2017). Exploring different dimensions of attention for uncertainty detection. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 1:22–34.

Agarwal, S. and Yu, H. (2010a). Biomedical negation scope detection with conditional random fields. *Journal of the American Medical Informatics Association*, 17(6):696–701.

Agarwal, S. and Yu, H. (2010b). Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of Biomedical Informatics*, 43(6):953–961.

Al-khawaldeh, F. T. (2016). Speculation and Negation Annotation for Arabic Biomedical Texts : BioArabic Corpus. *World of Computer Science and Information Technology Journal (WCSIT)*, 6(1):8–11.

Al-khawaldeh, F. T. (2019). Speculation and Negation Detection for Arabic Biomedical Texts. *World of Computer Science and Information Technology Journal (WCSIT)*, 9(3):12–16.

Alam, R., Cheraghi-Sohi, S., Panagioti, M., Esmail, A., Campbell, S., and Panagopoulou, E. (2017). Managing diagnostic uncertainty in primary care: A systematic critical review. *BMC Family Practice*, 18(1):1–13.

Bhatia, P., Celikkaya, B., and Khalilia, M. (2018). *Joint Entity Extraction and Assertion Detection for Clinical Text*. Springer International Publishing.

Bhise, V., Rajan, S. S., Sittig, D. F., Morgan, R. O., Chaudhary, P., and Singh, H. (2018). Defining and Measuring Diagnostic Uncertainty in Medicine: A Systematic Review. *Journal of General Internal Medicine*, 33(1):103–115.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Bordes, A., Chopra, S., and Weston, J. (2014). Question answering with subgraph embeddings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 615–620, Doha, Qatar. Association for Computational Linguistics.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

Chen, L. (2019). Attention-based deep learning system for negation and assertion detection in clinical notes. *International Journal of Artificial Intelligence and Applications (IJAIA)*, 10(1):1–9.

Clausen, D. (2010). HedgeHunter: A system for hedge detection and uncertainty classification. *CoNLL-2010: Shared Task - Fourteenth Conference on Computational Natural Language Learning, Proceedings of the Shared Task*, pages 120–125.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 12(null):2493–2537.
Costumero, R., Lopez, F., Gonzalo-Martín, C., Millan, M., and Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. *Lecture Notes in Computer Science (subseries in Artificial Intelligence)*, 8609 LNAI:366–375.
Cotik, V., Stricker, V., Vivaldi, J., and Rodriguez, H. (2016). Syntactic methods for negation detection in radiology reports in Spanish. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pages 156–165, Berlin, Germany. Association for Computational Linguistics.
Cruz Díaz, N. P. and Maña López, M. J. (2019). *Negation and Speculation Detection*. Editorial Assistant.
Cruz Díaz, N. P., Maña López, M. J., Vázquez, J. M., and Álvarez, V. P. (2012). A machine-learning approach to negation and speculation detection in clinical texts. *Journal of the American Society for Information Science and Technology*, 63(7):1398–1410.
Dalianis, H. (2018). *Clinical Text Mining*. Springer Open.
Dalloux, C., Claveau, V., and Grabar, N. (2019). Speculation and negation detection in French biomedical corpora. *International Conference Recent Advances in Natural Language Processing, RANLP*, pages 223–232.
De Albornoz, J. C., Plaza, L., Diaz, A., and Ballesteros, M. (2012). UCM-I: A rule-based syntactic approach for resolving the scope of negation. **SEM 2012 - 1st Joint Conference on Lexical and Computational Semantics*, 1:282–287.
Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm):4171–4186.
Elazhary, H. (2017). NegMiner: An automated tool for mining negations from electronic narrative medical documents. *International Journal of Intelligent Systems and Applications*, 9(4):14–22.
Fancellu, F., Lopez, A., Webber, B., and He, H. (2017). Detecting negation scope is easy, except when it isn't. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 58–63, Valencia, Spain. Association for Computational Linguistics.
Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010a). The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task*, pages 1–12, Uppsala, Sweden. Association for Computational Linguistics.
Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010b). The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning – Shared Task*, pages 1–12, Uppsala, Sweden. Association for Computational Linguistics.
Goldberg, Y. and Hirst, G. (2017). *Neural Network Methods in Natural Language Processing*. Morgan and Claypool Publishers.
Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., and Poon, H. (2020). Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *arXiv*, 1(1):1–24.
Harkema, H., Dowling, J. N., Thornblade, T., and Chapman, W. W. (2009). ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42(5):839–851.
Hochreiter, S. and Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. *Advances in Neural Information Processing Systems*, pages 473–479.
Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional LSTM-CRF Models for Sequence Tagging.
Jean, P. A., Harispe, S., Ranwez, S., Bellot, P., and Montmain, J. (2016). Uncertainty detection in natural language: A probabilistic model. *ACM International Conference Proceeding Series*, 13-15-June(July).
Jiménez-Zafra, S. M., Morante, R., Blanco, E., Martín Valdivia, M. T., and Ureña López, L. A. (2020). Detecting negation cues and scopes in Spanish. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6902–6911, Marseille, France. European Language Resources Association.
Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Martí, M. A. (2018).

SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation*, 52(2):533–569.
Jiménez-Zafra, S. M., Cruz-Díaz, N. P., Taboada, M., and Martín-Valdivia, M. T. (2021). Negation detection for sentiment analysis: A case study in spanish. *Natural Language Engineering*, 27(2):225–248.
Kesterson, J., Beesley, C., Dexter, P., Schmidt, C. M., and Liu, H. (2015). incorporating dependency relation into NegEx. *Journal of biomedical informatics*, 54:213–219.
Khandelwal, A. and Sawant, S. (2020). NegBERT: A transfer learning approach for negation detection and scope resolution. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, pages 5739–5748.
Konstantinova, N., de Sousa, S. C., Cruz, N. P., Maña, M. J., Taboada, M., and Mitkov, R. (2012). A review corpus annotated for negation, speculation and their scope. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3190–3195, Istanbul, Turkey. European Language Resources Association (ELRA).
Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, Williamstown, MA, USA. Morgan Kaufmann Publishers Inc.
Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.
Lima Lopez, S., Perez, N., Cuadros, M., and Rigau, G. (2020). NUBes: A corpus of negation and uncertainty in Spanish clinical texts. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5772–5781, Marseille, France. European Language Resources Association.
Liu, Q., Kusner, M. J., and Blunsom, P. (2020). A survey on contextual embeddings. *arXiv*.
Liu, R., Shi, Y., Ji, C., and Jia, M. (2019a). A Survey of Sentiment Analysis Based on Transfer Learning. *IEEE Access*, 7:85401–85412.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019b). Roberta: A robustly optimized bert pretraining approach. cite arxiv:1907.11692.
Lopez, M. M. and Kalita, J. (2017). Deep Learning applied to NLP. *ArXiv*.
Marimon, M., Vivaldi, J., and Bel, N. (2017). Annotation of negation in the IULA Spanish clinical record corpus. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 43–52, Valencia, Spain. Association for Computational Linguistics.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.
Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., and Gao, J. (2020). Deep Learning Based Text Classification: A Comprehensive Review. *arXiv*, 1(1):1–43.
Morante, R. and Blanco, E. (2012). SEM 2012 shared task: Resolving the scope and focus of negation. *SEM 2012 - 1st Joint Conference on Lexical and Computational Semantics*, 1:265–274.
Morante, R. and Daelemans, W. (2009a). Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '09, page 28–36, USA. Association for Computational Linguistics.
Morante, R. and Daelemans, W. (2009b). A metalearning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 21–29, Boulder, Colorado. Association for Computational Linguistics.
Morante, R. and Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2):223–260.
Névéol, A., Dalianis, H., Velupillai, S., Savova, G., and Zweigenbaum, P. (2018). Clinical Natural Language Processing in languages other than English: Opportunities and challenges. *Journal of Biomedical Semantics*, 9(1):1–13.
Nikfarjam, A., Emadzadeh, E., and Gonzalez, G. (2014). *Biomedical Informatics Insights*, volume 5. Springer.
Oronoz, M., Gojenola, K., Pérez, A., de Ilarraza, A. D., and Casillas, A. (2015). On the creation of

a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of Biomedical Informatics*, 56:318–332.

Ortiz Suarez, P., Romary, L., and Benoit, S. (2012). A monolingual approach to contextualized word embeddings for mid-resource languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 1703–1714, Online.

Panigrahi, S., Nanda, A., and Swarnkar, T. (2021). A Survey on Transfer Learning. *Smart Innovation, Systems and Technologies*, 194:781–789.

Peng, D. L., Wang, Y. R., Liu, C., and Chen, Z. (2020). TL-NER: A Transfer Learning Model for Chinese Named Entity Recognition. *Information Systems Frontiers*, 22(6):1291–1304.

Peng, Y., Wang, X., Lu, L., Bagheri, M., Summers, R., and Lu, Z. (2018). NegBio: a high-performance tool for negation and uncertainty detection in radiology reports. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, 2017:188–196.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:2227–2237.

Pires, T., Schlinger, E., and Garrette, D. (2019). How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., and Luo, W. (2016). Speculation and negation scope detection via convolutional neural networks. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 815–825.

Santamaria, J. (2019). Negex-mes: Negex para textos médicos en español. https://github.com/PlanTL-SANIDAD/NegEx-MES.

Santiso, S., Casillas, A., Pérez, A., and Oronoz, M. (2018). Word embeddings for negation detection in health records written in Spanish. *Soft Computing*.

Santiso, S., Pérez, A., Casillas, A., and Oronoz, M. (2020). Neural negated entity recognition in Spanish electronic health records. *Journal of Biomedical Informatics*, 105(December 2019):103419.

Shaitarova, A., Furrer, L., and Rinaldi, F. (2020). Cross-lingual transfer-learning approach to negation scope resolution. *CEUR Workshop Proceedings*, 2624(June).

Shaodian, Z., Tian, K., Xingting, Z., Dong, W., Noemie, E., and Jianbo, L. (2016). Speculation detection for chinese clinical notes: Impacts of word segmentation and embedding models. *Journal of Biomedical Informatics*, 60(February):334–341.

Skeppstedt, M., Paradis, C., and Kerren, A. (2016). Marker words for negation and speculation in health records and consumer reviews. *CEUR Workshop Proceedings*, 1650:64–69.

Soares, F., Villegas, M., Gonzalez-Agirre, A., Krallinger, M., and Armengol-Estapé, J. (2019). Medical word embeddings for Spanish: Development and evaluation. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 124–133, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Solarte-Pabón, O., Menasalvas, E., and Rodriguez-González, A. (2020). Spa-neg: An Approach for Negation Detection in Clinical Text Written in Spanish. In Rojas, I., Valenzuela, O., Rojas, F., Herrera, L. J., and Ortuño, F., editors, *Bioinformatics and Biomedical Engineering*, pages 323–337, Cham. Springer International Publishing.

Solarte Pabón, O., Torrente, M., Provencio, M., Rodríguez-Gonzalez, A., and Menasalvas, E. (2021). Integrating Speculation Detection and Deep Learning to Extract Lung Cancer Diagnosis from Clinical Notes. *Applied Sciences*, 11(2).

Stricke, r. V., Ignacio, I., and Cotik, V. (2015). Negated findings detection in radiology reports in spanish: an adaptation of negex to spanish.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 4(January):3104–3112.

Szarvas, G., Vincze, V., Farkas, R., , and Gurevych, I. (2012a). Cross-Genre and Cross-Domain Detection of Semantic Uncertainty. *Computational Linguistics*, 38(2):335–367.

Szarvas, G., Vincze, V., Farkas, R., Móra, G., and Gurevych, I. (2012b). Cross-Genre and Cross-Domain

Detection of Semantic Uncertainty. *Computational Linguistics*, 38(2):335–367.
Taylor, S. J. and Harabagiu, S. M. (2018). The Role of a Deep-Learning Method for Negation Detection in Patient Cohort Identification from Electroencephalography Reports. *AMIA Annu Symp Proc. AMIA Symposium*, 2018:1018–1027.
Uzuner, Ö., South, B. R., Shen, S., and DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.
Uzuner, O., Zhang, X., and Sibanda, T. (2009). Machine Learning and Rule-based Approaches to Assertion Classification. *Journal of the American Medical Informatics Association*, 16(1):109–115.
Velupillai, S., Dalianis, H., and Kvist, M. (2011). Factuality levels of diagnoses in Swedish clinical text. *Studies in Health Technology and Informatics*, 169:559–563.
Velupillai, S., Skeppstedt, M., Kvist, M., Mowery, D., Chapman, B. E., Dalianis, H., and Chapman, W. W. (2014). Cue-based assertion classification for Swedish clinical text-Developing a lexicon for pyConTextSwe. *Artificial Intelligence in Medicine*, 61(3):137–144.
Vincze, V. (2010). Speculation and negation annotation in natural language texts: What the case of bioscope might (not) reveal. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, NeSp-NLP '10, page 28–31, USA. Association for Computational Linguistics.
Vincze, V. (2014). Uncertainty detection in Hungarian texts. *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers*, pages 1844–1853.
Vincze, V., Szarvas, G., Farkas, R., Móra, G., and Csirik, J. (2008). The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(SUPPL. 11):1–9.
Wang, Y., Liu, S., Afzal, N., Rastegar-Mojarad, M., Wang, L., Shen, F., Kingsbury, P., and Liu, H. (2018). A comparison of word embeddings for the biomedical natural language processing. *Journal of Biomedical Informatics*, 87(April):12–20.
Wu, A. S., Do, B. H., Kim, J., and Rubin, D. L. (2011). Evaluation of negation and uncertainty detection and its impact on precision and recall in search. *Journal of Digital Imaging*, 24(2):234–242.
Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv preprint arXiv:1609.08144*, pages 1–23.
Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., and Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32(NeurIPS):1–18.
Zavala, R. R. and Martinez, P. (2020). The impact of pretrained language models on negation and speculation detection in cross-lingual medical text: Comparative study. *JMIR Medical Informatics*, 8(12):1–21.
Zhou, H., Deng, H., Huang, D., and Zhu, M. (2015). Hedge scope detection in biomedical texts: An effective dependency-based method. *PLoS ONE*, 10(7):1–16.
Zhou, H., Ning, S., Yang, Y., Liu, Z., and Xu, J. (2018). Chinese hedge scope detection based on phrase semantic representation. In *2017 International Conference on Asian Language Processing (IALP)*, volume 2018-Janua, pages 285–288.
Zou, B., Zhou, G., and Zhu, Q. (2016). Research on Chinese negation and speculation: corpus annotation and identification. *Frontiers of Computer Science*, 10(6):1039–1051.