Note: Using Causality to Mine Sjögren's Syndrome related Factors from Medical Literature

Pranav Dhananjay Gujarathi
pgujarat@iu.edu
Indiana University,
Bloomington, USA

Sai Krishna Reddy Gopi Reddy
Venkata Mani Babu Karri
Ananth Reddy Bhimireddy
Anushri Singh Rajapuri
sgopired@iu.edu
vkarri@iu.edu
anbhimi@iu.edu
anraja@iu.edu
Indiana University IUPUI
Indianapolis, USA

Manohar Reddy
Mounika Sabbani
Biju Cheriyan
Jack VanSchaik
peddires@iu.edu
msabbani@iu.edu
bijcheri@iu.edu
jtvansch@iu.edu
Indiana University IUPUI
Indianapolis, USA

Thankam P. Thyvalikakath
tpt@iu.edu
Indiana University School of Dentistry
Indianapolis, USA

Sunandan Chakraborty*
sunchak@iu.edu
Indiana University IUPUI
Indianapolis, USA

ABSTRACT

Research articles published in medical journals often present findings from causal experiments. In this paper, we use this intuition to build a model that leverages causal relations expressed in text to unearth factors related to Sjögren's syndrome. Sjögren's syndrome is an auto-immune disease affecting up to 3.1 million Americans. The uncommon nature of the disease, coupled with common symptoms with other autoimmune conditions make the timely diagnosis of this disease very hard. A centralized information system with easy access to common and uncommon factors related to Sjögren's syndrome may alleviate the problem. We use automatically extracted causal relationships from text related to Sjögren's syndrome collected from the medical literature to identify a set of factors, such as "signs and symptoms" and "associated conditions", related to this disease. We show that our approach is capable of retrieving such factors with a high precision and recall values. Comparative experiments show that this approach leads to 25% improvement in retrieval F1-score compared to several state-of-the-art biomedical models, including BioBERT and Gram-CNN.

*Corresponding author

COMPASS '22, June 29-July 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9347-8/22/06…$15.00
https://doi.org/10.1145/3530190.3534850

CCS CONCEPTS

• **Computing methodologies → Information extraction**; *Reinforcement learning*; • **Applied computing → Health informatics**.

KEYWORDS

causal relationships, Sjögren's syndrome, medical NLP

ACM Reference Format:
Pranav Dhananjay Gujarathi, Sai Krishna Reddy Gopi Reddy, Venkata Mani Babu Karri, Ananth Reddy Bhimireddy, Anushri Singh Rajapuri, Manohar Reddy, Mounika Sabbani, Biju Cheriyan, Jack VanSchaik, Thankam P. Thyvalikakath, and Sunandan Chakraborty. 2022. Note: Using Causality to Mine Sjögren's Syndrome related Factors from Medical Literature. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3530190.3534850

1 INTRODUCTION

Sjögren's syndrome is an autoimmune disorder where the immune system destroys glands that produce tears and saliva [38, 48] and is also associated with rheumatic disorders [5, 6, 18]. Most people with Sjögren's syndrome have limited symptoms, such as dry eyes and dry mouth and lack of timely intervention may affect other organs of the body [40]. Due to the common symptoms and symptoms relating to different specialities, such as dentistry, ophthalmology, and rheumatology, and the lack of communication between them, it becomes a challenge for clinicians to timely diagnose Sjögren's syndrome. Several research studies have published new findings about Sjögren's syndrome concerning new *signs and symptoms, risk factors, and associated conditions* [11, 23, 38, 39]. Careful inspection reveals that among them the relationships between "signs and symptoms" and "associated conditions" are often expressed using causal semantics, e.g. Sjögren's syndrome *may cause* a patient to develop an *<associated condition>*, or, Sjögren's syndrome *can cause*

<signs and symptoms>. Such information in research articles may appear as causal sentences. For example, *"sjogrens syndrome can cause not only corneal perforation but also mucosal perforation which may lead to a lacrimal fistula"* [23] - this causal sentence expresses the possibility of two symptoms of Sjögren's syndrome. In this paper, we present a novel method to identify causal sentences and use them to unearth factors related to Sjögren's syndrome, specifically, *associated conditions* and *signs and symptoms*.

Expressing causality through a natural language can take many different forms, using explicit markers (e.g. "causes", "caused by") or implicitly (e.g. Last week temperature rose significantly, there were several cases of heat stroke reported). In the second example, although it is apparent that the rising temperature *caused* the heatstroke cases but no explicit markers were used. Past works on causality extraction from the text have mostly focused on explicit causality [7, 9, 15, 24] We propose a new framework to detect causal sentences and entities using Deep Q Reinforcement Learning (RL) method [35]. Given a sentence or a document, we aim to extract two sets of words and phrases that are connected by a causal semantic (may not be explicit). To extract such words we propose an RL agent, who will iterate over multiple episodes (subsamples of data) and increase the chance of identifying right cause words or causal phrases along with the related effect words or phrases by maximizing a reward. We train and test the model on two separate datasets containing causal sentences, SemEval-2010 [20] and ADE [17], and apply the trained model on a separate corpus of scholarly articles related to Sjögren's syndrome collected from PubMed.

The causal relationship extraction model performed with a F1-score of 0.89 and 0.87 on the SemEval-2010 and ADE datasets respectively. We compared these results with several baseline models, as well as related works that used the same dataset, and found that our model's performance was slightly lower than just one model [30] (F1-score: 90.6) on SemEval-2010 but outperformed the state-of-the-art models trained on the ADE dataset. We observe similar patterns while extracting factors related to Sjögren's syndrome. The precision and recall for our method in extracting Sjögren's syndrome related factors were 0.85 and 0.78 respectively (F1-score: 0.81), which was at least 25% better than other state-of-the-art models targeted towards biomedical text, such as, Gram-CNN and BioBERT.

2 SJOGREN'S SYNDROME

Sjogren's syndrome (SS) is the second most common autoimmune connective tissue disease [48] affecting up to 3.1 million Americans [38]. It affects the salivary and lacrimal glands resulting in dry mouth and dry eyes. SS is common among middle-aged people, with a high prevalence in females [4] [19]. The exact etiology of SS is not known [37, 39], as a result, the diagnosis of SS is delayed [12] and frequently misdiagnosed. Sjögren's syndrome patients (SSP) also experience significantly impaired quality of life due to tooth loss, corneal scarring, fatigue, pain, and depression [45].

The fragmentation of care and suboptimal communication between dentists, physicians including rheumatologists and ophthalmologists is a crucial reason for the poor understanding of SSPs' disease characteristics. The clinical diagnosis of SS is based on a combination of symptoms and objective tests that includes the following: patient-reported symptoms of dry eyes and dry mouth along with objective evidence of ocular tests indicating dry eyes [8], and/or decreased salivary flow rate indicating dry mouth and a positive test for serum anti-Ro antibodies [46] or rheumatoid factor or a labial salivary gland showing lymphocyte infiltrates [25]. While dentists are trained to determine the salivary rate and labial salivary gland biopsy and ophthalmologists are trained to conduct ocular tests for dry eyes, rheumatologists who manage patients with autoimmune conditions may not feel comfortable doing these procedures.

The objective of this paper is to establish a novel entity extraction model that automates the retrieval of clinical findings relevant to SS from the scientific literature. Such a model will support mining relevant information from a large corpus of literature, which is infeasible through a manual process. The factors related to SS that we aim to extract from the scientific literature, such as symptoms, associated conditions, risks are expressed in the literature using some form of *causal* semantics. For example, "dry mouth *is caused by* Sjogren's syndrome". Thus, our goal is to design a causal sentences classifier that identifies sentences with causal semantics, furthermore identify the *cause* and *effect* event pairs from those sentences.

3 CAUSAL RELATIONSHIP EXTRACTION USING REINFORCEMENT LEARNING

We define the problem of identifying causal relationships from natural language text as a sequence labeling task. If an input sentence with n words is represented as $X = x_1, x_2, ..., x_n$, then produce an output sequence of length n, $Y = y_1, ..., y_n$, where $y_i \in \{CAUS, EFF, NONE\}$ – *CAUS* and *EFF* represents words *causing* and *effect* events or factors respectively, and *NONE* represents all other words. Figure 1 shows an example causal sentence with different labels. To identify factors related to Sjögren's syndrome, we analyze sentences where $\mathcal{L}$("Sjögren's syndrome") $\in$ {*CAUS,EFF*}, where $\mathcal{L}(w)$ represents the label of the word w. If $\mathcal{L}$("Sjögren's syndrome") = *CAUS* then $\forall w$, $\mathcal{L}(w)$ = *EFF* will represent factors caused by Sjögren's syndrome and vice versa.

Deep Reinforcement Learning (RL) in recent times has emerged as a promising approach which can utilize popular architectures (e.g. Transformers, CNNs, LSTMs, etc) while also going a step further than function approximation towards general Artificial Intelligence. This is possible due to the way RL tasks are formulated to be basically an optimization strategy, where we simulate an agent playing a finite sequential game to gradually improve the reward obtained at each step. The key difference is, this scalar reward neither needs ground truth labels nor has to be differentiable - as long as the reward magnitudes reflects the agent behaving in a favorable way. We propose a unsupervised framework for the causality extraction from sentences using A2C [1, 29] or Actor Advantage Critic Method.

A typical RL problem consists of the following setup: a sequential task, where an *agent* starts at a initial position(s_0), and has to navigate through different *steps* to eventually reach an end point(s_T), which is referred to as completing an *episode*. At every step, the

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3822229/

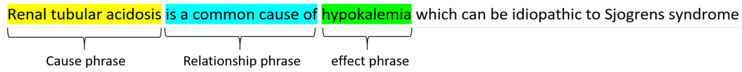


Figure 1: An example sentence [1]with causal relationship relationship that highlights a factor that may lead to Sjögren's syndrome

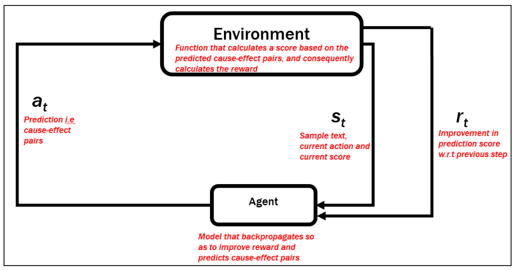


Figure 2: Generalized framework of reinforcement learning with explanation of how is used to extract cause-effect pairs

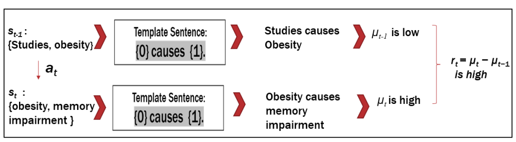


Figure 3: Overview of the scoring method using templates (reward)

agent receives feedback on the decisions taken. Based on the feedback, at time t it tries to take an action(a_t) that will maximize the reward(r_t). Eventually, after multiple simulations of an *episode* and using an optimization algorithm, the objective is to maximize the cumulative *reward* or $\sum_{t=0}^{T} r_t$ for an *episode*. We use this setup and define s_t, r_t and a_t to ensure that maximizing $\sum_{t=0}^{T} r_t$ will improve the prediction accuracy of labeling words in a sentence as *cause* and *effect*. Figure 2 visually depicts this process.

For a particular episode, we pick a random subsample(v) of sentences. At every step, the *agent*(in our case a neural network), takes s_t as an input and predicts a_t, also giving us θ_t. We score this prediction, and assign value μ_t to it. Accordingly, since we want to use previous feedback and results to guide current action, we define the *state*(s_t) as a collection of a time invariant variable (input sentence) as well as two time dependent variables (previous state and scores) incorporating the information of the trajectory after start: $s_t = [v, .a_{t-1}, \mu_{t-1}]$. Since RL algorithms optimize $\sum_{t=0}^{T} r_t$, we define our reward as: $r_t = \mu_t - \mu_{t-1}$. Thus, $\sum_{t=0}^{T} r_t$ is $\mu_t - \mu_0$, meaning optimizing cumulative reward is the same as improving the prediction score compared to random walk (based on our definition). An overview of the scoring method is shown in Figure 3.

We used the actor critique algorithm based on Deep Q-learning Network (DQN) [35] algorithm. This network uses Value function and Q-values at each state to compute the usefulness and quality of the state. At each state s_t consisting of v , a_{t-1} and μ_{t-1} where sentence v stays constant where as a_{t-1} and μ_{t-1} are the feedback terms. The μ_{t-1} is a scalar output and a_{t-1} is a vector of $4\times$ $maxlen(v)$. We fix that the maximum length of a sentence is 80 words for our experimentation and we estimate the probability of every word to be a cause or an effect word. The output vectors for each word will have a size of four and each element will represent the probability of the word to be start ($\phi^s(\kappa)$) and end ($\phi^e(\kappa)$) of the cause phrase or the effect phrase ($\phi^s(\epsilon)$,$\phi^e(\epsilon)$) respectively. At each iteration of the state a sentence of length $len(d_i)$ is passed through Albert [31] a lighter version of BERT [10] based transformer model with 12 million parameters to generate sentence embeddings of size ($len(d_i)$,768). This output is then batch normalized [43] and is reduced by taking a mean across the length l resulting in vector of size (1,768). Then the action $a_t - 1$ output from previous state of size (80, 4) is reduced to (1,128) and batch normalized. This output $a^{|}_{t-1}$ and μ_{t-1} are combined to one single vector of size (1, 896), this output is further reduced and normalized to (1, 128) and combined with the scalar epsilon from previous state ϵ_{t-1}.

4 EVALUATION

We evaluate this work in two phases - (Task 1) evaluate the performance of the causal relationship extraction model, and (Task 2) validate the findings after applying this model to extract factors for Sjögren's syndrome. We use SemEval-2010 Task 8 [20] and Adverse Drug Effects (ADE) [16, 17] for Task 1 and a custom dataset built from the PubMed database for Task 2. We created this dataset from 2,350 PubMed abstracts retrieved using the keyword "Sjögren's syndrome" and its variants. This dataset has 1,058 sentences and the words/phrases were annotated simultaneously by two annotators with a 90.7% agreement (See Appendix B). We refer this dataset as Sjögren's syndrome Dataset (SSD).

4.1 Performance of the Causal Relationship Extraction Model

We evaluated our model on the SemEval and ADE datasets by comparing our findings with the ground truth. We compare our findings with several baseline models, such as Long Term Short Term memory neural network (LSTM) [21], Bidirectional LSTM (BiLSTM) [22], BERT language model [10] with a fully connected network as the final classifier. Table 1 presents the performance across all these models. Our approach outperformed all these baseline models and the F1-score is almost 6% better than the next best (BERT based) model. We also compared our results with the best-performing models from the literature for each dataset – SemEval: Kyriakakis et al. [30], Li et al. [34], Wang et al. [50] and ADE: Gurulingappa et al. [17], Wang and Lu [49], Zhao et al. [53]. Table 2 presents the summary of this comparative analysis and shows the F1 scores in comparison to our approach. Our model outperformed other top models trained on ADE. On the other hand, for SemEval-2010, our model was marginally poorer than Kyriakakis et al. [30].

Table 1: Comparison of our reinforcement learning method with baseline models

	SemEval-2010			ADE		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Our approach (Reinforcement Learning)	**0.93**	**0.86**	**0.89**	**0.88**	**0.85**	**0.86**
LSTM-Glove	0.78	0.82	0.79	0.80	0.77	0.78
BiLSITM-Glove	0.82	0.80	0.81	0.82	0.84	0.82
BERT	0.87	0.83	0.84	0.86	0.831	0.84

Table 2: Comparison with selected related works

Dataset	Model	F1 Score	Dataset	Model	F1 Score
SemEval-2010	Li et al. [34]	84.6	ADE	Gurulingappa et al. [17]	70.0
	Wang et al. [50]	88.0		Wang and Lu [49]	80.1
	Kyriakakis et al. [30]	**90.6**		Zhao et al. [53]	81.1
	Our approach	89.4		**Our approach**	**86.4**

Table 3: Comparative performance

Model	**Precision**	**Recall**	**F1-score**
Bi LSTM	0.45	0.84	0.59
Glove Embedings + CNN	0.47	0.72	0.56
Bi LSTM + CRF	0.05	0.4	0.1
BioWordVec + CNN [26, 52]	0.48	0.74	0.58
BioBERT [33]	0.39	0.55	0.46
Gram-CNN [55]	0.52	0.74	0.61
Our approach	**0.85**	**0.78**	**0.81**

4.2 Identification of Factors related to Sjögren's syndrome

We apply the causal relationship extraction model tested on SemEval-2010 and ADE datasets on the Sjögren's syndrome dataset (SSD) to identify causal sentences and the corresponding cause and effect phrases to extract *signs and symptoms* and *associated conditions* related to Sjögren's syndrome. We present a set of selected causal-effect pairs extracted through our model in Table 4 in Appendix A. In these examples, we see that "Sjögren's syndrome " and the factors, such as signs and symptoms (e.g. "loss of secretion", "xerophthalmia") and associated conditions (e.g. "annular erythema", "non-Hodgkin's lymphoma") can appear as a cause as well as an effect. To verify the above claim, we applied our model to the labeled dataset where 1,058 sentences were annotated (See Appendix B for details about the annotation process). We created a test set containing 100 sentences out of the 383 causal sentences found in this dataset. We collected the *cause* (or effect) associated with the term "Sjögren's syndrome" when it is the *effect* (or cause) and computed the retrieval accuracy of those two labels. The model performed with a precision of 0.87 and recall of 0.71 (Table 3). We compared our findings with several baseline models designed for sequence labeling. All these were supervised models and trained on a set of 283 annotated sentences and tested on the same test set. Baseline models included BiLSTM and a modified BiLSTM with an additional CRF model similar to Li et al. [34], biomedical models, including BioBERT [33], BioWordVec [52], and gram-CNN [55] The results from these experiments are summarized in Table 3.

The results (Table 3) show the central hypothesis of this work that causal relations can be used to extract certain factors associated with Sjögren's syndrome holds. Retrieval performance is better than the baseline methods but on many occasions, *associated factors* or *signs and symptoms* are present in a sentence without any causal semantics. To achieve the long-term goals and improve the recall of the model, it is important to identify other relations that bind these factors with the disease. For example, the sentence "Two years after presentation the patient developed **dyspnea cough and xerostomia**" contains *symptoms* but due to the absence of a causal semantic, our present model will add this to the list of false negatives. Moreover, detecting and using other relations in the future will also help to extract the other two labels – "diagnostic tests" and "risk factors". As these two labels do not associate with the disease as a causality, we need to investigate the relations that will help to discover those factors. We will keep these tasks as part of the future directions of this work.

5 RELATED WORK

Causality is an important problem and has been addressed across many domains [42]. As an extension, mining causal relationships from text has been extensively explored. A large body of such work focused on domain-specific texts[32, 47, 51, 54] or have used rule-based models [2, 15, 27]. Meuller et al. [36] presented a novel method and a working prototype that automatically extracts not only causes and effects but also signs, mediators, and conditions from scientific papers. Kim et al. [28] proposed a method to extract technological data from patents, to identify technological causes and effect relations. Egami et al. [13] provided a conceptual framework for text-based causal inferences. CausalTriad [54] is a minimally supervised approach, based on focused distributional similarity methods and discourse connectives, for identifying causality. Other approaches focused on extracting causal relationships from text, exploited linguistic structures, such as, multi-word expressions [44], N-grams, topics and sentiments [27], lexical patterns [2, 15]. Paul et al. [41] used causal inference to find causal relationships between word features and document labels for better feature engineering. Although, there have been many different directions in identifying causal statements from text, however, most of the methods are directed at domain-specific text or statements that are express causality explicitly.

6 CONCLUSION

This paper presents an innovative approach to extract factors related to Sjögren's syndrome from medical journal articles. We present a novel reinforcement learning-based method to identify causal relations from text and show that it outperforms most other similar models. We apply this model on a dataset of 383 sentences extracted from a larger set of 2,530 abstracts taken from articles on Sjögren's syndrome. Using causal relationships, we aimed to extract two labels, "signs and symptoms" and "associated conditions" and show that our retrieval method has better performance compared to several supervised baseline models.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the National Science Foundation under Grant No. 1948322.

REFERENCES

[1] Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2016. An actor-critic algorithm for sequence prediction. *arXiv preprint arXiv:1607.07086* (2016).

[2] Danushka Bollegala, Simon Maskell, Richard Sloane, Joanna Hajne, and Munir Pirmohamed. 2018. Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach. *JMIR public health and surveillance* 4, 2 (2018).

[3] SJ Bowman, GH Ibrahim, G Holmes, John Hamburger, and JR Ainsworth. 2004. Estimating the prevalence among Caucasian women of primary Sjögren's syndrome in two general practices in Birmingham, UK. *Scandinavian journal of rheumatology* 33, 1 (2004), 39–43.

[4] SJ Bowman, GH Ibrahim, G Holmes, J Hamburger, and JR Ainsworth. 2004. Estimating the prevalence among Caucasian women of primary Sjögren's syndrome in two general practices in Birmingham, UK. *Scandinavian Journal of Rheumatology* 33, 1 (2004), 39–43. https://doi.org/10.1080/03009740310004676 arXiv:https://doi.org/10.1080/03009740310004676

[5] Lindsay E. Brown, Michelle L. Frits, Christine K. Iannaccone, Michael E. Weinblatt, Nancy A. Shadick, and Katherine P. Liao. 2014. Clinical characteristics of RA patients with secondary SS and association with joint damage. *Rheumatology* 54, 5 (10 2014), 816–820. https://doi.org/10.1093/rheumatology/keu400 arXiv:https://academic.oup.com/rheumatology/article-pdf/54/5/816/6699339/keu400.pdf

[6] Lindsay E Brown, Michelle L Frits, Christine K Iannaccone, Michael E Weinblatt, Nancy A Shadick, and Katherine P Liao. 2015. Clinical characteristics of RA patients with secondary SS and association with joint damage. *Rheumatology* 54, 5 (2015), 816–820.

[7] Quoc-Chinh Bui, Breanndán Ó Nualláin, Charles A Boucher, and Peter MA Sloot. 2010. Extracting causal relations on HIV drug resistance from literature. *BMC bioinformatics* 11, 1 (2010), 101.

[8] Xiaoyun Chen, Huaxun Wu, and Wei Wei. 2018. Advances in the diagnosis and treatment of Sjogren's syndrome. *Clinical rheumatology* 37, 7 (2018), 1743–1749.

[9] Tirthankar Dasgupta, Rupsa Saha, Lipika Dey, and Abir Naskar. 2018. Automatic Extraction of Causal Relations from Text using Linguistically Informed Deep Neural Networks. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 306–316.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]

[11] Lingli Dong, Yu Chen, Yasufumi Masaki, Toshiro Okazaki, and Hisanori Umehara. 2013. Possible mechanisms of lymphoma development in Sjogren's syndrome. *Current immunology reviews* 9, 1 (2013), 13–22.

[12] Linda Douglas. 2018. Facilitating timely diagnosis of Sjögren's syndrome. *BDJ Team* 5, 2 (2018), 18026.

[13] Naoki Egami, Christian J Fong, Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. 2018. How to make causal inferences using texts. *arXiv preprint arXiv:1802.02163* (2018).

[14] George E Fragoulis, Sofia Fragkioudaki, James H Reilly, Shauna C Kerr, Iain B McInnes, and Haralampos M Moutsopoulos. 2016. Analysis of the cell populations composing the mononuclear cell infiltrates in the labial minor salivary glands from patients with rheumatoid arthritis and sicca syndrome. *Journal of Autoimmunity* 73 (2016), 85–91.

[15] Roxana Girju and Dan I. Moldovan. 2002. Text Mining for Causal Relations. In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*. AAAI Press, 360–364. http://dl.acm.org/citation.cfm?id=646815.708596

[16] Harsha Gurulingappa, Abdul Mateen-Rajpu, and Luca Toldo. 2012. Extraction of potential adverse drug events from medical case reports. *Journal of biomedical semantics* 3, 1 (2012), 15.

[17] Harsha Gurulingappa, Abdul Mateen Rajput, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. 2012. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of biomedical informatics* 45, 5 (2012), 885–892.

[18] Charles G Helmick, David T Felson, Reva C Lawrence, Sherine Gabriel, Rosemarie Hirsch, C Kent Kwoh, Matthew H Liang, Hilal Maradit Kremers, Maureen D Mayes, Peter A Merkel, et al. 2008. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. *Arthritis & Rheumatism* 58, 1 (2008), 15–25.

[19] Charles G. Helmick, David T. Felson, Reva C. Lawrence, Sherine Gabriel, Rosemarie Hirsch, C. Kent Kwoh, Matthew H. Liang, Hilal Maradit Kremers, Maureen D. Mayes, Peter A. Merkel, Stanley R. Pillemer, John D. Reveille, John H. Stone, and National Arthritis Data Workgroup. 2008. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. *Arthritis & Rheumatism* 58, 1 (2008), 15–25. https://doi.org/10.1002/art.23177 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/art.23177

[20] Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, Uppsala, Sweden, 33–38. https://www.aclweb.org/anthology/S10-1006

[21] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[22] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv:1508.01991* (2015).

[23] Sho Ishikawa, Takuhei Shoji, Yuri Nishiyama, and Kei Shinoda. 2019. A case with acquired lacrimal fistula due to Sjögren's syndrome. *American Journal of Ophthalmology Case Reports* 15 (2019), 100526.

[24] Ashwin Ittoo and Gosse Bouma. 2011. Extracting explicit and implicit causal relations from sparse, domain-specific texts. In *International Conference on Application of Natural Language to Information Systems*. Springer, 52–63.

[25] Roland Jonsson, Karl A Brokstad, Malin V Jonsson, Nicolas Delaleu, and Kathrine Skarstein. 2018. Current concepts on Sjögren's syndrome–classification criteria and biomarkers. *European journal of oral sciences* 126 (2018), 37–48.

[26] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. 2016. Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* (2016).

[27] Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, and Eduard Hovy. 2017. Detecting and explaining causes from text for a time series event. *arXiv preprint arXiv:1707.08852* (2017).

[28] Hongbin Kim, Junegak Joung, and Kwangsoo Kim. 2018. Semi-automatic extraction of technological causality from patents. *Computers and Industrial Engineering* 115 (2018), 532 – 542. https://doi.org/10.1016/j.cie.2017.12.004

[29] Vijay Konda and John Tsitsiklis. 1999. Actor-critic algorithms. *Advances in neural information processing systems* 12 (1999).

[30] Manolis Kyriakakis, Ion Androutsopoulos, Artur Saudabayev, and Joan Ginés i Ametllé. 2019. Transfer Learning for Causal Sentence Detection. In *Proceedings of the 18th BioNLP Workshop and Shared Task*. Association for Computational Linguistics, Florence, Italy, 292–297. https://doi.org/10.18653/v1/W19-5031

[31] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).

[32] Dong-gi Lee and Hyunjung Shin. 2017. Disease causality extraction based on lexical semantics and document-clause frequency from biomedical literature. *BMC medical informatics and decision making* 17, 1 (2017), 53.

[33] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240.

[34] Zhaoning Li, Qi Li, Xiaotian Zou, and Jiangtao Ren. 2019. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings. *Neurocomputing* 423 (2019), 207–219.

[35] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[36] Roland Mueller and Sebastian Hüttemann. 2018. Extracting Causal Claims from Information Systems Papers with Natural Language Processing for Theory Ontology Learning *(Hawaii International Conference on System Sciences)*.

[37] M. M. Muñoz, J. Sebastián, R. Roda, Y. J. Soriano, and María Gracia Sarrión Pérez. 2009. Sjögren's syndrome of the oral cavity. Review and update.

[38] Cuong Q. Nguyen and Ammon B. Peck. 2009. Unraveling the Pathophysiology of Sjogren Syndrome-Associated Dry Eye Disease. *The Ocular Surface* 7, 1 (2009), 11–27. https://doi.org/10.1016/S1542-0124(12)70289-6

[39] Gaëtane Nocturne and Xavier Mariette. 2013. Advances in understanding the pathogenesis of primary Sjögren's syndrome. *Nature Reviews Rheumatology* 9, 9 (2013), 544–556.

[40] Ruchika Patel and Anupama Shahane. 2014. The epidemiology of Sjögren's syndrome. *Clinical epidemiology* 6 (2014), 247.

[41] Michael J Paul. 2017. Feature Selection as Causal Inference: Experiments with Text Classification. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*. 163–172.

[42] Judea Pearl. 2009. Causal inference in statistics: An overview. *Statist. Surv.* 3 (2009), 96–146. https://doi.org/10.1214/09-SS057

[43] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. 2018. How does batch normalization help optimization? *arXiv preprint arXiv:1805.11604* (2018).

[44] Shota Sasaki, Sho Takase, Naoya Inoue, Naoaki Okazaki, and Kentaro Inui. 2017. Handling multiword expressions in causality estimation. In *IWCS 2017—12th International Conference on Computational Semantics—Short papers*.

[45] Barbara Segal, Simon J Bowman, Philip C Fox, Frederick B Vivino, Nandita Murukutla, Jeff Brodscholl, Sarika Ogale, and Lachy McLean. 2009. Primary Sjögren's Syndrome: health experiences and predictors of health quality among patients in the United States. *Health and quality of life outcomes* 7, 1 (2009), 1–9.

[46] Ana-Luisa Stefanski, Christian Tomiak, Uwe Pleyer, Thomas Dietrich, Gerd Rüdiger Burmester, and Thomas Dörner. 2017. The diagnosis and treatment of Sjögren's syndrome. *Deutsches Ärzteblatt International* 114, 20 (2017), 354.

[47] Santosh Tirunagari. 2015. Data Mining of Causal Relations from Text: Analysing Maritime Accident Investigation Reports. *arXiv preprint arXiv:1507.02447* (2015).

[48] Frederick B. Vivino. 2017. Sjogren's syndrome: Clinical aspects. *Clinical Immunology* 182 (2017), 48–54. https://doi.org/10.1016/j.clim.2017.04.005 Special issue: Sjogren's Syndrome.

[49] Jue Wang and Wei Lu. 2020. Two are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 1706–1721. https://doi.org/10.18653/v1/2020.emnlp-main.133

[50] Linlin Wang, Zhu Cao, Gerard De Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention cnns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1298–1307.

[51] Yujia Zhai, Shaojing Sun, Fang Wang, and Ying Ding. 2017. Multiplicity and uncertainty: Media coverage of autism causation. *Journal of Informetrics* 11, 3 (2017), 873 – 887. https://doi.org/10.1016/j.joi.2017.07.005

[52] Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. 2019. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific data* 6, 1 (2019), 1–9.

[53] Shan Zhao, Minghao Hu, Zhiping Cai, and Fang Liu. 2020. Modeling Dense Cross-Modal Interactions for Joint Entity-Relation Extraction. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, Christian Bessiere (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4032–4038. https://doi.org/10.24963/ijcai.2020/558 Main track.

[54] Sendong Zhao, Meng Jiang, Ming Liu, Bing Qin, and Ting Liu. 2018. CausalTriad: Toward Pseudo Causal Relation Discovery and Hypotheses Generation from Medical Text Data. (2018).

[55] Qile Zhu, Xiaolin Li, Ana Conesa, and Cécile Pereira. 2018. GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics* 34, 9 (2018), 1547–1554.

APPENDIX

A EXAMPLES OF FACTORS RELATED TO SJÖGREN'S SYNDROME USING CAUSAL RELATIONSHIPS

Table 4: Selected examples of extracting factors by mining causal relationships

	Sentence	Cause	Effect
1	Hypokalemic paralysis is a rare presentation of Fanconi syndrome (FS) caused by sjogrens Syndrome.	sjogrens Syndrome	Hypokalemic paralysis
2	Primary sjogrens syndrome (pSS) is a chronic systemic autoimmune disease that leads to sicca symptoms, mainly xerophtalmia and xerostomia.	Primary sjogrens syndrome	sicca symptoms, mainly xerophtalmia and xerostomia
3	sjogrens syndrome (SjS) is an autoimmune condition that primarily affects salivary and lacrimal glands, causing loss of secretion.	sjogrens syndrome	loss of secretion
4	71-year old woman in whom diagnosis of possible causes of the development of annular erythema, led the team to identify primary sjogrens syndrome (SS).	development of annular erythema	primary sjogrens syndrome
5	Primary sjogrens syndrome (pSS) is characterized by lymphocytic infiltration of the exocrine glands resulting in decreased saliva and tear production.	Primary Sjogrens Syndrome	decreased saliva and tear production
6	Development of non-Hodgkin's lymphoma (NHL) is the major adverse outcome of sjogrens syndrome affecting both morbidity and mortality.	sjogrens syndrome	non-Hodgkin's lymphoma

B DATA ANNOTATION

B.1 Data Extraction and Preprocessing

We collected around 2,530 abstracts with 25,525 sentences. These abstracts were extracted from the PubMed database using keywords "Sjogren's Syndrome", "Sjogren" from 2016 to December 2020. Duplicates were removed, and the abstracts were downloaded. The downloaded data had further additional information such as PMID, Title, Authors, Citation, NIHMS ID, DOI, and abstract text. The abstract text was further cleaned to ASCII text to remove all non-Latin words and letters, and the resulting abstract text was saved to an excel sheet for further usage. Each sentence of the abstract was further broken down and converted into individual text files for annotations. We selected a set of 1,058 sentences for annotation and to be used in all the experiments.

B.2 Annotations Guidelines and References Standards

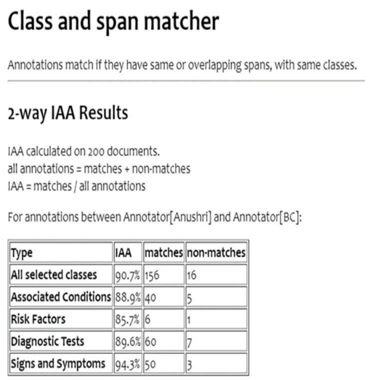
We created annotation guidelines for manually annotating Sjogren's Syndrome information that typically dentists seek for their diagnosis of the disease during patient care. We created these guidelines based on the existing literature in dentistry and medicine [3, 6, 14, 18]. Sjogren's related information address by our annotation schema included concepts of Signs and Symptoms, Associated Conditions, Diagnostic Tests, and Risk Factors. Two annotators (A and B) participated in this task and both have advanced knowledge and prior experience with Sjögren's syndrome. We chose the extensible Human Oracle Suite of Tools (eHOST) for this annotation task. Table 5 summarizes the label and corresponding examples.

B.3 Annotation Task

Practice Phase: For this phase, annotators A and B first selected a set of 100 sentences then 501 and lastly 200 from the given dataset and independently annotated them based on the minimal guidelines created. After every set Inter-Annotator Agreements (IAA) were calculated and disagreements between the annotators were resolved through discussion and consensus, and the guidelines were updated subsequently. After this phase concluded, the first author analyzed each annotation set to identify annotation patterns. This cycle continued till a good score of IAA was achieved thus representing an excellent agreement between the two researchers. The analysis results were then discussed among the annotators and served to refine the guidelines.
Adjudication phase: Finally, the final set of annotations were adjudicated and overseen by the annotator C. To create the gold standard to be used on the remaining 2000 annotations. During this phase, annotator C was free and discussed the annotations with the actual annotator to understand his/her reasoning.

Table 5: Labels with examples

Sjogren's Syndrome Concepts	Examples of the literal text match from the sentences.
Signs and Symptoms	"xerostomia", "xeropthalmia", "hyposalivation", "dry eyes", "dry mouth", "joint pain", vasculitis
Associated Conditions	"Rheumatoid arthritis"," Systemic lupus Erythematosus,", "Squamous cell carcinoma," "Hodgkin's lymphoma"
Diagnostic Tests	"Schirmer Test", "Rose Bengal Test", "Abnormal Flow rate", "Scintigram",
Risk Factors	"Women", "Postmenopausal", "Mean age 40" "Rheumatic Disease"



Class and span matcher

Annotations match if they have same or overlapping spans, with same classes.

2-way IAA Results

IAA calculated on 200 documents.
all annotations = matches + non-matches
IAA = matches / all annotations

For annotations between Annotator[Anushri] and Annotator[BC]:

Type	IAA	matches	non-matches
All selected classes	90.7%	156	16
Associated Conditions	88.9%	40	5
Risk Factors	85.7%	6	1
Diagnostic Tests	89.6%	60	7
Signs and Symptoms	94.3%	50	3

Figure 4: Screenshot of eHOST tool summarizing the inter-annotator performance and agreement

Results: After the first set of 100 and 501 sentences, the IAA score was a fair 48.4% and 53.5% with a moderate increase of 5.5%. In discussing the disagreements, the annotators' existing domain knowledge and inference were playing a key role in identifying the concepts. Therefore, for the next set of 200 sentences, a strict ground rule was set, as "The annotations should be text-bound. The annotators domain knowledge and interpretation should play a minimal role in annotation and the annotator should be only concerned with what is explicitly stated in the text. The annotators should also provide basis and justify the annotation and its concept". Following this and the updated guidelines IAA was recorded to be 90.7% (Figure 4.