A Systematic Study on Extraction of Temporal Relation from Clinical Free Text

Vani K.B.
Research Scholar
VTU-RC
AJIET, Mangalore
INDIA

Lokesh M.R.
Professor
Dept of ISE
AJIET, Mangalore
INDIA

ABSTRACT

Clinical Assessment and decision making mainly depends on Temporal Relations that exists between clinical event and activities of treatment prescribed. Temporal Relation extraction is a challenging task due to complexities associated with natural language processing techniques, representational ways for temporal data related to clinical activities, methodical approaches followed to extract temporal relations and temporal reasoning. In this work, we propose review of temporal relation extraction in clinical text. We analyzed around 118 articles via an exhaustive search of semanticscholar.org, PubMed, DBLP computer science Bibliography between 2018 to 2023. Relevant studies were made concerned to data sets and methodical approaches incorporated to extract temporal information. A thorough examination of selected papers was made to collect information on TLINK types, data sources, features selection methods used, DocTimeRel, Candidate pair generations and reported performance. Most state of the art is based on attention-based models, with contextualized word representations being fine-tuned for temporal relation extraction. Performance of Tlink extraction is dependent parameter of underlying mechanisms involved in temporal expression identification, temporal events recognitions and mechanisms used to extract temporal relations. F-score for identifying the temporal relation is observed to be in the range of 80% to 91.1%. Most works frequently used TLINKS are ‘before’, ‘after’, ‘overlap’ and ‘contains’ leaving a scope to extend the use of other TLINKS such as ‘started by’, ‘finished by’ ’precedes’ and so on. Machine learning based models and Deep learning-based models were the most commonly adopted techniques for extraction of temporal relations. Dataset Imbalance because of candidate pair generation and task complexity affects system’s performance leaving a scope for research. Most publications worked so far resides on same datasets, which shows a need for design of experiments on new kind of annotations.

Keywords

TLINKS, temporal relation, extraction, complexity, candidate pair generation, Doctimerel, HER (Electronic Health Record)

1.INTRODUCTION

Health care data can be routinely collected and recorded as Electronic Health Records(EHRs) to facilitate clinical and research activities. Such health care data collected can be in different forms of structured, semi- and unstructured data. Structured EHRs normally comprises coded and numerical time stamped information such as clinical visits, diagnoses, biomarkers, results, treatments and so on. Unstructured EHRs normally comprises clinical free text or semi-structured narratives data, clinical imaging etc. Clinical Free text such as clinical notes, letters, reports and observations and so on can be to record patient’s medical history, duration of symptoms, patients’ condition, rejected and diagnosed hypothesis, treatment experience. Over the past times, research activities have shown that key clinical information can be extracted form free text or narratives. Clinical events embedded within clinical text can be organized in the temporal context to understand the time line order of clinical procedures there byfacilitating enhanced diagnostics and treatments. Natural Language tools such as Clinical NLP can be used extract rich and contextual information to obtain current status of a patient’s past, temporal relations and also provide information about patient’s future.Temporal Relation Extraction provides the chronological order among mention over texts, representing clinical events or Temporal Expressions. Clinical event can be discharge summery, treatment, reports and so on. A temporal Expression can be a time mentioned in free text or document creation Time(DCT).

Although, many research works have contributed methodologies to improve the performance of temporal relation extraction, there exist a potential scope to advance further. This works mainly aims to conduct a systematic study on clinical text mining to survey existing methods for extracting temporal relations form clinical free text in English language there by establishing state of the art in this field. However, difficulties with Annotation and extraction of temporal relation is a complicated process due to lower inter-annotator agreement(IAA) than other clinical annotation tasks such as vent and temporal expression annotation tasks.Annotating clinical data may require specific medical expertise which can be quite expensive. In addition to that clinical text exhibits specific characteristics which directly impact the text pre-processing steps and extraction results.The interest in temporal relation extraction from clinical narratives began to grow i2b2 challenge and clinical TempEval in SemEval2015 [16]. With a focus into approaches used, main aspects and choosing of best method in studies, we have performed a systematic review which follows PRISMA statement [19].Despite previous review works on temporal relation extraction in clinical texts, there is still a scope for possibly on some areas. The authors of [20] highlighted some preliminary studies between 2006 and 2012, while the authors of [21] presented studies between 2006 and 2018. Owing to recent discoveries, the state-of-the-art changed over these two years, which was not covered by the authors of [21].

Authors of [22] have discussed state of the art of temporal relation extraction covering both TLINK and DocTimeRel types. However, there is still a scope to address performance of Temporal relation extraction methodologies with enhanced candidate pair generation based on ontologies and semantics. In this regard we performed a systematic review that follows work of [22] with emphasis given to temporal extraction methodologies, state of the art, frameworks, tools used and enhanced candidate pair generation methodologies. Additionally, using our publication criteria, we analyze most extensive set of articles that contemplated by [22], covering the research topic evolution over recent years.

Objective of this study is to present state-of the art temporal relation extraction. Reader of this work can derive answers for questions on "effectiveness of machine learning and rule-based techniques in identifying temporal relations in clinical texts". Our secondary objective is to provide insights into the domain evolution over time by leveraging temporal relation extraction objectives and developing frameworks. A reader of this review can expect an analysis of temporal relations and investigate the best performing techniques and frameworks for temporal relation extraction. A reader of this review can expect an analysis of temporal relations and investigate the best performing techniques and frameworks for temporal relation extraction. Finally, recent development in candidate pair generation methodologies and an acceptable influence of that temporal relation extraction are thoroughly surveyed and addressing this gap to enhanced performance.

The remainder of this article is structured as follows. Section 2 provides an overview of temporal relation extraction in detail. Section 3 provides analyisis od TLINK extraction strategies Datasets and Tools used. Section 4 provides extraction of DocTimeRel relations. Section 5 provides Extraction of Tlink Relations. Section 6 provides concusion on candidate pair generation to enhance performance of temporal relation extraction.

2.TEMPORAL RELATION EXTRACTION

Temporal relation extraction can be summarized in two steps: (i) identifying a relation between pairs of mentions (e.g., event and temporal expressions) and (ii) classifying this relation into a temporal relation type among a predefined set. In Section 2.1, we explain temporal relation representations and discuss the differences between temporal relation sets. In Section 2.2, we explain the event and temporal expression characteristics in both clinical and general domains.

2.1 Temporal Relation Representation

Allen's Interval–based algebra is the basis for Time-ML temporal Mark-up language developed exclusively to annotate even, temporal expressions and relations in the text [22]. The TLINK tag represents a temporal relationship between events and temporal expressions. The TimeML relations are displayed in Table 1 (TimeML column). In THYME-TimeML temporal history of medical events are annotated. Temporal expression definitions are similar to TimeML with addition ofnew category for preoperative, intraoperative and post-operative mentions. THYME-TimeML created DocTimeRel Category and a narrative container concept. The DocTimeRel relations are considered an event attribute and have the following relation set: BEFORE, AFTER, OVERLAP, BEFORE/OVERLAP, and AFTER. The THYME-TimeML DocTimeRel relations are displayed in Table 1 (THYME-TimeML DocTimeRel column). BEFORE/OVERLAP indicates that the event occurred in the past and still occurs in the DCT. The narrative container concept is used to annotate the TLINKs. The THYME-TimeML TLINK relation set are BEFORE, OVERLAPS, BEGINS_ON, ENDS_ON, and CONTAINS. The THYME-TimeML TLINK relations are displayed in Table 1 (THYME-TimeML TLINKs column). Several events or temporal expressions can be connected to the same anchor, which contains them (represented in the CONTAINS row in Table 1). Events and temporal expressions in the same narrative container can be related, as a single element, with other containers [28]. The most significant advantage is a reduction in the number of required annotations [28]. The narrative container strategy is suitable in the clinical domain because there are central mentions of the texts, such as temporal expressions of date and time types, or more comprehensive events, such as mentions of exams.Different annotation schemes will have a temporal relation set based on the annotation requirements. For instance, the temporal relation OVERLAP is generic, implying that the two mentions somehow overlap. However, specific relations such as IDENTITY and SIMULTANEOUS indicate a particular OVERLAP case in which both events coincide, having the same start and endpoints. There is a trade-off between the amount of information represented by a relation set and the task complexity in both the annotation and extraction steps.To distinguish between close temporal relation types, additional information or specific knowledge may be required. In ClinicalTempEval only 'CONTAINS 'relation Type was used [22]. In i2b2 2012 shared task for annotation process 'BEFORE, BEFORE/OVERLAP, OVERLAPS, DURING, ENDS_BY, AFTER, BEGINS_BY, and SIMULTANEOUS relation types are used for annotation [22].Thus, always an extended relation set is ideal. However, trade-off lies between temporal information and task complexity.

2.2 Temporal Relation Extraction processes

In this section we describe the process of temporal relation extraction process. There has been an increasing interest among researchers to extract TIMEXs, Events and TLINKS form clinical free text [16-22]. Given a clinical text known to contain time expression, DCT and clinical events, to extract temporal relation both of TLINK and DocTimeRel it needs to undergo certain stages viz as follows and as shown in the figure 1. Fig 1 describes the different stages involved in the process of temporal relation extraction.

2.2.1 Pre-processing

This stage enriches clinical text with lexical, syntactic and semantic information and coverts them in to representation form required for subsequent stages in sown stream steps. This stages typically comprises tokenization, sentence splitting, part of speech tagging etc. A range of existing have been used for pre-processing in the selected studies [20] Most frequently used tool for pre processing is cTAKES, Stanford's cornel and MetaMAP.

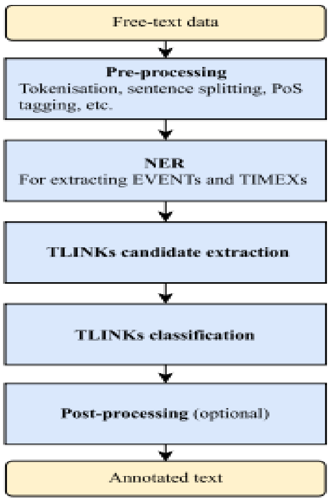

Fig 1: Process of Temporal Extraction.

2.2.2 NER (Extracting Events and Timex's)

This stage involves extracting EVENTs and Timex's form unstructured clinical text using machine learning classifiers. In the study conducted most works used Support Vector machine, Conditional Random Fields(CRFs) and rule based methods. Deep learning is also a current probable choice for extracting EVENTs and TIMEXs [28,57].

2.2.3 TLINK and DocTimeRel candidate extraction

In this stage, TLINK candidates from possible pairs of EVENTS and TIMEXs that should be linked through a temporal relation. In the research review conducted it is found that strategies commonly used relies upon on the following [20,22,73]: a) for TLINKs between Event and Document Creation Time (DCT) with inclusion of all pairs. b) TLINKs within one sentence c) TLINKs across sentences and EVENTS in consecutive sentences. However, there is no significant works from recent times about ontologies and sematic based candidate pair generation and also methodologies to enhance performance of temporal relation extraction via enhanced lexicographic features-oriented candidate pair generation. There is tremendous scope of research in extended candidate pair generation.

2.2.4 TLINK classification

In this stage heterogeneous TLINKS types will be assigned to pair of entities. research review conducted outlines 3 categories of techniques viz as follows a) Rule based Methods b) Machine Learning Algorithm based classification c) Hybrid Approaches [20,22]. Figure 3 shows plot of distribution of classification approaches between 2017 to 2023.

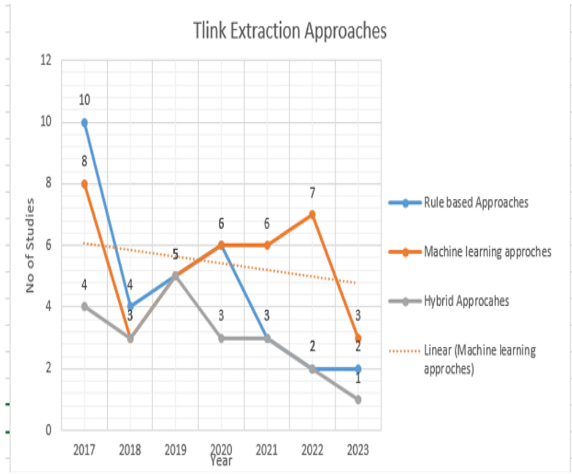

Fig 2. Plot of studies against TLINK Extraction approaches

Table 1. Features used in machine learning based methods

Features used in machine learning-based methods.

Feature	Description	Frequency
Syntactic features	Features relate to the structure of sentences, phrases and words in the text.	41
Entity attributes	The attributes that is extracted from given corpora, e.g., modality, polarity, normalized TIMEXs, etc.	36
Lexical features	This includes n-grams of characters in a word, n-grams of words in a context window and n-grams of lemma, stem, PoS, etc.	35
Positional features	The distance between related words and the surrounding context	24
Dependency parsing	Features from dependency parser	23
Semantic features	Features relate to the meaning in the language and the relationships among the words.	20
Word representation	Word representation features based on Brown clustering, word2vec, etc.	16
Dictionary feature	Whether an n-gram matches with EVENT in dictionaries. This also includes features extracted from external sources (i.e., NLP tools).	15
Orthographic feature	The structure of sentences, phrases and words in text, e.g., starting with a lowercase letter, containing only alphanumeric characters, containing a symbols, etc.	11
Rules	The use of rules as features	9

Features used in machine learning approaches in the research review are tabulated in the table 1. With inspection of few studies Shared Task events such as i2b2 and TempEval happened in previous years are predominately follows NLP approaches as shown figure 3

Method on i2b2 THYME	2017	2018	2019	2020	2021	2022	2023
Rule-based	✓	✓			✓		
Naïve Bayes							
Support Vector Machine (SVM)		✓	✓	✓	✓	✓	
Conditional random fields (CRFs)		✓	✓	✓		✓	
Maximum Entropy Markov Logic Network Decision tree	✓			✓			
Logistic Regression K-nearest neighbours (KNN) Neural Networks (NNs)	✓	✓					
Convolutional NNs (CNNs)	✓	✓				✓	

Fig 3. Shared task events.

2.2.4 Post Pre-processing

Studies on Post Pre-processing focused on enhancing the performance of their methods or to deal with conflicting Hybrid Approaches [20,22,31,34,39,59,63,68,69,73]. In this stage, if any conflicts arise for classification due sentences that may contradicts relation, such relations are removed off by hand-rule.

2.2.5 Methodology

SemanticScholar.org, PubMed, DibLIb databases were selected for this review. The inclusion and exclusion criteria for the title and abstract analysis and the full-text analysis are provided in the Table 5. The search criteria "temporal relation extraction in clinical text". Chronological study on the same topic are shown in figure 4.

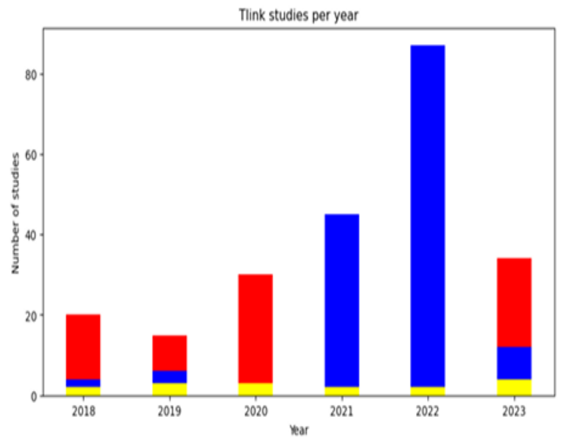

Fig 4. Plot of number of studies on TLINK Extraction approaches.

All articles published up to 2023 since 2012 are reviewed. A Rul total of 3208 studies are retained after eliminating duplication. Around 2908 article were excluded from the study after making full text analysis with retention of 214 journals. By a careful inspection of each work out of 214, 96 journals were dropped off from the review process. Thus around 118 articles or journals were finally considered for the review study. The inclusion and exclusion of any journal to this study is based on criteria as shown Table 4 below.

Table 2. Inclusive and exclusive criteria for TLINK extraction

Criteria	Inclusion criteria	Exclusion Criteria
Title and Abstract	Must mention temporality extraction in the abstract Must mention working with clinical free text	Review or update articles Articles not written in English
Full text	Must provide information about the method used to address temporality extraction Must provide at least one quantitative measure to evaluate the experiments	Do not provide information about the dataset size and data source

All 118 article were analyzed and most important and summarized in the tables. Methodological steps involved are depicted through figure 5

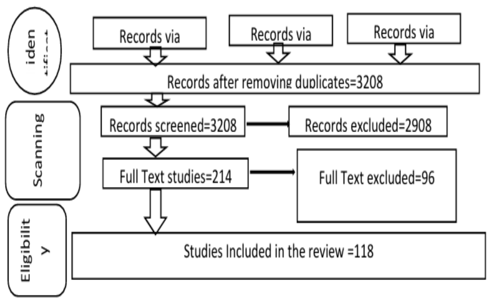

Fig 5: Methodological steps adopted in the study

Datasets

Datasets used for temporal relation extraction form various shared tasks, reports, prescription, cases sheets and so on providing information such as document origin, temporal annotation schema and all related studies among the reviewed articles.There exists a clear difference between TLINKs and DocTimeRel with separate annotations and evaluations in evaluation scripts. The ClinicalTempEval 2015 dataset has around 440 documents averaging 136.05 events,12,43 temporal expressions and 37.43 TLINKs per documents [22]. The ClinicalTempEval 2016 has produced more annotated data with a total of 591 documents averaging 133.42 events ,13.30 temporal Expressions and 39.93 TLINKs. The ClinicalTempEval 2017 dataset contains 769 documents, averaging 120.83 events,12.70 temporal extraction and 33.28 TLINKs per document.

3. Analysis of TLINKs extraction strategies

Analysis of strategies adopted to extract TLINK are analyzed in this section. Reviewed works are sorted and categorized based on strategies used deal with temporality. Three Categories which can be predominantly identified are a) Rule Based Systems b) Machine learning based systems and c) Hybrid systems. An in-depth analysis of all selected article and compiled summary of those articles are also provided here

3.1 Rule Based Systems

Rule based system do not require any classifiers to feed upon.

3.2 Machine learning based systems

In this section we have analyzed the articles that used machine learning based systems for TLINKs extraction. Most articles reviewed in this section are related to shared task datasets. Ordering of temporal segments is the key agenda in [123] and [124]. Pairwise classification and event ranking were tested in [79] and achived better results. Temporal indexing, predicting TLINKs between events and temporal expressions while keeping the most relevant pair for each event. BERT model is used for extracting temporal relation form clinical texts [42]. Training a classifier with required number of positive sample and negative samples depends on candidate generation process. It is evidential that events such i2b2, THYME corpus and ClinicalTempEval produced more negative samples [22,23,24]. Strategy of restricting within sentence relations, was widely used in [6,12,31,34,77,84,86,87,89,90,99,101,104,108,111]. The strategy of using all possible pairs within sentence and specific heuristics to cover cross-sentence was used in [28,83,94,96,105,112,113,127]. In traditional Machine learning, most approaches use SVM. SVM classifier are used in [6,12,28,31,33,43, 94,100,108,127,130]. SVM classifiers outperformed the previously mentioned traditional machine learning. Approaches based on MTL focuses on both DocTimeRel and TLINKs [97]. [12] and [100] developed a strategy for the training set expansion. However, machine learning classifier do not perform as well as deep learning classifiers. Despite all previous there is still scope for research to enhance the candidate pair generation thereby improving the efficacy of temporal relation extraction. Articles related to use of machine learning are provided in figure 6.

Authors	Best strategy	Candidate pair selection	WS	CS	NS	SE	Results
		I2b2 2012 dataset					
[127]	WS: 3 SVM clfs; CS: 3 SVM clfs	WS: APP, CS: rules	√	√			**Fm 0.6849**
[35]	WS: NB clf; CS: NB clf	WS: rules, CS: rules	√	√			Fm 0.671
[130]	2 SVM clfs	Rules	√	√			Fm 0.5594
[37]	BI-LSTM; TS expansion	-			√		Fm 0.6217 (NTC)
		Clinical TempEval 2015 dataset					
[12]	2 SVM clfs. CSL; TS expansion	WS: APP	√			√	Fm 0.321
		Clinical TempEval 2016 dataset					
[103]	BERT; 3 class; MTL	APP over TK	√	√		√	**Fm 0.686**
[102]	BERT; TS expansion; 3-class	APP over TK	√	√		√	**Fm 0.684**
[107]	Context segmentation; Associated ATT; Position weights	APP over TK	√	√		√	Fm 0.643
[89]	WS: tree-based Bi-LSTM-RNN	WS: APP	√			√	Fm 0.633
[99]	Bi-LSTM; TS expansion; 3-class; XML markup	WS: APP	√			√	Fm 0.630
[90]	Tree-based Bi-LSTM-RNN	WS: APP	√			√	Fm 0.629
[95]	LSTM; MTL	APP over TK	√	√		√	Fm 0.628
[101]	SVM clf + CNN; XML markup	WS: APP	√			√	Fm 0.621

Fig 6: Articles adopting Machine learning to extract TLINKs

3.3 Hybrid systems

Figure 7 gives research article employeed hybrid system to extract TLINKs.

Authors	Best strategy	Candidate pair selection	WS	CS	NS	SE	Results
		I2b2 2012 dataset					
[74]	[11] + rules + additional features	WS: rules, CS: rules	√	√			**Fm 0.702**
[72]	WS: 2 ME clfs; CS: 1 ME clf + rules	WS: APP, CS: rules	√	√			**Fm 0.6954**
[12]	WS: 2 SVM clfs; CS: 2 SVM clfs + rules; CSL; TS expansion	WS: APP, CS: rules	√	√			**Fm 0.695**
[11]	WS: 2 SVM clfs; CS: 2 SVM clfs; Rules	WS: rules, CS: rules	√	√			Fm 0.6932
[73]	[11] + rules + additional features	WS: rules, CS: rules	√	√			Fm 0.693
[7]	WS: 2 SVM clfs + temporal graph; CS: rules	WS: APP	√	√			Fm 0.63
[76]	9 clfs + rules	Cross-product	√	√			Fm 0.6231
[71]	2 ME clfs + rules	Rules	√	√			Fm 0.5628
[135]	SVM clf + rules	WS: APP, CS: rules	√	√			Fm 0.537
[23]	WS: ME clf + conflict resolution, CS: rules	WS: rules	√	√			Fm 0.43
[75]	[73, 74]	[73, 74]		√			Fm 0.341 (NTC)
[38]	SVM clfs + rules + CSL	WS: APP	√				Fm 0.6377 (NTC)
		Clinical TempEval 2015 dataset					
[128]	CRF clf + rules	Rules	√	√		√	Fm 0.181
		Clinical TempEval 2016 dataset					
[106]	WS: SVM clf; CS: SVM clf; Rules; 3-class	WS: APP, CS: rules	√	√		√	Fm 0.538

Legend: WS, within-sentence; CS, cross-sentence; NS, not specified; SE, separate evaluation; Fm, f-measure; ME, maximum entropy; clf, classifier; APP, all possible pairs; CSL, cost-sensitive learning; TS, training set; CTE, Clinical TempEval; CRF, conditional random fields; 3-class, transforming to a 3-class classification.

Fig 7: Research articles that used Hybrid system

4. Analysis of DocTimeRel extraction strategies

A thorough analysis of strategies adopted to extract DocTimeRel are analyzed in this section. Reviewed works are sorted and categorized based on strategies used deal with temporality. Three Categories which can be predominantly identified are a) Rule Based Systems b) Machine learning based systems and c) Hybrid systems. An in-depth analysis of all selected article and compiled summary of those articles are also provided here in.

4.1 Rule Based Systems

Rule based system to extract DocTimeRel are categorized in two types in to two types [i] those which derives relationship between the event and the DCT by connecting both and [ii] those which classify the relationships.

Single step temporal information extraction methodology was adopted in [48,50-57]. Detailed temporal information extraction depending on the specific tasks such as regular expression based Tool-Context [56] for the task of extracting event attributes. Context Tool extracts experiencer, negation and Temporality(DocTimeRel). ConText tool is adopted with enhanced rules in [45,57]. It has been observed that performance of rule based systems are not better than machine learning system and hybrid system for I2b2 datasets. Figure 9 gives systems that extracted DocTimeRel with Rule Based Approach

4.2 Machine learning based systems

Use of Machine learning system to extract DocTimeRel are summarized in figure 8. In the research review it is evident most works used support vector machines(SVMs) and Conditional Random Field (CRF) among traditional machine learning classifiers [47,106,67,86]. For shared task – related datasets, i2b2 datasets and regular dataset SVM showed better performance. Research articles that used Machine learning systems to extract DocTimeRel are summarized in figure 10.

Authors	Best Strategy	SE	Results
	I2b2 2012 dataset		
[71]	Rules		Fm 0.5628
[32]	Rules	√	Match ratio 0.69 (NTC)
	I2b2/UTHealth 2014 dataset		
[70]	df + specific rules		Fm 0.915
[67]	Rules		Fm 0.907
[58]	df + context-aware refinement approach		Fm 0.897
[69]	Df		Fm 0.890
[64]	df + specific rules		Fm 0.8776
[65]	Df		Fm 0.875

Fig 8: Research articles that used Machine learning

4.3 Hybrid systems

Review shows hybrid systems to extract DocTimeRel are fewer compared to Machine learning systems and rule based system. There is lack of substantial evidence to support superiority of hybrid systems over machine learning system and rule based system in performance. Figure 9 gives the summarization of hybrid system used to extract DocTimeRel

Authors	Best strategy	SE	Results
	I2b2 2012 dataset		
[7]	SVM clf + rules		Fm 0.63
[130]	SVM clf + rules		Fm 0.5594
	I2b2/UTHealth 2014 dataset		
[66]	3 SVM clfs + df + rules + ann refinement		**Fm 0.9277**
[60]	CART DT + df		Fm 0.917
[61]	Markov networks + rules		Fm 0.9098
[63]	NB clf + rules		Fm 0.8302
	Clinical TempEval 2016 dataset		
[88]	LR clf + rules	√	Fm 0.743

Fig 9: Research articles that used Hybrid system to extract DocTime relations

5. CONCLUSION AND FUTURE WORK

This work reviews the methods, datasets and outcomes of extraction of Temporal Relations and DocTimeRel relations. By the careful examination of 118 papers following conclusions are derived.

1. Most research works use publicly available datasets (i2b2 and THYME) and focuses on a limited TLINK types such as 'before', 'after', 'overlap' and 'contains'.
2. Deep Learning models, State-of-the-art based on **Contextual Embedding and approaches based on** BERT have shown performance improvements over traditional Machine learning models
3. F-score has reached 91.1 % for some particular task. However Temporal relation extraction still has scope for improvements.
4. DocTimeRel relation extraction (a secondary research topic) approaches are mostly relied upon Machine learning models.
5. Most publications on TLINKs are based on a single dataset which limits evaluation of approaches in scenarios of different medical treatment scenarios.
6. Enhanced Candidate pair generation improves overall performance temporal relation extraction.

This study identifies a few topics which potential enough to pose a need for additional research works, which are as follows:

1. Additional TLINK types are to be incorporated for training classifier and analyses the performance.
2. Research can be extended to considering extended Unannoted corpus and report the performance.
3. Extended ways such as Heuristics and semantic ontologies set to enhance Candidate Generation Techniques which may improve F-score of TLINK extraction process.
4. Fine tuning pre trained models with clinical text such as BERT model to understand the clinical context.
5. Exploring state of the art text mining approaches that have not yet used such as attention based neural networks model, minwise hashing and many others.
6. Research on extraction of temporal relations would improve if additional datasets with different medical specialties and clinical text types.

7. REFERENCES

[1] Peter B. Jensen, Lars J. Jensen, and Søren Brunak. 2012. Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics 13, 6 (2012), 395–405. DOI:https://doi.org/10.1038/nrg3208

[2] Daniel Capurro, Meliha Yetisgen, Erik van Eaton, Robert Black, and Peter Tarczy-Hornoch. 2014. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: A multisite assessment. EGEMS 2, 1 (2014), 1079. DOI:https://doi.org/10.13063/2327-9214.1079

[3] Yanshan Wang, Liwei Wang, Majid Rastegar-Mojarad, Sungrim Moon, Feichen Shen, Naveed Afzal, Sijia Liu, Yuqun Zeng, Saeed Mehrabi, Sunghwan Sohn, and Hongfang Liu. 2018. Clinical information extraction applications: A literature review. Journal of Biomedical Informatics 77 (2018), 34–49. DOI:https://doi.org/10.1016/j.jbi.2017.11.011

[4] Kory Kreimeyer, Matthew Foster, Abhishek Pandey, Nina Arya, Gwendolyn Halford, Sandra F Jones, Richard Forshee, Mark Walderhaug, and Taxiarchis Botsis. 2017. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. Journal of Biomedical Informatics 73 (2017), 14–29. DOI:https://doi.org/10.1016/j.jbi.2017.07.012

[5] Sumithra Velupillai, Danielle Mowery, Brett South, Maria Kvist, and Hercules Dalianis. 2015. Recent advances in clinical natural language processing in support of semantic analysis. Yearbook of medical informatics 10 (2015), 183– 93. DOI:https://doi.org/10.15265/IY-2015-009 ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. 144:28 Y. B. Gumiel et al.

[6] William F. Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky. 2014. Temporal annotation in the clinical domain. Transactions of the Association for Computational Linguistics 2 (2014), 143–154. DOI:https://doi.org/10.1162/ tacl_a_00172

[7] Azadeh Nikfarjam, Ehsan Emadzadeh, and Graciela Gonzalez. 2013. Towards generating a patient's timeline: Extracting temporal relationships from clinical notes. Journal of Biomedical Informatics 46 (2013), S40–S47. DOI:https: //doi.org/10.1016/j.jbi.2013.11.001

[8] Seyedmostafa Sheikhalishahi, Riccardo Miotto, Joel T. Dudley, Alberto Lavelli, Fabio Rinaldi, and Venet Osmani. 2019. Natural language processing of clinical notes on chronic diseases: Systematic review. JMIR Med Inform 7, 2 (2019), e12239. DOI:https://doi.org/10.2196/12239

[9] World Health Organization. 2020. Noncommunicable Diseases: Progress Monitor 2020. World Health Organization. vi, 224 pages.

[10] James F. Allen. 1983. Maintaining knowledge about temporal intervals. Communications of the ACM 26, 11 (1983), 832–843. DOI:https://doi.org/10.1145/182.358434

[11] Buzhou Tang, Yonghui Wu, Min Jiang, Yukun Chen, Joshua C. Denny, and Hua Xu. 2013. A hybrid system for temporal information extraction from clinical text. Journal of the American Medical Informatics Association 20, 5

(2013), 828–835. DOI:https://doi.org/10.1136/amiajnl-2013-001635

[12] Chen Lin, Dmitriy Dligach, Timothy A. Miller, Steven Bethard, and Guergana K. Savova. 2016. Multilayered temporal modeling for the clinical domain. Journal of the American Medical Informatics Association 23, 2 (2016), 387–395. DOI:https://doi.org/10.1093/jamia/ocv113

[13] Robert Leaman, Ritu Khare, and Zhiyong Lu. 2015. Challenges in clinical natural language processing for automated disorder normalization. Journal of Biomedical Informatics 57, C (2015), 28–37. DOI:https://doi.org/10.1016/j.jbi.2015. 07.010

[14] Stephane M. Meystre, Guergana K. Savova, Karin C. Kipper-Schuler, and John F. Hurdle. 2008. Extracting information from textual documents in the electronic health record: A review of recent research. Yearbook of Medical Informatics (2008), 128–144.

[15] Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. Journal of the American Medical Informatics Association 20, 5 (2013), 806–813. DOI:https://doi.org/10. 1136/amiajnl-2013-001628

[16] Steven Bethard, Leon Derczynski, Guergana Savova, James Pustejovsky, and Marc Verhagen. 2015. SemEval-2015 Task 6: Clinical TempEval. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval'15). Association for Computational Linguistics, Denver, Colorado, 806–814. DOI:https://doi.org/10.18653/v1/S15-2136

[17] Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen. 2016. SemEval-2016 Task 12: Clinical TempEval. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1052–1062. DOI:https://doi.org/10. 18653/v1/S16-1165

[18] Steven Bethard, Guergana Savova, Martha Palmer, and James Pustejovsky. 2017. SemEval-2017 Task 12: Clinical TempEval. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 565–572. DOI:https://doi.org/10.18653/v1/S17-2093

[19] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLOS Medicine 6, 7 (2009), 1–6. DOI:https://doi.org/10.1371/journal.pmed.1000097

[20] Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Temporal reasoning over clinical text: The state of the art. Journal of the American Medical Informatics Association 20, 5 (2013), 814–819. DOI:https://doi.org/10.1136/amiajnl2013-001760

[21] Ghada Alfattni, Niels Peek, and Goran Nenadic. 2020. Extraction of temporal relations from clinical free text: A systematic review of current approaches. Journal of Biomedical Informatics 108 (2020), 103488. DOI:https://doi.org/ 10.1016/j.jbi.2020.103488

[22] James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. ISO-TimeML: An international standard for semantic annotation. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta, 394–397.

[23] Yao Cheng, Peter Anick, Pengyu Hong, and Nianwen Xue. 2013. Temporal relation discovery between events and temporal expressions identified in clinical narrative. Journal of Biomedical Informatics 46 (2013), S48–S53. DOI:https: //doi.org/10.1016/j.jbi.2013.09.010

[24] James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. Language Resources and Evaluation 39, 2 (2005), 123–164. DOI:https://doi.org/10.1007/s10579- 005-7882-7 ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. Temporal Relation Extraction in Clinical Texts: A Systematic Review 144:29 [25] Marc Verhagen. 2005. Temporal closure in an annotation environment. Language Resources and Evaluation 39, 2 (2005), 211–241. DOI:https://doi.org/10.1007/s10579-005-7884-5

[26] Leon Derczynski. 2016. Representation and learning of temporal relations. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. The COLING 2016 Organizing Committee, Osaka, Japan, 1937–1948.

[27] James Pustejovsky and Amber Stubbs. 2011. Increasing informativeness in temporal annotation. In Proceedings of the 5th Linguistic Annotation Workshop. Association for Computational Linguistics, Portland, Oregon, 152–160.

[28] Abdulrahman Khalifa, Sumithra Velupillai, and Stephane Meystre. 2016. UtahBMI at SemEval-2016 Task 12: Extracting temporal information from clinical text. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1256–1262. DOI:https: //doi.org/10.18653/v1/S16-1195

[29] Guergana K. Savova, James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. 2010. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. Journal of the American Medical Informatics Association 17, 5 (2010), 507– 513. DOI:https://doi.org/10.1136/jamia.2009.001560

[30] Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proceedings AMIA Symposium (2001), 17–21.

[31] Julien Tourille, Olivier Ferret, Xavier Tannier, and Aurélie Névéol. 2017. Temporal information extraction from clinical text. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, Valencia, Spain, 739–745.

[32] Wei Wang, Kory Kreimeyer, Emily Jane Woo, Robert Ball, Matthew Foster, Abhishek Pandey, John Scott, and Taxiarchis Botsis. 2016. A new algorithmic approach for the extraction of temporal associations from clinical narratives with an application to medical product safety surveillance reports. Journal of Biomedical Informatics 62

(2016), 78–89. DOI:https://doi.org/10.1016/j.jbi.2016.06.006

[33] Timothy Miller, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, Chen Lin, and Guergana Savova. 2013. Discovering temporal narrative containers in clinical text. In Proceedings of the 2013 Workshop on Biomedical Natural Language Processing. Association for Computational Linguistics, Sofia, Bulgaria, 18–26. [34] Chen Lin, Timothy Miller, Alvin Kho, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, and Guergana Savova. 2014. Descending-path convolution kernel for syntactic structures. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Baltimore, Maryland, 81–86. DOI:https://doi.org/10.3115/v1/P14-2014

[35] Gandhimathi Moharasan and Tu-Bao Ho. 2019. Extraction of temporal information from clinical narratives. Journal of Healthcare Informatics Research 3, 2 (2019), 220–244. DOI:https://doi.org/10.1007/s41666-019-00049-0

[36] Thanat Chokwijitkul, Anthony Nguyen, Hamed Hassanzadeh, and Siegfried Perez. 2018. Identifying risk factors for heart disease in electronic medical records: A deep learning approach. In Proceedings of the BioNLP 2018 Workshop. Association for Computational Linguistics, Melbourne, Australia, 18–27. DOI:https://doi.org/10.18653/v1/W18-2303

[37] Zixu Wang, Julia Ive, Sumithra Velupillai, and Lucia Specia. 2019. Is artificial data useful for biomedical Natural Language Processing algorithms?. In Proceedings of the 18th BioNLP Workshop and Shared Task. Association for Computational Linguistics, Florence, Italy, 240–249. DOI:https://doi.org/10.18653/v1/W19-5026

[38] Hee Jin Lee, Yaoyun Zhang, Min Jiang, Jun Xu, Cui Tao, and Hua Xu. 2018. Identifying direct temporal relations between time and events from clinical notes. BMC Medical Informatics and Decision Making 18, Suppl 2 (2018), 49. DOI:https://doi.org/10.1186/s12911-018-0627-5

[39] Dong Xu, Meizhuo Zhang, Tianwan Zhao, Chen Ge, Weiguo Gao, Jia Wei, and Kenny Q. Zhu. 2015. Data-driven information extraction from Chinese electronic medical records. PLOS ONE 10, 8 (2015), 1–18. DOI:https://doi.org/ 10.1371/journal.pone.0136270

[40] Jia Su, Jinpeng Hu, Jingchi Jiang, Jing Xie, Yang Yang, Bin He, Jinfeng Yang, and Yi Guan. 2019. Extraction of risk factors for cardiovascular diseases from Chinese electronic medical records. Computer Methods and Programs in Biomedicine 172 (2019), 1–10.

[41] Zengjian Liu, Xiaolong Wang, Qingcai Chen, Buzhou Tang, and Hua Xu. 2019. Temporal indexing of medical entity in Chinese clinical notes. BMC Medical Informatics and Decision Making 19, 1 (2019), 17. DOI:https://doi.org/10.1186/ s12911-019-0735-x

[42] Xiaohui Zhang, Yaoyun Zhang, Qin Zhang, Yuankai Ren, Tinglin Qiu, Jianhui Ma, and Qiang Sun. 2019. Extracting comprehensive clinical information for breast cancer using deep learning methods. International Journal of Medical Informatics 132 (2019), 103985. DOI:https://doi.org/10.1016/j.ijmedinf.2019.103985

[43] Jae-Wook Seol, Wangjin Yi, Jinwook Choi, and Kyung Soon Lee. 2017. Causality patterns and machine learning for the extraction of problem-action relations in discharge summaries. International Journal of Medical Informatics 98 (2017), 1–12. DOI:https://doi.org/10.1016/j.ijmedinf.2016.10.021
ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. 144:30 Y. B. Gumiel et al.

[44] Wangjin Lee and Jinwook Choi. 2018. Temporal segmentation for capturing snapshots of patient histories in Korean clinical narrative. Healthcare Informatics Research 24, 3 (2018), 179–186. DOI:https://doi.org/10.4258/hir.2018.24.3.179

[45] Zubair Afzal, Ewoud Pons, Ning Kang, Miriam C. J. M. Sturkenboom, Martijn J. Schuemie, and Jan A. Kors. 2014. ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. BMC Bioinformatics 15, 1 (2014), 373. DOI:https://doi.org/10.1186/s12859-014- 0373-3

[46] Natalia Viani, Timothy A. Miller, Carlo Napolitano, Silvia G. Priori, Guergana K. Savova, Riccardo Bellazzi, and Lucia Sacchi. 2019. Supervised methods to extract clinical events from cardiology reports in Italian. Journal of Biomedical Informatics 95 (2019), 103219. DOI:https://doi.org/10.1016/j. jbi.2019.103219

[47] Aron Henriksson, Maria Kvist, Hercules Dalianis, and Martin Duneld. 2015. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. Journal of Biomedical Informatics 57 (2015), 333–349. DOI:https://doi.org/10.1016/j.jbi.2015.08.013

[48] Marjan Najafabadipour, Massimiliano Zanin, Alejandro Rodríguez-González, Maria Torrente, Beatriz Nuñez García, Juan Luis Cruz Bermudez, Mariano Provencio, and Ernestina Menasalvas. 2020. Reconstructing the patient's natural history from electronic health records. Artificial Intelligence in Medicine 105 (2020), 101860. DOI:https://doi.org/10. 1016/j.artmed.2020.101860

[49] Leonardo Campillos, Louise Deléger, Cyril Grouin, Thierry Hamon, Anne-Laure Ligozat, and Aurélie Névéol. 2018. A French clinical corpus with comprehensive semantic annotations: Development of the Medical Entity and Relation LIMSI annOtated Text corpus (MERLOT). Language Resources and Evaluation 52, 2 (2018), 571–601. DOI:https://doi. org/10.1007/s10579-017-9382-y

[50] Srinivasan V. Iyer, Rave Harpaz, Paea LePendu, Anna Bauer-Mehren, and Nigam H. Shah. 2014. Mining clinical text for signals of adverse drug-drug interactions. Journal of the American Medical Informatics Association 21, 2 (2014), 353–362. DOI:http://dx.doi.org/10.1136/amiajnl-2013-001612

[51] Srinivasan V. Iyer, Paea Lependu, Rave Harpaz, Anna Bauer-Mehren, and Nigam H. Shah. 2013. Learning signals of adverse drug-drug interactions from the unstructured text of electronic health records. AMIA Joint Summits on Translational Science Proceedings 2013 (2013), 83–87.

[52] Yi Liu, Paea Lependu, Srinivasan Iyer, and Nigam H. Shah. 2012. Using temporal patterns in medical records to discern adverse drug events from indications. AMIA Joint

Summits on Translational Science Proceedings 2012 (2012), 47–56.

[53] Hee-Jin Lee, Min Jiang, Yonghui Wu, Christian M. Shaffer, John H. Cleator, Eitan A. Friedman, Joshua P. Lewis, Dan M. Roden, Josh Denny, and Hua Xu. 2017. A comparative study of different methods for automatic identification of clopidogrel-induced bleedings in electronic health records. AMIA Joint Summits on Translational Science Proceedings 2017 (2017), 185–192.

[54] Cosmin A. Bejan, Lucy Vanderwende, Heather L. Evans, Mark M. Wurfel, and Meliha Yetisgen-Yildiz. 2013. Ontime clinical phenotype prediction based on narrative reports. AMIA Annual Symposium Proceedings 2013 (2013), 103–110.

[55] Stephen T. Wu, Young J. Juhn, Sunghwan Sohn, and Hongfang Liu. 2014. Patient-level temporal aggregation for textbased asthma status ascertainment. Journal of the American Medical Informatics Association 21, 5 (2014), 876–884. DOI:https://doi.org/10.1136/amiajnl-2013-002463

[56] Henk Harkema, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. 2009. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. Journal of Biomedical Informatics 42, 5 (2009), 839–851. DOI:https://doi.org/10.1016/j.jbi.2009.05.002

[57] Rob Gaizauskas, Henk Harkema, Mark Hepple, and Andrea Setzer. 2006. Task-oriented extraction of temporal information: The case of clinical narratives. In 13th International Symposium on Temporal Representation and Reasoning (TIME'06). 188–195. DOI:https://doi.org/10.1109/TIME. 2006.27

[58] Nai-Wen Chang, Hong-Jie Dai, Jitendra Jonnagaddala, Chih-Wei Chen, Richard Tzong-Han Tsai, and Wen-Lian Hsu. 2015. A context-aware approach for progression tracking of medical concepts in electronic medical records. Journal of Biomedical Informatics 58, S (2015), S150–S157. DOI:https://doi.org/10.1016/j.jbi.2015.09.013

[59] Qingcai Chen, Haodi Li, Buzhou Tang, Xiaolong Wang, Xin Liu, Zengjian Liu, Shu Liu, Weida Wang, Qiwen Deng, Suisong Zhu, Yangxin Chen, and Jingfeng Wang. 2015. An automatic system to identify heart disease risk factors in clinical texts over time. Journal of Biomedical Informatics 58 (2015), S158–S163. DOI:https://doi.org/10.1016/j.jbi. 2015.09.002

[60] James Cormack, Chinmoy Nath, David Milward, Kalpana Raja, and Siddhartha R. Jonnalagadda. 2015. Agile text mining for the 2014 i2b2/UTHealth Cardiac Risk Factors Challenge. Journal of Biomedical Informatics 58 (2015), S120–S127. DOI:https://doi.org/10.1016/j.jbi.2015.06.030

[61] Travis Goodwin and Sanda M. Harabagiu. 2015. A probabilistic reasoning method for predicting the progression of clinical findings from electronic medical records. AMIA Joint Summits on Translational Science Proceedings 2015 (2015), 61–65. ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. Temporal Relation Extraction in Clinical Texts: A Systematic Review 144:31

[62] Cyril Grouin, Véronique Moriceau, and Pierre Zweigenbaum. 2015. Combining glass box and black box evaluations in the identification of heart disease risk factors and their temporal relations from clinical records. Journal of Biomedical Informatics 58 (2015), S133–S142. DOI:https://doi.org/10.1016/j.jbi.2015.06.014

[63] Jitendra Jonnagaddala, Siaw-Teng Liaw, Pradeep Ray, Manish Kumar, Hong-Jie Dai, and Chien-Yeh Hsu. 2015. Identification and progression of heart disease risk factors in diabetic patients from longitudinal electronic health records. BioMed Research International 2015 (2015), 636371. DOI:https://doi.org/10.1155/2015/636371

[64] George Karystianis, Azad Dehghan, Aleksandar Kovacevic, John A. Keane, and Goran Nenadic. 2015. Using local lexicalized rules to identify heart disease risk factors in clinical notes. Journal of Biomedical Informatics 58 Suppl, Suppl (2015), S183–S188. DOI:https://doi.org/10.1016/j.jbi.2015.06.013

[65] Abdulrahman Khalifa and Stéphane Meystre. 2015. Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes. Journal of Biomedical Informatics 58, S (2015), S128–S132. DOI:https://doi.org/10.1016/j.jbi.2015.08.002

[66] Kirk Roberts, Sonya E. Shooshan, Laritza Rodriguez, Swapna Abhyankar, Halil Kilicoglu, and Dina DemnerFushman. 2015. The role of fine-grained annotations in supervised recognition of risk factors for heart disease from EHRs. Journal of Biomedical Informatics 58, S (2015), S111–S119. DOI:https://doi.org/10.1016/j.jbi.2015.06.010

[67] Chaitanya P. Shivade, Pranav Malewadkar, Eric Fosler-Lussier, and Albert M. Lai. 2015. Comparison of UMLS terminologies to identify risk of heart disease using clinical notes. Journal of Biomedical Informatics 58 (2015), S103–S110. DOI:https://doi.org/10.1016/j.jbi.2015.08.025

[68] Manabu Torii, Jung wei Fan, Wei li Yang, Theodore Lee, Matthew T. Wiley, Daniel S. Zisook, and Yang Huang. 2015. Risk factor detection for heart disease by applying text analytics in electronic medical records. Journal of Biomedical Informatics 58 (2015), S164–S170. DOI:https://doi.org/10.1016/j.jbi.2015.08.011

[69] Jay Urbain. 2015. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. Journal of Biomedical Informatics 58 (2015), S143–S149. DOI:https://doi.org/10.1016/j.jbi.2015.08. 009

[70] Hui Yang and Jonathan M. Garibaldi. 2015. A hybrid model for automatic identification of risk factors for heart disease. Journal of Biomedical Informatics 58 Suppl, Suppl (2015), S171–S182. DOI:https://doi.org/10.1016/j.jbi.2015. 09.006

[71] Yung-Chun Chang, Hong-Jie Dai, Johnny Chi-Yang Wu, Jian-Ming Chen, Richard Tzong-Han Tsai, and Wen-Lian Hsu. 2013. TEMPTING System: A hybrid method of rule and machine learning for temporal relation extraction in patient discharge summaries. Journal of Biomedical Informatics 46 (2013), S54–S62. DOI:https://doi.org/10.1016/j.jbi. 2013.09.007

[72] Colin Cherry, Xiaodan Zhu, Joel Martin, and Berry de Bruijn. 2013. À la recherche du temps perdu: Extracting temporal relations from medical text in the 2012 i2b2 NLP challenge. Journal of the American Medical Informatics Association 20, 5 (2013), 843–848. DOI:https://doi.org/10.1136/amiajnl-2013-001624 [73] Jennifer D'Souza and Vincent Ng. 2013. Classifying temporal relations in clinical data: A hybrid, knowledge-rich approach. Journal of Biomedical Informatics 46 (2013), S29–S39. DOI:https://doi.org/10.1016/j.jbi.2013.08.003

[74] Jennifer D'Souza and Vincent Ng. 2014. Knowledge-rich temporal relation identification and classification in clinical notes. Database 2014 (11 2014). DOI:https://doi.org/10.1093/database/bau109

[75] Jennifer D'Souza and Vincent Ng. 2014. Annotating inter-sentence temporal relations in clinical notes. In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland, 2758–2765.

[76] Cyril Grouin, Natalia Grabar, Thierry Hamon, Sophie Rosset, Xavier Tannier, and Pierre Zweigenbaum. 2013. Eventual situations for timeline extraction from clinical reports. Journal of the American Medical Informatics Association 20, 5 (2013), 820–827. DOI:https://doi.org/10.1136/amiajnl-2013-001627

[77] Zhijing Li, Chen Li, Yu Long, and Xuan Wang. 2020. A system for automatically extracting clinical events with temporal information. BMC Medical Informatics and Decision Making 20, 1 (2020), 198. DOI:https://doi.org/10.1186/ s12911-020-01208-9

[78] Danielle Mowery, Henk Harkema, John Dowling, Jonathan Lustgarten, and Wendy Chapman. 2009. Distinguishing historical from current problems in clinical reports—which textual features help?. In Proceedings of the BioNLP 2009 Workshop. Association for Computational Linguistics, Boulder, Colorado, 10–18.

[79] Preethi Raghavan, Albert Lai, and Eric Fosler-Lussier. 2012. Learning to temporally order medical events in clinical text. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Jeju Island, Korea, 70–74.

[80] Preethi Raghavan, Eric Fosler-Lussier, and Albert Lai. 2012. Temporal classification of medical events. In BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. Association for Computational Linguistics, Montréal, Canada, 29–37. ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. 144:32 Y. B. Gumiel et al.

[81] Preethi Raghavan, Eric Fosler-Lussier, and Albert Lai. 2012. Exploring semi-supervised coreference resolution of medical concepts using semantic and temporal features. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Montréal, Canada, 731–741.

[82] Chen Lin, Elizabeth W. Karlson, Dmitriy Dligach, Monica P. Ramirez, Timothy A. Miller, Huan Mo, Natalie S. Braggs, Andrew Cagan, Vivian Gainer, Joshua C. Denny, and Guergana K. Savova. 2015. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. Journal of the American Medical Informatics Association 22, e1 (2015), e151–161. DOI:https://doi.org/10.1136/amiajnl-2014- 002642

[83] Marcia Barros, Andre Lamurias, Gonçalo Figueiro, Marta Antunes, Joana Teixeira, Alexandre Pinheiro, and Francisco M. Couto. 2016. ULISBOA at SemEval-2016 Task 12: Extraction of temporal expressions, clinical events and relations using IBEnt. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1263–1267. DOI:https://doi.org/10.18653/v1/S16-1196

[84] Tommaso Caselli and Roser Morante. 2016. VUACLTL at SemEval 2016 Task 12: A CRF pipeline to clinical TempEval. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1241–1247. DOI:https://doi.org/10.18653/v1/S16-1193 [85] Veera Raghavendra Chikka. 2016. CDE-IIITH at SemEval-2016 Task 12: Extraction of temporal information from clinical documents using machine learning techniques. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1237–1240. DOI:https: //doi.org/10.18653/v1/S16-1192

[86] Arman Cohan, Kevin Meurer, and Nazli Goharian. 2016. GUIR at SemEval-2016 Task 12: Temporal information processing for clinical narratives. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1248–1255. DOI:https://doi.org/10.18653/v1/S16- 1194

[87] Dmitriy Dligach, Timothy Miller, Chen Lin, Steven Bethard, and Guergana Savova. 2017. Neural temporal relation extraction. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, Valencia, Spain, 746–751.

[88] Jason Fries. 2016. Brundlefly at SemEval-2016 Task 12: Recurrent neural networks vs. Joint Inference for Clinical Temporal Information Extraction. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1274–1279. DOI:https://doi.org/10. 18653/v1/S16-1198

[89] Diana Galvan, Koji Matsuda, Naoaki Okazaki, and Kentaro Inui. 2020. Empirical exploration of the challenges in temporal relation extraction from clinical text. Journal of Natural Language Processing 27, 2 (2020), 383–409. DOI:https://doi.org/10.5715/jnlp.27.383

[90] Diana Galvan, Naoaki Okazaki, Koji Matsuda, and Kentaro Inui. 2018. Investigating the challenges of temporal relation extraction from clinical text. In Proceedings of the 9th International Workshop on Health Text Mining and Information Analysis. Association for Computational Linguistics, Brussels, Belgium, 55–64. DOI:https://doi.org/10. 18653/v1/W18-5607

[91] Cyril Grouin and Véronique Moriceau. 2016. LIMSI at SemEval-2016 Task 12: Machine-learning and temporal information to identify clinical events and time expressions. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1225–1230. DOI:https://doi.org/10.18653/v1/S16-1190

[92] Serena Jeblee and Graeme Hirst. 2018. Listwise temporal ordering of events in clinical notes. In Proceedings of the 9th International Workshop on Health Text Mining and Information Analysis. Association for Computational Linguistics, Brussels, Belgium, 177–182. DOI:https://doi.org/10.18653/v1/W18-5620

[93] Catherine Kerr, Terri Hoare, Paula Carroll, and Jakub Mareček. 2020. Integer programming ensemble of temporal relations classifiers. Data Mining and Knowledge Discovery 34, 2 (2020), 533–562. DOI:https://doi.org/10.1007/s10618- 019-00671-x

[94] Hee-Jin Lee, Hua Xu, Jingqi Wang, Yaoyun Zhang, Sungrim Moon, Jun Xu, and Yonghui Wu. 2016. UTHealth at SemEval'16 Task 12: An end-to-end system for temporal information extraction from clinical notes. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1292–1297. DOI:https://doi.org/10.18653/v1/S16-1201

[95] Artuur Leeuwenberg and Marie-Francine Moens. 2018. Word-level loss extensions for neural temporal relation classification. In Proceedings of the 27th International Conference on Computational Linguistics. Association for Computational Linguistics, Santa Fe, New Mexico, 3436–3447.

[96] Artuur Leeuwenberg and Marie-Francine Moens. 2016. KULeuven-LIIR at SemEval 2016 Task 12: Detecting narrative containment in clinical records. In Proceedings of the 10th International Workshop on Semantic Evaluation ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. Temporal Relation Extraction in Clinical Texts: A Systematic Review 144:33 (SemEval'16). Association for Computational Linguistics, San Diego, California, 1280–1285. DOI:https://doi.org/10. 18653/v1/S16-1199 [97] Artuur Leeuwenberg and Marie-Francine Moens. 2017. Structured learning for temporal relation extraction from clinical records. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Association for Computational Linguistics, Valencia, Spain, 1150–1158.

[98] Peng Li and Heng Huang. 2016. UTA DLNLP at SemEval-2016 Task 12: Deep learning based natural language processing system for clinical information identification from clinical notes and pathology reports. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1268–1273. DOI:https://doi.org/10.18653/v1/S16-1197

[99] Chen Lin, Timothy Miller, Dmitriy Dligach, Hadi Amiri, Steven Bethard, and Guergana Savova. 2018. Self-training improves recurrent neural networks performance for temporal relation extraction. In Proceedings of the 9th International Workshop on Health Text Mining and Information Analysis. Association for Computational Linguistics, Brussels, Belgium, 165–176. DOI:https://doi.org/10.18653/v1/W18-5619

[100] Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova. 2016. Improving temporal relation extraction with training instance augmentation. In Proceedings of the 15th Workshop on Biomedical Natural Language Processing. Association for Computational Linguistics, Berlin, Germany, 108–113. DOI:https://doi.org/10. 18653/v1/W16-2914

[101] Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova. 2017. Representations of time expressions for temporal relation extraction with convolutional neural networks. In BioNLP 2017. Association for Computational Linguistics, Vancouver, Canada, 322–327. DOI:https://doi.org/10.18653/v1/W17-2341

[102] Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova. 2019. A BERT-based universal model for both within- and cross-sentence clinical temporal relation extraction. In Proceedings of the 2nd Clinical Natural Language Processing Workshop. Association for Computational Linguistics, Minneapolis, Minnesota, 65–71. DOI:https://doi.org/10.18653/v1/W19-1908

[103] Chen Lin, Timothy Miller, Dmitriy Dligach, Farig Sadeque, Steven Bethard, and Guergana Savova. 2020. A BERTbased one-pass multi-task model for clinical temporal relation extraction. In Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing. Association for Computational Linguistics, Online, 70–75. DOI:https: //doi.org/10.18653/v1/2020.bionlp-1.7

[104] Sijia Liu, Liwei Wang, Vipin Chaudhary, and Hongfang Liu. 2019. Attention neural model for temporal relation extraction. In Proceedings of the 2nd Clinical Natural Language Processing Workshop. Association for Computational Linguistics, Minneapolis, Minnesota, 134–139. DOI:https://doi.org/10.18653/v1/W19-1917

[105] Julien Tourille, Olivier Ferret, Aurélie Névéol, and Xavier Tannier. 2017. Neural architecture for temporal relation extraction: A Bi-LSTM approach for detecting narrative containers. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Vancouver, Canada, 224–230. DOI:https://doi.org/10.18653/v1/P17-2035

[106] Julien Tourille, Olivier Ferret, Aurélie Névéol, and Xavier Tannier. 2016. LIMSI-COT at SemEval-2016 Task 12: Temporal relation identification using a pipeline of classifiers. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16). Association for Computational Linguistics, San Diego, California, 1136–1142. DOI:https://doi.org/10.18653/v1/S16-1175

[107] Shiyi Zhao, Lishuang Li, Hongbin Lu, Anqiao Zhou, and Shuang Qian. 2019. Associative attention networks for temporal relation extraction from electronic health records. Journal of Biomedical Informatics 99 (2019), 103309. DOI:https://doi.org/10.1016/j.jbi.2019.103309

[108] Po-Yu Huang, Hen-Hsen Huang, Yu-Wun Wang, Ching Huang, and Hsin-Hsi Chen. 2017. NTU-1 at SemEval-

2017 Task 12: Detection and classification of temporal events in clinical data with domain adaptation. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 1010–1013. DOI:https://doi.org/10.18653/v1/S17-2177 [109] Artuur Leeuwenberg and Marie-Francine Moens. 2017. KULeuven-LIIR at SemEval'17 Task 12: Cross-domain temporal information extraction from clinical records. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 1030–1034. DOI:https: //doi.org/10.18653/v1/S17-2181

[110] Yu Long, Zhijing Li, Xuan Wang, and Chen Li. 2017. XJNLP at SemEval-2017 Task 12: Clinical temporal information extraction with a hybrid model. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 1014–1018. DOI:https://doi.org/10.18653/ v1/S17-2178

[111] Sean MacAvaney, Arman Cohan, and Nazli Goharian. 2017. GUIR at SemEval-2017 Task 12: A framework for cross-domain clinical temporal information extraction. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 1024–1029. DOI:https: //doi.org/10.18653/v1/S17-2180 ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. 144:34 Y. B. Gumiel et al.

[112] Sarath P. R. Manikandan Ravikiran, and Yoshiki Niwa. 2017. Hitachi at SemEval-2017 Task 12: System for temporal information extraction from clinical notes. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 1005–1009. DOI:https://doi.org/10. 18653/v1/S17-2176

[113] Julien Tourille, Olivier Ferret, Xavier Tannier, and Aurélie Névéol. 2017. LIMSI-COT at SemEval-17 Task 12: Neural architecture for temporal information extraction from clinical narratives. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17). Association for Computational Linguistics, Vancouver, Canada, 597– 602. DOI:https://doi.org/10.18653/v1/S17-2098

[114] Li Rumeng, Jagannatha Abhyuday N, and Yu Hong. 2017. A hybrid neural network model for joint prediction of presence and period assertions of medical events in clinical notes. AMIA Annual Symposium Proceedings 2017 (2017), 1149–1158.

[115] Min Li and Jon Patrick. 2012. Extracting temporal information from electronic patient records. AMIA Annual Symposium Proceedings 2012 (2012), 542–551.

[116] Joshua C. Denny, Josh F. Peterson, Neesha N. Choma, Hua Xu, Randolph A. Miller, Lisa Bastarache, and Neeraja B. Peterson. 2010. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. Journal of the American Medical Informatics Association 17, 4 (2010), 383–388. DOI:https://doi.org/10.1136/jamia. 2010.004804

[117] Sijia Liu, Liwei Wang, Donna Ihrke, Vipin Chaudhary, Cui Tao, Chunhua Weng, and Hongfang Liu. 2017. Correlating lab test results in clinical notes with structured lab data: A case study in HbA1c and glucose. AMIA Joint Summits on Translational Science Proceedings 2017 (2017), 221–228.

[118] Guergana K. Savova, Janet E. Olson, Sean P. Murphy, Victoria L. Cafourek, Fergus J. Couch, Matthew P. Goetz, James N. Ingle, Vera J. Suman, Christopher G. Chute, and Richard M. Weinshilboum. 2012. Automated discovery of drug treatment patterns for endocrine therapy of breast cancer within an electronic medical record. Journal of the American Medical Informatics Association 19, e1 (2012). DOI:https://doi.org/10.1136/amiajnl-2011-000295

[119] Natalia Viani, Joyce Kam, Lucia Yin, Somain Verma, Robert Stewart, Rashmi Patel, and Sumithra Velupillai. 2019. Annotating temporal relations to determine the onset of psychosis symptoms. Studies in Health Technology and Informatics 264 (2019), 418–422. DOI:https://doi.org/10.3233/SHTI190255 [120] Amber Stubbs and Benjamin Harshfield. 2010. Applying the TARSQI toolkit to augment text mining of EHRs. In Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. Association for Computational Linguistics, Uppsala, Sweden, 141–143.

[121] Li Zhou, Simon Parsons, and George Hripcsak. 2008. The evaluation of a temporal reasoning system in processing clinical discharge summaries. Journal of the American Medical Informatics Association 15, 1 (2008), 99–106. DOI:https: //doi.org/10.1197/jamia.M2467

[122] Daniel Capurro, Mario Barbe, Claudio Daza, Josefa Santa María, Javier Trincado, and Ignacio Gomez. 2015. ClinicalTime: Identification of patients with acute kidney injury using temporal abstractions and temporal pattern matching. AMIA Joint Summits on Translational Science Proceedings 2015 (2015), 46–50.

[123] Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006. Finding temporal order in discharge summaries. AMIA Annual Symposium Proceedings 2006 (2006), 81–85.

[124] Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006. Inducing temporal graphs. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Sydney, Australia, 189–198.

[125] Preethi Raghavan, Eric Fosler-Lussier, Noémie Elhadad, and Albert M. Lai. 2014. Cross-narrative temporal ordering of medical events. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Baltimore, Maryland, 998–1008. DOI:https://doi.org/10. 3115/v1/P14-1094

[126] Amber Stubbs and Özlem Uzuner. 2015. Annotating risk factors for heart disease in clinical narratives for diabetic patients. Journal of Biomedical Informatics 58 (2015), S78–S91. DOI:https://doi.org/10.1016/j.jbi.2015.05.009

[127] Yan Xu, Yining Wang, Tianren Liu, Junichi Tsujii, and Eric I-Chao Chang. 2013. An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge. Journal of the American Medical Informatics Association 20, 5 (2013), 849–858. DOI:https://doi.org/10.1136/amiajnl-2012-001607

[128] Sumithra Velupillai, Danielle L Mowery, Samir Abdelrahman, Lee Christensen, and Wendy Chapman. 2015. BluLab: Temporal information extraction for the 2015 clinical TempEval challenge. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval'15). Association for Computational Linguistics, Denver, Colorado, 815– 819. DOI:https://doi.org/10.18653/v1/S15-2137

[129] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. DOI:https://doi.org/10.18653/v1/N19-1423 ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. Temporal Relation Extraction in Clinical Texts: A Systematic Review 144:35

[130] Kirk Roberts, Bryan Rink, and Sanda M. Harabagiu. 2013. A flexible framework for recognizing events, temporal expressions, and temporal relations in clinical text. Journal of the American Medical Informatics Association 20, 5 (2013), 867–875. DOI:https://doi.org/10.1136/amiajnl-2013-001619

[131] Roser Saurí, Jessica Moszkowicz, Bob Knippen, Rob Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. TimeML Annotation Guidelines Version 1.2.1. (2006).

[132] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: From pairwise approach to listwise approach. In Proceedings of the 24th International Conference on Machine Learning. ACM, New York, NY, 129–136. DOI:https://doi.org/10.1145/1273496.1273513

[133] Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Berlin, Germany, 1105–1116. DOI:https://doi.org/10.18653/v1/P16-1105

[134] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36, 4 (2019), 1234–1240. DOI:https://doi.org/10.1093/bioinformatics/btz682

[135] Sunghwan Sohn Dr., Kavishwar B. Wagholikar, Dingcheng Li, Siddhartha R. Jonnalagadda, Cui Tao, Ravikumar Komandur Elayavilli, and Hongfang Liu. 2013. Comprehensive temporal information detection from clinical text: Medical events, time, and TLINK identification. Journal of the American Medical Informatics Association 20, 5 (2013), 836–842. DOI:https://doi.org/10.1136/amiajnl-2013-001622

[136] Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. Publicly available clinical BERT embeddings. In Proceedings of the 2nd Clinical Natural Language Processing Workshop. Association for Computational Linguistics, Minneapolis, Minnesota, 72–78. DOI:https://doi.org/ 10.18653/v1/W19-1909

[137] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. Scientific Data 3, 1 (2016), 160035. DOI:https://doi.org/10.1038/sdata.2016.35

[138] Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2020. A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge Data Engineering 01 (2020), 1–1. DOI:https://doi.org/10.1109/TKDE.2020.2981314

[139] James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Rob Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003. The TimeBank corpus. Proceedings of Corpus Linguistics (Jan 2003).

[140] Taylor Cassidy, Bill McDowell, Nathanael Chambers, and Steven Bethard. 2014. An annotation framework for dense event ordering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Baltimore, Maryland, 501–506. DOI:https://doi.org/10. 3115/v1/P14-2082

[141] Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard. 2014. Dense event ordering with a multipass architecture. Transactions of the Association for Computational Linguistics 2 (2014), 273–284. DOI:https://doi. org/10.1162/tacl_a_00182 [142] Siddharth Vashishtha, Benjamin Van Durme, and Aaron Steven White. 2019. Fine-grained temporal relation extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, 2906–2919. DOI: https://doi.org/10.18653/v1/P19-1280

[143] Rujun Han, I-Hung Hsu, Mu Yang, Aram Galstyan, Ralph Weischedel, and Nanyun Peng. 2019. Deep structured neural network for event temporal relation extraction. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). Association for Computational Linguistics, Hong Kong, China, 666–676. DOI:https://doi.org/10.18653/v1/K19-1062

[144] Qiang Ning, Sanjay Subramanian, and Dan Roth. 2019. An improved neural baseline for temporal relation extraction. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'19). Association for Computational Linguistics, Hong Kong, China, 6203–6209. DOI:https://doi.org/10.18653/v1/D19-1642

[145] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics, New Orleans, Louisiana, 2227–2237. DOI:https://doi.org/10.18653/v1/N18-1202

[146] Greg Irving, Ana Luisa Neves, Hajira Dambha-Miller, Ai Oishi, Hiroko Tagashira, Anistasiya Verho, and John Holden. 2017. International variations in primary care physician consultation time: A systematic review of 67 countries. BMJ Open 7, 10 (2017). DOI:https://doi.org/10.1136/bmjopen-2017-017902

[147] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. (2019). arXiv:1910.01108. Retrieved from https://arxiv.org/abs/1910.01108. ACM Computing Surveys, Vol. 54, No. 7, Article 144. Publication date: September 2021. 144:36 Y. B. Gumiel et al.

[148] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. (2019). arXiv:1907.11692. Retrieved from https://arxiv.org/abs/1907.11692.

[149] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. In Advances in Neural Information Processing Systems, 32. Curran Associates, Inc