Detecting Fall Incidents in Clinical Notes: Vocabulary-Based Approach

Dganit Jenia Kim
Computational Lexicology and Terminology Lab (CLTL)
Vrije Universiteit Amsterdam
jenka@protonmail.com

Abstract

The experiment described in this paper is part of a text mining project aiming to detect in-hospital fall incidents in the free text of medical records in Dutch. A search method based on a hand-crafted vocabulary and simple post-processing rules is used to detect mentions of falls (of any type) in a random sample of clinical notes. The method is quite sensitive (recall 0.72), but its precision is low (0.35-0.39). Nonetheless, the method can be utilized in the next steps of the project to pre-select a fall-biased sample of clinical notes. This will significantly reduce the annotation effort, allowing to create a gold-label dataset of fall incidents typology that can be used to train machine learning algorithms.

1 Introduction

In-hospital falls are a major clinical, legal and regulatory problem. They are among the most common adverse events in hospitals, ranging from 3 to 12 falls per 1,000 patient-days (see Hempel et al., 2013, Hitcho et al., 2004, Shiner et al., 2016 and references therein). Approximately 30% of inpatient falls result in injury, with 4% to 6% resulting in serious injury, including fractures, excessive bleeding and even death (Hitcho et al., 2004).

Consequently, fall prevention is a significant concern for hospitals; for example, Joint Commission International (JCI)[1] defines fall prevention as one of its six International Patient Safety Goals. To establish an effective prevention program, healthcare staff relies on measurement of inpatient fall incidents. However, falls are difficult to reliably measure since they are often under-reported in incident reporting systems and/or misclassified (Simon et al., 2013, Shiner et al., 2016).

To overcome this inadequacy of the structured data, an additional source of information can be explored: the free text of electronic medical records. Clinical documents, such as patient progress reports, nurse notes and discharge summaries, might mention falls that have not been officially reported or coded into the structured data. Identifying these unreported incidents can therefore provide additional valuable insight into the scope and the characteristics of the phenomenon.

The experiment described in this paper is part of a text mining project conducted at Amsterdam UMC, which aims to detect in-hospital fall incidents in the free text of medical records in Dutch. A vocabulary of words describing a falling event was constructed based on interviews with nurses, descriptions of falls in the incident reporting system, and lexical enrichment from a dictionary. Using regular expressions, vocabulary matches were detected in a random sample of clinical notes. The notes were then annotated by a team of healthcare professionals who indicated whether a fall is mentioned in the note, and if so, classified it according to its location (in or out of the hospital) and consequence (injurious or not).

The main goals of the experiment were (a) to check how accurately can mentions of falls (of any type) be detected using a small vocabulary and simple post-processing rules, and (b) to provide the project team with insights regarding the data, such as how prevalent mentions of falls are in various types of clinical notes and how many of these falls are in-hospital incidents. The results of the experiment are meant to inform the next steps of the project.

Section 2 offers a short overview of existing work on detection of fall incidents in clinical text. Section 3 presents the dataset used for the exper-

[1] www.jointcommissioninternational.org

iment. Section 4 summarizes the results of the annotation effort. Sections 5 and 6 focus on the method and the results of the experiment. Section 7 provides discussion of the results and recommendations for the next steps of the project.

2 Related Work

Extraction of information from the free text of medical records is a fertile research discipline, with numerous applications, including detection of specific clinical conditions (see Ford et al., 2016 for an overview), risk factors (e.g. Jonnagaddala et al., 2015), adverse drug events (e.g. Eriksson et al., 2013) and many others.

Clinical text mining has also been applied to detection of falls and fall-related injuries. In a series of studies performed on medical records of the US Veterans Health Administration (VHA), Tremblay et al. (2009), McCart et al. (2012) and Luther et al. (2015) successfully train various supervised machine learning models to detect clinical notes containing injurious falls. For this approach, a large training set is needed in which documents are labeled as containing / not-containing fall incidents; the training set is usually obtained through manual annotation by human reviewers. While annotation is a very labor-intensive task in general, it is especially challenging in the context of clinical text, since it often requires medical expertise and therefore must be done by healthcare professionals, who have little time to spare. To facilitate an effective annotation process, the set used for annotation is often not a randomly selected sample of clinical notes, but rather a biased sample, in which a reasonable percentage of relevant documents is ensured. In the VHA studies, this is achieved by utilizing relevant ICD codes.[2] For example, McCart et al. (2012) create the dataset for their study by selecting patients who have an injury code as the primary reason for the ambulatory visit (ICD-9-CM 800-999) and who additionally have a fall-related E-code (ICD-9-CM E880-889).[3] They then match these patients' records with comparable control records that are not likely to contain falls, i.e. patients of the same age and gender who are treated for a similar, but not fall-related, injury. This carefully selected dataset is then annotated (fall / no-fall) and used for training and testing machine learning models.

This approach is suitable for the VHA studies since the main goal of their line of research was to automatically predict, based on the textual description, which records should be coded as fall-related injuries. The motivation behind this research direction is to alleviate the administrative burden of coding and to reduce underreporting and misclassification. However, the approach is not suitable for the current project. First, ICD codes focus on injurious falls only, while our project is interested in all falls, including those that did not result in an injury. Second, ICD codes are mainly used to reflect the primary reason of the visit / hospitalization, meaning that they point to falls that occurred outside of the hospital. Our project, on the other hand, is interested specifically in in-hospital falls, occurring among patients who might have been admitted for a different reason (i.e. might have a primary diagnosis code that is not injury-related). In other words, the falling incidents that the current project aims to detect are not likely to be reflected in the ICD codes; moreover, there is no reason to assume at this stage of the project that certain document types or patients are more likely than others to be involved in in-hospital falls. Therefore, pre-selection of data is not desirable for our purposes; rather, we choose to work with a random sample of patients and all available types of clinical notes, including progress reports, nurse notes, discharge summaries and others (see Section 3).

There are two previous studies that attempt to detect falls in a random sample of clinical notes: Toyabe (2012) and Shiner et al. (2016). Toyabe (2012) utilizes 170 manually constructed rules to detect injurious in-hospital fall incidents. Each of the rules consists of a set of Japanese morphemes and syntactic / semantic relationships between them.[4] The rules, which were created based on analysis of incident reports, perform well when applied to certain types of text, specifically in-

[2] The International Classification of Diseases (ICD) is the standard diagnostic tool for epidemiology, health management and clinical purposes, managed by the WHO. See the current version at http://apps.who.int/classifications/icd10/browse/2016/en

[3] E-codes are secondary (and usually non-mandatory) codes that describe the external causes of injuries. For example, E880 refers to "Accidental fall on or from stairs or steps".

[4] For example, the rule [*"sasaeru" (support someone)* + *"nai" (negation)* ⇒ *"tentoh" (fall)*] consists of three morphemes. The sign + indicates a syntactic dependency relationship, and the sign ⇒ indicates a causal or temporal relationship. This rule detects descriptions of a situation in which someone cannot support a staggering patient and the patient consequently falls down.

cident reports (with F-measure values between 0.84 and 1.0) and image order forms (F=0.91). They perform poorly, however, when applied to progress notes (F=0.12) and discharge summaries (F=0.24), suggesting that they are over-fitted for specific, possibly more standard, text type.

Shiner et al. (2016) train a supervised machine learning algorithm on a set of 2,730 progress notes that were annotated as containing / not containing a fall description. The dataset consists of 3 days of progress notes from a random sample of 100 inpatients. The authors use the Automated Retrieval Console (ARC, see D'avolio et al., 2010), which iteratively calculates performance of combinations of various NLP features and supervised classification algorithms. ARC is developed specifically for retrieving information from clinical text (see D'avolio et al., 2010); it processes text to more than 98 different data types, including mapping text to unique medical concepts in the National Library of Medicine's Unified Medical Language System (UMLS).[5] The best performing algorithm in Shiner et al. (2016)'s study reached 0.64 recall, 0.75 precision, and an F-score of 0.67. The most useful features for classification were UMLS unique concepts and noun phrases. This study shows that supervised machine learning can be successfully applied to detection of falls in a similar dataset to ours (in terms of document and patient composition, and size). However, it is important to note that the study relies heavily on English-language resources that are not available for Dutch. Specifically, the most influential features for classification were UMLS unique concepts, which are derived from mapping phrases in the text into a controlled vocabulary. This provides a single representation to phrases that are synonymous in a medical (and general) context, e.g. "fall prevention", "preventing falls" and "falling prevention" are mapped to the same concept. Since there is no comparable tool for Dutch at this time, extensive pre-processing is needed to obtain similar features for our study. Therefore, it was decided to experiment with a simpler method first: a search algorithm based on a simple hand-crafted vocabulary.

Vocabulary-based approaches to clinical text mining have been shown to be successful for various tasks. For example, Eriksson et al. (2013) use such approach to identify adverse drug events in the free text of electronic patient records in Danish. They construct a vocabulary of adverse drug events in Danish (based on summaries of product characteristics (SPCs) of 7,446 drugs), use a named-entity-recognition tagger to identify vocabulary matches in the text, and apply post-processing rules to filter-out, for example, negations and sentences about subjects other than the patient. Their method manages to detect adverse drug events with precision of 0.89 and recall of 0.75. In this paper, a comparable method is applied, as described in Sections 5 and 6.

3 Data

The data available for this experiment is an export from Epic, the electronic medical record application used in the hospital since March 2016. The dataset contains the records of 8,885 patients (both inpatients and outpatients) who have at some point had contact with the rehabilitation department of the hospital.[6] The medical record of each of these patients contains various clinical notes; the number varies from 1 to 1,051 notes per patient (median: 11, mean: 34). In total, the dataset contains 305,242 clinical notes of different types, including progress reports (*'voortgangsverslag'*, 45% of all notes), nurse notes (*'zorgplan/vpk rapportage'*, 26% of all notes), summaries of telephone calls (*'telefonisch contact'*, 8% of all notes) etc. The notes date from July 2015 to May 2017.[7]

4 Annotation

A subset of the data described in Section 3 was manually annotated by a team of healthcare professionals. This was done in order to evaluate the vocabulary-based detection method and provide the project team with initial insights about the prevalence of different types of falls in the clinical notes.

The annotation was done in two rounds: in the first round, 1,980 randomly selected notes were annotated; in the second round, 1,193 notes

[5] https://www.nlm.nih.gov/research/umls/quickstart.html

[6] This is not an ideal dataset for drawing general conclusions, as various biases might arise from the fact that it involves only rehabilitation patients. However, at this point this was the only dataset available to the project team. The next steps of the project will be performed on a bigger dataset that represents the whole patient population of the hospital.

[7] As mentioned above, Epic is in use at the hospital only from May 2016; the earlier-dated documents were probably entered into the system retrospectively.

Sample 1	Positive	Negative	Total
No. of notes	40	1,940	1,980
%	2%	98%	
Sample 2	**Positive**	**Negative**	**Total**
No. of notes	176	1,017	1,193
%	15%	85%	

Table 1: Overview of the annotation results

Location	Injury	%	abs
In-hospital	Totals	8.8	26
	- injury	0.3	1
	- no injury	5.1	15
	- unspecified	3.4	10
Out-hospital	Totals	86.4	254
	- injury	44.6	131
	- no injury	8.2	24
	- unspecified	33.7	99
Unspecified	Totals	4.8	14
	- injury	1.4	4
	- no injury	0.3	1
	- unspecified	3.1	9
Totals	Totals	100.0	294
	- injury	46.3	136
	- no injury	13.6	40
	- unspecified	40.1	118

Table 2: Falls by type in the annotated sample

that were suspected as containing falls were annotated. The positively-biased sample used for the second annotation round was compiled using the vocabulary-based search (see Section 6.1); the search was performed on a random sample of 10,000 notes and the 1,193 notes that contained vocabulary items were sent to annotation. This was done for practical consideration; the annotators had limited time and we needed as many as possible positive gold-label examples in order to perform meaningful analysis.

The first annotation round was performed by a team of 14 medical experts from the hospital (incl. physiotherapists, nurses, doctors, etc.); each of them annotated a sample of 50 to 200 notes. The second round was performed by 8 medical experts (a sub-group of the initial 14), each of whom annotated a sample of about 150 notes.[8]

The annotators were instructed to mark the piece of text that indicates that a fall event has happened; they were encouraged to mark as much text as they think necessary to describe the fall event. In addition, each marked text segment had to be classified into one of the 9 categories shown in Table 2, i.e. according to the location of the falling event and whether it lead to an injury.

Table 1 shows an overview of the annotation results. In the random Sample 1, 2% of the notes contain falls; in the positively-biased Sample 2, 15% of the notes contain falls. In total, 294 segments mentioning falls were marked by the annotators (it is more than the number of positive notes since one note may contain more than one mention of falling). Of these 294 falls, 9% took place in the hospital, 86% happened outside of the hospital, and for the rest the location was unspecified (see Table 2).

[8]In a setting with multiple annotators, it is desirable to check for inter-annotator agreement (i.e. let different people annotate the same subset of notes and check whether they come to the same result). For the current experiment, it had been decided not to do this because the annotators time was very scarce.

Mentions of falls tend to appear in specific types of clinical notes. There are 36 different types of notes in the annotated sample, but 97% of all fall mentions are found in 7 specific note types: progress reports (*'voortgangsverslag'*, 53% of fall mentions), consultations (*'consulten'*, 13% of fall mentions), letters (*'brief'*, 9% of fall mentions), ER doctor notes (*'SEH arts notitie'*, 7% of fall mentions), nurse notes (*'zorgplan/vpk rapportage'*, 7% of fall mentions), paramedic notes (*'consultnotitie paramedici'*, 4% of fall mentions), discharge summaries (*'ontslagsamenvattingen'*, 3% of fall mentions).

The 26 in-hospital falls appear in three types of notes: progress reports (*'voortgangsverslag'*, 62% of in-hospital fall mentions), nurse notes (*'zorgplan/vpk rapportage'*, 31% of in-hospital fall mentions), consultations (*'consulten'*, 7% of in-hospital fall mentions).

The numbers of in-hospital falls in general, and injurious in-hospital falls in particular, are lower in the annotated sample than what is reported in the literature (see Section 1). The sample is too small, however, to draw any conclusions at this point.

5 Method

5.1 Vocabulary

There are numerous ways to describe a falling event in free text. In addition to a variety of applicable verbs and nominalizations (e.g. fall, stumble, slip, collapse, etc.), indirect descriptions might also be used (e.g. 'I heard a loud clap; when I entered the room, I saw the patient on the floor near the bed.'). Our goal was to create a vocabulary that captures a wide variety of descriptions and phrasings. For this purpose we utilized two main sources: (a) descriptions of fall incidents that were reported in the hospital's incident reporting system (DIM - *Decentraal Incidenten Melden*), (b) interviews with nurses. The vocabulary was further enriched with synonyms and hyponyms of the words obtained from (a) and (b).

DIM data: We had access to all incidents labeled as 'Fall' that were reported in the hospital's incident reporting system between April 2010 and October 2017; 609 unique records in total. The records include a field for 'cause of incident', where one can either choose from a standard list of causes (e.g. *'motorische onrust patiënt'* - 'motorically restless patient') or enter free text. All free text entries were manually examined; 37 of them contained descriptions of the falling event and were therefore selected for further processing. An example of such a description is *'dhr lag te dicht bij het randje, bedhekken waren niet omhoog, is tijdens het slapen uit bed gegleden'* ('Mr was lying too close to the edge, side rails were not up, slid out of bed while sleeping'). In this case, the phrase describing the falling event is 'slid out of bed while sleeping' and the cue that most directly indicates falling is the verb *'glijden'* ('to slide, to slip'). This "fall cue" is added to the vocabulary.[9]

Interviews: A number of interviews with nurses working at the hospital were conducted by a master's student of medical informatics.[10] During an interview, the nurse was presented with scenarios describing falls; after each scenario, s/he was asked to answer a few questions about it (see Appendix D for the full list of the materials used in the interviews). In one of the questions the nurse was asked to simulate entering a note about the scenario in the electronic medical record of the patient; s/he was requested to type the note on the interviewer's laptop. The purpose of typing the answer was to encourage a more realistic description, including abbreviations, typos, etc. The interviews provided 22 descriptions of falling events. These descriptions were processed similarly to the DIM data; the words and phrases identified as the main linguistic cue that indicates falling were added to the vocabulary.

Lexical augmentation: The vocabulary was then augmented by systematically adding the synonyms and hyponyms of the "fall cues" collected from the interviews and DIM. The procedure was as follows: (a) each word was looked-up in the online *Van Dale Hedendaags Nederlands* dictionary[11], (b) the relevant senses were selected from all available senses of the word, (c) the synonyms and hyponyms of the relevant senses were added to the vocabulary (if not already there).

The procedure described above resulted in a list of 71 vocabulary items; the full list of the items is found in Appendix A. Of these 71 items, 60% are verbs (e.g. *'vallen'* - 'to fall', *'glijden'* - 'to slide, to slip'), 32% are verbal phrases (e.g. *'evenwicht verliezen'* - 'to lose balance', *'klap horen'* - 'to hear a clap'), and 8% are nouns or nominalizations (e.g. *'val'* - 'fall'). 18 out of the 71 items (25%) originate from the interviews and/or DIM, and the rest are synonyms and hyponyms. The relevant grammatical inflections of the 71 items were added to the vocabulary as well; for details about this procedure, see Appendix B. This resulted in a total of 353 inflected forms.

To sum up, the vocabulary was very carefully and manually crafted. It was decided to do it in this way is order to minimize noise and mistakes in this initial phase of the project. The manual procedure was feasible since the vocabulary is rather small; for a bigger vocabulary, it would be reasonable to exploit automated resources (e.g. lemmatize instead of adding inflected forms, automatically add synonyms and hyponyms with WordNet, etc.).

5.2 Regular Expressions

To perform the search, vocabulary items were converted into three types of regular expressions (regex):

- For one-word vocabulary items, a simple regex of the item surrounded by word boundaries was used: `\b<text>\b`. For example, the regex `\bgevallen\b` detects *'Zij*

[9]During the analysis of the the descriptions, additional information was extracted and stored, such as the cause of the fall (e.g. 'side rails not up'), locations (e.g. 'in the bathroom'), etc. For the current experiment, only the "fall cues" were utilized; the additional information might prove useful for the next steps of the project.

[10]Steyn Kahrel, s.k.kahrel@amc.uva.nl

[11]The online dictionary is not publicly available; the resource was accessed through the university license.

*is **gevallen**.'* ('She has **fallen**.') because the text *gevallen* in this sentence is between two word boundaries - a space and a punctuation mark.

- For two-word items, the regex (\b<text1>\b[^.]*?\b<text2>\b|\b<text2>\b[^.]*?\b<text1>\b) was used. It allows the two words of the item to appear in different order and to have any character except for a full stop in between them.[12] In other words, it detects the cases where the two words appear in the same sentence, regardless of their order and whether there are additional words in between them. For example, *'Hij **verloor** plotseling zijn **evenwicht**.'* ('He suddenly lost his balance.') and *'Hij zegt dat hij zijn **evenwicht** plotseling **verloor**.'* ('He says that he suddenly lost his balance.').

- For three-word items, the various word orders were incorporated as separate items in the dictionary (to avoid the need for a very complex regex). The regex that was used is: \b<text1>\b[^.]*?\b<text2>\b[^.]*?\b<text3>\b. This setup detects the cases where the three words appear in the same sentence, regardless of their order and whether there are additional words in between them. For example, *'Zij **ging** hard **tegen** de **grond**'* ('She fell hard on the floor.') and *'Zij zegt dat zij **tegen** de **grond** **ging**.'* ('She says that she fell on the floor').

Each vocabulary item was converted into a regular expression of the correct type, according to the number of words it comprises. This resulted in 353 regex's that were used in the experiments described in the following section.

6 Experiments

6.1 Experiment 1: full vocabulary

In Experiment 1, a search based on the full vocabulary described in Section 5.1 (71 items, 353 inflected forms) was performed on two randomly selected samples of 2,000 and 10,000 notes. The notes were searched for matches to the 353 regular expressions described in Section 5.2; any note containing one or more matches was marked as positive. In both samples, 12% of the notes were marked positive by the search procedure. The results were then evaluated using the gold-label annotated data described in Section 4. As shown in Table 3, the recall of the method is pretty high: 82% of the notes containing falls in Sample 1 were correctly retrieved by the method.[13] However, the precision of the method is low: only 14-15% of the notes detected by the method indeed contain falls. In other words, most of the notes retrieved by the search were false positives.

6.2 Experiment 2: modified vocabulary

Next, the individual contribution of the various vocabulary items was analyzed. Of the 71 vocabulary items, 34 were found in the clinical notes. However, about half of these 34 items were "doing more harm than good", i.e. most of their matches were false positives. For example, the verb *'zakken'* ('to fall, to drop, to sink, to fail') with its various inflections was found 257 times in the clinical notes; however, only 10% of these matches were in true positive notes. Based on that, it was decided to remove from the vocabulary the 19 items for which less than 20% of the matches were true positives (for the list of the removed items, see Table 5 in Appendix A). The results are shown in Table 3 as Experiment 2; this modification significantly improves the precision (from 14-15% to 23-24%) without impacting the recall.

6.3 Experiment 3: post-processing rules

The next step was to manually inspect a sample of about 30 randomly selected false positives. The analysis revealed three main types of false positives in the sample:

- other meanings of the verb *'vallen'* ('to fall'), such as *'in slaap vallen'* ('to fall asleep') and *'afvallen'* ('to lose weight')[14],

- mentions of habitual falls rather than specific incidents, e.g. *'Kijkt niet uit, **valt** veel, kijkt niet gericht.'* ('Doesn't watch out, falls a lot, unfocused gaze.'),

[12]The *? quantifier means that there can be either zero or more characters between the words and that the shortest match should be considered (i.e. non-greediness).

[13]In Sample 2, the suspected negatives were not annotated (see Section 4), therefore it is unknown how many of them are true and how many are false, and recall cannot be calculated.

[14]The verb *'afvallen'* is detected by the search method because it has a separable prefix, as in the sentence *'Mw voelt zich goed, **valt** goed **af** richting streef gewicht.'* ('Ms. feels good, loses weight towards the target weight.'). In this case the form *valt* is detected since it means 'falls'. More on verbs with a separable prefix in Appendix B.

		FN	TN	FP	TP	Precision	Recall
Experiment 1	Sample 1	7	1758	202	33	0.14	0.82
	Sample 2	n/a	n/a	1017	176	0.15	n/a
Experiment 2	Sample 1	7	1847	113	33	0.23	0.82
	Sample 2	n/a	n/a	544	173	0.24	n/a
Experiment 3	Sample 1	11	1915	45	29	0.39	0.72
	Sample 2	n/a	n/a	244	134	0.35	n/a

Table 3: Evaluation of the three experiments

- negated mentions of falls, e.g. *'Is niet duizelig, **valt** niet.'* ('Not dizzy, doesn't fall.').

Based on these findings, a set of post-processing rules was developed; the rules filter out from the search results these three types of false positives. The full list of words and phrases used for the post-processing rules is found in Appendix C. The post-processing boils down to checking whether one of the words indicating a false positive (e.g. 'not') is present in the same sentence with a vocabulary item. If this is the case, the original match is removed from the search results. For example, the sentence *'Bij het lopen sleept hij met zijn voeten, waardoor hij regelmatig **valt** en **struikelt**.'* ('He drags his feet when walking, which causes him to stumble and fall regularly.') is a false positive since it does not describe a specific fall incident.[15] It is detected by the search because there are two vocabulary items in it: *'valt'* ('falls') and *'struikelt'* ('stumbles'). The post processing rule identifies that the word *regelmatig* ('regularly'), which is in the post-processing vocabulary (Appendix C) is also in the sentence; this removes the sentence from the list of search results (and if it is the only match in the note, it removes the note from the list of positive notes).

Clearly, this first version of the rules is very simple and prone to inaccuracies; for example, if a word like 'regularly' or 'not' appears in a sentence, it does not necessarily relate to the falling event. Therefore, the method impacts the recall, which drops from 0.82 to 0.72 (see Table 3). In other words, some true positives are filtered out by this coarse method. However, it does manage to filter out many false positives as well, which improves the precision from 0.23-0.24 to 0.35-0.39.

[15]However, see Section 7 about inconsistencies in annotation for this type of cases.

7 Discussion

The experiments reported in the previous section and the annotation effort performed to evaluate them provide valuable insights about the task at hand in general, and the next steps of the project in particular.

Since the project aims to detect *in-hospital* fall incidents, an important finding is that less than 10% of the 294 fall mentions detected in the annotated sample happened inside the hospital. In terms of notes, about 11% of the positive notes (0.22% of all notes) contain in-hospital falls. In terms of patients, 13% of the patients who fell (1% of all patients in the dataset) fell inside the hospital. These numbers make clear that annotation of random samples of notes is not a viable method to obtain gold-label data for the task at hand; to get 100 notes with in-hospital falls, one would have to annotate 50,000 clinical notes.

One way to assist the annotation effort is by preselecting a fall-biased sample using the method presented in this paper. In its current version, the method can retrieve between 72-82% of the notes that mention falls (of any type) from a random sample. In the retrieved sample, between 23-39% of the notes will contain mentions of falls (of any type). Using this method, one would have to annotate about 2,500 notes to get 100 notes with in-hospital falls. Although some information is inevitably lost in this way, since some positive notes are not retrieved by the method, its current recall is high enough for the purpose of creating a gold-label dataset that can be used to train machine learning algorithms.

Furthermore, the vocabulary-based search method can be improved in a number of ways, depending on its intended use. To improve the recall (i.e. make sure that the method misses less fall mentions), the vocabulary can be extended based on a false negatives analysis. In the current experiment, the suspected negatives were annotated only in Sample 1, where 7 notes were identified as false negatives. Analysis of these 7 notes yielded about 4 items that could be added to the vocabulary. Although it is not practical at this point to annotate the 8,807 suspected negative notes of Sample 2, a subset of them can be annotated to estimate the recall and perform a false negatives analysis.

To improve the method's precision (i.e. make sure that a bigger portion of the retrieved notes contain fall mentions), the post-processing rules can be refined. For example, the number of words that may appear between 'not' and the falling cue can be restricted to assure that the negation indeed relates to the falling event and not another element in the sentence.

Alternatively, the vocabulary-based method can be fine-tuned to focus specifically on in-hospital falls. This can be done by revisiting the data collected from the interviews with nurses and the DIM dataset and identifying the linguistic cues that are specific for in-hospital incidents (e.g. *'tijdens mobiliseren'* - 'during transfer', *'van de behandeltafel'* - 'from the treatment table').

Additional important insights gained from the current experiment relate to the annotation process. Crucially, it needs to be clearly defined what exactly is considered a falling event. For the annotation effort described in this paper, the WHO definition of fall was used, namely "an event which results in a person coming to rest inadvertently on the ground or floor or other lower level". An injurious fall was defined as a fall that has a consequence which requires medical attention. However, these definitions led to quite a few inconsistencies, both inter-annotator and intra-annotator. For example, habitual falls (e.g. *'<name> valt nog ongeveer 1x per week'* - '<name> falls about once a week') and falls that happened many years ago were sometimes marked as falls and sometimes not. Other controversial examples involved e.g. car accidents, where a bike-rider was hit by a car and consequently found himself on the ground. These types of borderline cases need to be discussed and the decisions need to be incorporated into the annotation guidelines. Interestingly, in the existing literature on similar annotation efforts (see Section 2), the exact definitions are not provided and potential issues are not mentioned.

8 Conclusion

Detecting in-hospital falls in the free text of medical records in Dutch is a novel and challenging task. In this paper, the first exploratory steps were taken towards the ultimate goal. The vocabulary-based search method described in this paper detects mentions of falls with a recall of 0.72 and precision of 0.39. It can be utilized to assist the annotation effort in the next steps of the project.

References

Leonard W D'avolio, Thien M Nguyen, Wildon R Farwell, Yongming Chen, Felicia Fitzmeyer, Owen M Harris, and Louis D Fiore. 2010. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (arc). *Journal of the American Medical Informatics Association*, 17(4):375–382.

Robert Eriksson, Peter Bjødstrup Jensen, Sune Frankild, Lars Juhl Jensen, and Søren Brunak. 2013. Dictionary construction and identification of possible adverse drug events in danish clinical narrative text. *Journal of the American Medical Informatics Association*, 20(5):947–953.

Elizabeth Ford, John A Carroll, Helen E Smith, Donia Scott, and Jackie A Cassell. 2016. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*, 23(5):1007–1015.

Susanne Hempel, Sydne Newberry, Zhen Wang, Marika Booth, Roberta Shanman, Breanne Johnsen, Victoria Shier, Debra Saliba, William D Spector, and David A Ganz. 2013. Hospital fall prevention: a systematic review of implementation, components, adherence, and effectiveness. *Journal of the American Geriatrics Society*, 61(4):483–494.

Eileen B Hitcho, Melissa J Krauss, Stanley Birge, William Claiborne Dunagan, Irene Fischer, Shirley Johnson, Patricia A Nast, Eileen Costantinou, and Victoria J Fraser. 2004. Characteristics and circumstances of falls in a hospital setting: a prospective analysis. *Journal of general internal medicine*, 19(7):732–739.

Jitendra Jonnagaddala, Siaw-Teng Liaw, Pradeep Ray, Manish Kumar, Nai-Wen Chang, and Hong-Jie Dai. 2015. Coronary artery disease risk assessment from unstructured electronic health records using text mining. *Journal of biomedical informatics*, 58:S203–S210.

Stephen L Luther, James A McCart, Donald J Berndt, Bridget Hahm, Dezon Finch, Jay Jarman, Philip R Foulis, William A Lapcevic, Robert R Campbell, Ronald I Shorr, et al. 2015. Improving identification of fall-related injuries in ambulatory care using statistical text mining. *American journal of public health*, 105(6):1168–1173.

James A McCart, Donald J Berndt, Jay Jarman, Dezon K Finch, and Stephen L Luther. 2012. Finding falls in ambulatory care clinical documents using statistical text mining. *Journal of the American Medical Informatics Association*, 20(5):906–914.

Brian Shiner, Julia Neily, Peter D Mills, and Bradley V Watts. 2016. Identification of inpatient falls using automated review of text-based medical records. *Journal of patient safety*.

Michael Simon, Susan Klaus, Byron J Gajewski, and Nancy Dunton. 2013. Agreement of fall classifications among staff in us hospitals. *Nursing research*, 62(2):74–81.

Shin-ichi Toyabe. 2012. Detecting inpatient falls by using natural language processing of electronic medical records. *BMC health services research*, 12(1):448.

Monica Chiarini Tremblay, Donald J Berndt, Stephen L Luther, Philip R Foulis, and Dustin D French. 2009. Identifying fall-related injuries: Text mining the electronic medical record. *Information Technology and Management*, 10(4):253.

A Vocabulary Items

Tables 4 and 5 contain the 71 vocabulary items described in Section 5.1. For each item, it is indicated whether it was collected from the DIM dataset/interviews with nurses or added from a dictionary as a synonym or hyponym of one of the collected items. The items in Table 5 were eventually removed from the final version of the vocabulary, as described in Section 6.

B Vocabulary Construction: Inflections

The items listed in Tables 4 and 5 are the infinitive form for verbs and the singular form for nouns. The relevant inflections of these base forms were also added to the vocabulary:

- For nouns and nominalizations, the plural form was added;
- For verbs and verbal phrases whose subject is the falling person (e.g. stumble, lose balance), the 3rd person singular forms in all tenses were added. In Dutch, this includes
  - past participle (*'hij/zij is **gevallen**'*-'he/she has **fallen**'),
  - past simple (*'hij/zij **viel**'*-'he/she **fell**'),
  - present simple (*'hij/zij **valt**'*-'he/she **falls**').
- For verbal phrases whose subject is someone who witnesses a fall (e.g. hear a clap, find on the floor), both 1st and 3rd person singular and plural forms in all tenses were added. In Dutch, this includes
  - past participle (*'ik/hij/zij/wij/zij heb/heeft/hebben **een klap gehoord**'*-'I/he/she/we/they has/have **heard a clap**'),
  - past simple (*'ik/hij/zij/wij/zij **hoorde/hoorden een klap**'*-'I/he/she/we/they **heard a clap**'),
  - present simple (*'ik/hij/zij/wij/zij **hoor/hoort/horen een klap**'*-'I/he/she/we/they **hear a clap**').

An additional consideration is related to verbs with a separable prefix. In main clauses, the prefix is separated from the verb (*'hij/zij **gleed uit**'*-'he/she **slipped**'), while in embedded clauses and past participle form, the prefix is glued to the verb (*'ik zag dat hij/zij **uitgleed**'*-'I saw that he/she

Table 4: Vocabulary items (final version, after removal)

Item	Source	Item	Source
vallen	interviews/DIM	omkukelen	dictionary
omvallen	interviews/DIM	lazeren	dictionary
flauwvallen	interviews/DIM	mieteren	dictionary
glijden	interviews/DIM	neerkletteren	dictionary
uitglijden	interviews/DIM	ploffen	dictionary
afglijden	interviews/DIM	sodemieteren	dictionary
wegglijden	interviews/DIM	kieperen	dictionary
struikelen	interviews/DIM	omkiepen	dictionary
klap maken	interviews/DIM	tuimelen	dictionary
evenwicht verliezen	interviews/DIM	omtuimelen	dictionary
klap horen	interviews/DIM	kwakken	dictionary
grond aantreffen	interviews/DIM	omkwakken	dictionary
collaps	interviews/DIM	bezwijmen	dictionary
val	interviews/DIM	schuiver maken	dictionary
doodsmak	dictionary	dreun maken	dictionary
schuiver	dictionary	bons maken	dictionary
smak	dictionary	smak maken	dictionary
struikeling	dictionary	evenwicht kwijtraken	dictionary
neervallen	dictionary	balans kwijtraken	dictionary
donderen	dictionary	equilibrium kwijtraken	dictionary
omdonderen	dictionary	equilibrium verliezen	dictionary
donderstralen	dictionary	vloer aantreffen	dictionary
duvelen	dictionary	vloer terugvinden	dictionary
flikkeren	dictionary	tegen grond gaan	dictionary
omflikkeren	dictionary	tegen vlakte gaan	dictionary
kukelen	dictionary	van stokje gaan	dictionary

Table 5: Vocabulary items removed in Experiment 2

Item	Source	Item	Source
collaberen	interviews/DIM	ineenzakken	dictionary
zakken	interviews/DIM	terugzakken	dictionary
rollen	interviews/DIM	wegzakken	dictionary
balans verliezen	interviews/DIM	onderuitgaan	dictionary
smakken	dictionary	storten	dictionary
neersmakken	dictionary	gestrekt gaan	dictionary
omgaan	dictionary	grond tegenkomen	dictionary
kantelen	dictionary	grond terugvinden	dictionary
omkantelen	dictionary	vloer tegenkomen	dictionary
omrollen	dictionary		

slipped'). The first case is captured by the main verb forms; i.e. since the verb *'glijden'* without prefixes (and its past simple form *'gleed'*) is also in the vocabulary, the example *'hij/zij **gleed uit**'* will be detected anyway. Therefore, the only forms that need to be added to the dictionary are the ones with the prefix attached (e.g. *'uitgleed'*).

C Post Processing Rules

Table 6 shows the items that were used for the post-processing rules described in Section 6.3. In the table, only the infinitive form of verbs and the singular form of nouns is shown; the relevant inflections (according to the procedure in Appendix B) were also used in the post processing rules. Most of the items come directly from the analysis of a sample of false positives described in Section 6.3. To these, some synonyms were added using the same procedure described in Section 5.1.

For the habitual adverbs and the words indicating negation, the post processing rule is: if one of them co-occurs in the same sentence with a vocabulary item (Table 4) ⇒ remove the match from the search results. For *(in) slaap vallen* ('to fall asleep') and *lastig vallen* ('to bother') the rule is: if both words (e.g. *'slaap'* and *'vallen'*) co-occur in the same sentence (no matter the order) ⇒ remove the match from the search results. For *'vallen op'* ('to strike'), *'vallen af'* ('to lose weight') and *'vallen voor'* ('to occur') the rule is: if both words (e.g. *'vallen'* and *'op'*) co-occur in the same sentence (in this order) ⇒ remove the match from the search results. The item *'Vallen - Johns Hopkins'* is part of a structured form that is used in some notes; the rule for it is: remove this match from the search results.

D Interview Materials

Table 8 shows the 14 scenarios used in the interviews described in Section 5.1. The scenarios were written and categorized by healthcare professionals from the project team; they are meant to reflect common fall-related situations of different types. The situations include falls in the hospital and outside of it, as well as situations where it is unspecified where the fall occurred and situations where there is no fall (marked 'n/a'). Some of the scenarios involve an injury while others do not; for some situations it is unspecified whether there was an injury. For the vocabulary construction, the responses to scenarios 1-9 were used (i.e. all scenarios that involve a falling event); the responses to the non-fall scenarios were not utilized in the current experiment.

Table 7 shows the questions the nurses had to answer after they were presented with a scenario. For the vocabulary construction, the responses to the second question (i.e. 'How would you describe the incident in your report?') were used. The responses to the other questions were not utilized in the current experiment.

Table 6: Items for post-processing rules

Group	Item	Source
Habitual adverbs	vaak, veel, meestal, wel eens, soms, regelmatig	false positives sample
	weleens, dikwijls, geregeld	dictionary
Negation	geen, dreigen, niet, bang om, bijna, risico	false positives sample
'Vallen'	vallen op, vallen af, vallen voor, slaap vallen, lastig vallen, Vallen - Johns Hopkins	false positives sample

Table 7: Questions used in interviews with nurses

Original Dutch text	English translation
Waar zou u dit incident/deze situatie noteren? In DIM of in Epic? Waarom?	Where would you report this incident/situation? In DIM or in Epic? Why?
Hoe zou u de situatie omschrijven in uw rapportage?	How would you describe the incident in your report?
Hoe zou u de oorzaak en/of gevolgen van die situatie omschrijven?	How would you describe the cause and/or consequences of the situation?
Hoe zou u de acties van u of uw collegas in woord brengen bij een dergelijk incident?	How would you describe the actions of you or your coworkers in such a situation?
Zijn er nog andere dingen die u zou noteren in een dergelijke situatie?	Are there any additional notes you would write down in such a situation?

Table 8: Scenarios used in interviews with nurses

Hospital (in/out)	Injury (yes/no)	Original Dutch text	English translation
out	no	U ziet een blauwe plek bij een oudere patiënt. Ze vertelt dat ze thuis gevallen is waardoor u zich zorgen maakt dat dit vaker kan voorkomen.	You see a bruise on an elderly patient. She tells you that she fell at home; you are worried she might fall again.
in	no	U ziet een oudere patiënt uitglijden bij een pas gedweilde vloer. De patiënt heeft geen zichtbare verwondingen maar loopt wel wat mank.	You see a patient slip and fall on a recently mopped floor. The patient has no visible injuries but does limp a bit.
in	yes	Een 35-jarige patiënt struikelt in zijn kamer en loopt een schaafwond op. De wond is niet ernstig maar moet wel behandeld worden.	A 35-year old patient trips over in his room and consequently has a scrape. The wound is not serious but does require treatment.
in	no	Een patiënt valt na een operatie. De patiënt loopt geen verwondingen op maar is wel aangeslagen.	A patient falls after undergoing surgery. The patient did not suffer any injuries but is visibly shook.
unspecified	unspecified	Een patiënt die medicatie slikt waarvan bekend is dat het duizeligheid kan veroorzaken valt bij het lopen naar de wc.	A patient that takes medication known to cause dizziness falls on the way to the bathroom.
unspecified	unspecified	Een oudere patiënt die een rollator nodig heeft om te lopen valt met de rollator.	An elderly patient who needs a walker falls while walking with the walker.
unspecified	yes	Een oudere patiënt valt in de badkamer en breekt haar heup.	An elderly patient falls in the bathroom and fractures her hip.
in	unspecified	Een patiënt zakt plots in elkaar in zijn/haar kamer.	A patient suddenly collapses in his/her room.
in	unspecified	Een patiënt wordt glijdt tijdens een gesprek plots van zijn/haar stoel.	A patient suddenly falls out of his/her chair during a conversation.
n/a	n/a	Een patiënt valt terug op de stoel bij het opstaan.	A patient falls back into his/her chair while standing up.
n/a	n/a	U ziet een patiënt wankel lopen op de afdeling.	You see a patient walk unsteadily at the department.
n/a	n/a	U ziet een patiënt struikelen maar niet vallen.	You see a patient trip but not fall.
n/a	n/a	Een patiënt heeft veel pijn in de benen en u bent bang dat de patiënt hierdoor zou kunnen vallen.	A patient has a lot of pain in his/her legs which causes you to worry that the patient might fall.
n/a	n/a	Een patiënt heeft last van uitvalsverschijnselen waardoor u zich zorgen maakt dat de patiënt mogelijk kan vallen.	A patient suffers from cognitive loss which causes you to worry the patient might fall.