International Journal of Medical Informatics 112 (2018) 149–157

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

# Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: A feasibility study

Sheng-Feng Sung[a,b,c], Kuanchin Chen[d], Darren Philbert Wu[a], Ling-Chien Hung[a], Yu-Hsiang Su[a], Ya-Han Hu[b,*]

[a] Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi City, Taiwan
[b] Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan
[c] Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan
[d] Department of Business Information Systems, Western Michigan University, Kalamazoo, MI, United States

ARTICLE INFO

*Keywords:*
Acute ischemic stroke
Electronic medical record
Intravenous thrombolysis
Natural language processing

ABSTRACT

*Objective:* To reduce errors in determining eligibility for intravenous thrombolytic therapy (IVT) in stroke patients through use of an enhanced task-specific electronic medical record (EMR) interface powered by natural language processing (NLP) techniques.

*Materials and methods:* The information processing algorithm utilized MetaMap to extract medical concepts from IVT eligibility criteria and expanded the concepts using the Unified Medical Language System Metathesaurus. Concepts identified from clinical notes by MetaMap were compared to those from IVT eligibility criteria. The task-specific EMR interface displays IVT-relevant information by highlighting phrases that contain matched concepts. Clinical usability was assessed with clinicians staffing the acute stroke team by comparing user performance while using the task-specific and the current EMR interfaces.

*Results:* The algorithm identified IVT-relevant concepts with micro-averaged precisions, recalls, and F1 measures of 0.998, 0.812, and 0.895 at the phrase level and of 1, 0.972, and 0.986 at the document level. Users using the task-specific interface achieved a higher accuracy score than those using the current interface (91% versus 80%, $p = 0.016$) in assessing the IVT eligibility criteria. The completion time between the interfaces was statistically similar (2.46 min versus 1.70 min, $p = 0.754$).

*Discussion:* Although the information processing algorithm had room for improvement, the task-specific EMR interface significantly reduced errors in assessing IVT eligibility criteria.

*Conclusion:* The study findings provide evidence to support an NLP enhanced EMR system to facilitate IVT decision-making by presenting meaningful and timely information to clinicians, thereby offering a new avenue for improvements in acute stroke care.

## 1. Introduction

Stroke is one of the leading causes of death and disability, and places a huge economic burden on healthcare systems worldwide [1]. Of all strokes, about 75% to 90% are ischemic strokes [2]. Currently, intravenous thrombolysis (IVT) is a standard treatment for acute ischemic stroke (AIS) [3], significantly reducing the chance of disability following stroke if patients are treated within 4.5 h of stroke onset [4,5]. Because the effect of IVT on functional outcomes is time-dependent, with better outcomes being associated with earlier treatment [6,7], patients with AIS should be evaluated and treated as soon as possible upon arrival at the emergency department (ED). Various strategies, such as hospital pre-notification, prompt stroke team notification, rapid evaluation, and pre-acquisition of history, have been used to expedite the administration of IVT [8,9].

However, IVT is not without side effects, of which symptomatic intracranial hemorrhage (SICH) is the most serious and life-threatening. SICH potentially increases the risk of poor and fatal outcomes [10], and might consequently reduce the net benefit from IVT. Worst of all, such hemorrhagic complications may lead to litigation and malpractice

*Abbreviations:* AIS, acute ischemic stroke; CUI, concept unique identifier; ED, emergency department; EMR, electronic medical record; IVT, intravenous thrombolytic therapy; NLP, natural language processing; SICH, symptomatic intracranial hemorrhage; SUS, System Usability Scale; UMLS, Unified Medical Language System

* Corresponding author at: Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, 168 University Road, Min-Hsiung, Chiayi County 62102, Taiwan.
*E-mail address:* yahan.hu@mis.ccu.edu.tw (Y.-H. Hu).

https://doi.org/10.1016/j.ijmedinf.2018.02.005
Received 7 November 2017; Received in revised form 31 January 2018; Accepted 4 February 2018
1386-5056/ 

claims [11]. Many factors, including older age, greater stroke severity, and hyperglycemia, increase the risk of SICH [12]. Therefore, stroke guidelines generally specify eligibility criteria for patients who can be treated with IVT [3,13], and violations of the treatment protocol were found to increase the risk of SICH [14]. Although most of the eligibility criteria for IVT simply involve medical history, it is challenging for physicians to accurately analyze a patient's medical history within a very short time [8]. Additionally, patient self-report history is known to be subject to errors [15], especially for the elderly [16], not to mention those with aphasia or consciousness disturbance.

With the design of the Taiwan Electronic Medical Record Template [17], electronic medical records (EMRs) have replaced traditional paper records in many hospitals in Taiwan. EMRs typically contain clinical notes, laboratory results, radiology and pathology records, and prescriptions. Because the EMR interface is where interactions between clinicians and the EMR system occur, an ideal design of the EMR interface should reduce errors and support clinicians [18]. Although previous studies have indicated that EMR use improved quality of decisions made in the ED [19], it is time-consuming to distill relevant information from a large collection of clinical notes and other documents within the time constraints of IVT. Furthermore, in the already overcrowded EDs in Taiwan [20], physicians may suffer from cognitive overload because their work involves constant multitasking and frequent interruptions, resulting in the possibility of medical errors [21].

Recently, natural language processing (NLP) techniques have shown considerable promise in extracting meaningful information from EMRs to support clinical decision-making and to facilitate secondary use of EMRs for research [22–25]. NLP is a field of computer science and artificial intelligence, and involves the analyzing, understanding and generation of the languages that humans use naturally. NLP techniques enable various kinds of text processing tasks including named entity recognition, information extraction, text classification, and information retrieval [26]. The most commonly used general-purpose NLP tools for extracting information from clinical text are cTAKES, MetaMap, and MedLee [25]. However, such techniques have only recently received attention in the management of stroke. Very scarce evidence has been accumulated in the literature regarding NLP and stroke. For example, Mowery et al. tested an NLP tool called pyContext for its ability to distinguish clinical reports with significant carotid stenosis from those without [27]. Although not specifically designed for strokes, Giang et al. compared their text extraction algorithm with expert judgment of the Barthel index to measure quality of life [28]. Because of the dire need for a timely assessment of medical history to evaluate patient eligibility for IVT, we believe that acute stroke care can be improved through the use of EMRs with NLP techniques.

Therefore, this study proposed a new information processing algorithm to automatically extract key medical concepts from clinical notes that are relevant to the IVT eligibility criteria, and developed an enhanced task-specific EMR interface to streamline the presentation of the above analysis for a timely treatment of AIS. Additionally, experiments were conducted to (a) evaluate the accuracy of the proposed algorithm to identify relevant medical concepts from clinical notes, and (b) assess the improvement in completion time and accuracy of the task-specific EMR interface in helping clinicians evaluate clinical notes. Our aim was to build on the existing literature of NLP in healthcare to explore the feasibility of NLP in clinical decisions for timely stroke thrombolysis. To the best of our knowledge, this study was the first to apply NLP techniques to this clinical problem.

## 2. Materials and methods

Our approach consisted of three main phases: (1) medical concept extraction from IVT eligibility criteria, (2) medical concept identification from EMRs, and (3) development of a task-specific EMR interface to support clinical decision-making. Below, we first detail the data collection procedure and then give in-depth descriptions of each text processing task in our approach.

### 2.1. Dataset

This study was conducted in Ditmanson Medical Foundation Chiayi Christian Hospital, a 1000-bed hospital with an ED volume of approximately 100,000 patient visits per year. About 400 patients with AIS are admitted via ED each year. Upon arrival of a patient presenting with stroke-like symptoms within 3 h of onset, a rapid thrombolysis protocol (code stroke) will be activated [29], and a nurse practitioner, an ED physician, and an on-call stroke physician on the acute stroke team will collaboratively manage the patient [30].

Cases for patient review were selected from the stroke registry of the study hospital. Adult patients with AIS who presented to the ED within 3 h of onset but were not treated with IVT between October 2007 and December 2015 were enrolled. Patients who were not treated because of age over 80 years but otherwise eligible for IVT were also included. We focused on complex patients with multiple underlying diseases and frequent interactions with the healthcare system; only patients with more than five visits to the study hospital in the previous two years along with at least one inpatient visit were included. An evaluation corpus was assembled by sampling one discharge summary note and two outpatient notes for each patient from the clinical notes recorded prior to the ED visit for AIS. The study protocol was approved by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Institutional Review Board (CYCH-IRB No.104105). Patient identifiers were replaced by sequential numeric identifiers to ensure confidentiality. An informed consent was thus exempted.

### 2.2. Natural language processing

This study used MetaMap, an NLP tool developed by the National Library of Medicine, to extract biomedical concepts from free-form text [31]. MetaMap breaks down inputted text into words or phrases through a lexical/syntactic analysis, including tokenization, sentence boundary determination, part-of-speech tagging, and parsing, and generates variants of the phrase words [32]. MetaMap then identifies all the possible candidate terms in the Unified Medical Language System (UMLS) and evaluates how each candidate matches the phrase retrieved in the previous process based on measures of centrality, variation, coverage, and cohesiveness. For each matched phrase, MetaMap categorizes it into a semantic type and then returns a concept unique identifier (CUI) with a score between 0 and 1000 based on the strength of the mapping [32].

### 2.3. Phase I: extraction of CUIs from IVT eligibility criteria

Although IVT eligibility criteria varied across guidelines [3,13], they generally consisted of medical conditions for which IVT is contraindicated. Some of the medical conditions are contraindications to IVT only if they occur within a time frame before stroke (Supplemental Table 1). The most recent guidelines by the Taiwan Stroke Society [13] do not categorize these contraindications as absolute or relative. The first phase in the development of the algorithm was to determine which target concepts were to be mined for each medical condition. For example, the medical condition of "use of oral anticoagulants" was determined by domain experts to consist of the target concept of "oral anticoagulant" plus specific drug classes of oral anticoagulants (Supplemental Table 1). A "start list" of CUIs was generated using MetaMap to map each target concept (Table 1) to the UMLS Metathesaurus, version 2016AB. To include all CUIs related to the target concepts in the list, we further adopted a method proposed by Davis et al. [33] to perform multiple queries from tables MRCONSO and MRREL within the UMLS schema. We then eliminated duplicates and manually removed CUIs that were considered unrelated to the original medical condition. Finally, a total of 361 distinct CUIs were produced in the "target list" of

**Table 1**
Performance of automatic identification of contraindications to intravenous thrombolysis in experiment I.

| Target concepts | Phrase level | | | | Document level | | | |
|---|---|---|---|---|---|---|---|---|
| | N | P | R | F1 | N | P | R | F1 |
| Oral anticoagulant | 73 | 1.000 | 1.000 | 1.000 | 31 | 1.000 | 1.000 | 1.000 |
| Heparin | 18 | 1.000 | 0.833 | 0.909 | 5 | 1.000 | 1.000 | 1.000 |
| Arterial puncture | 0 | NA | NA | NA | 0 | NA | NA | NA |
| Major surgery | 8 | 1.000 | 0.125 | 0.222 | 6 | 1.000 | 0.167 | 0.286 |
| Serious trauma | 0 | NA | NA | NA | 0 | NA | NA | NA |
| Gastrointestinal hemorrhage | 34 | 1.000 | 0.618 | 0.764 | 18 | 1.000 | 0.667 | 0.800 |
| Urinary tract hemorrhage | 22 | 1.000 | 0.955 | 0.977 | 16 | 1.000 | 0.938 | 0.968 |
| Head trauma | 17 | 1.000 | 1.000 | 1.000 | 14 | 1.000 | 1.000 | 1.000 |
| Stroke | 300 | 0.996 | 0.756 | 0.859 | 118 | 1.000 | 0.915 | 0.956 |
| Intracranial surgery | 1 | NA | 0 | NA | 1 | NA | 0 | NA |
| Intraspinal surgery | 0 | NA | NA | NA | 0 | NA | NA | NA |
| Acute myocardial infarction | 4 | 1.000 | 0.250 | 0.400 | 3 | 1.000 | 0.333 | 0.500 |
| Intracranial hemorrhage | 43 | 1.000 | 0.767 | 0.868 | 26 | 1.000 | 0.769 | 0.870 |
| Intracranial neoplasm | 5 | 1.000 | 0.400 | 0.571 | 5 | 1.000 | 0.400 | 0.571 |
| Arteriovenous malformation | 0 | NA | NA | NA | 0 | NA | NA | NA |
| Aneurysm | 4 | 1.000 | 1.000 | 1.000 | 4 | 1.000 | 1.000 | 1.000 |
| Diabetes mellitus | 144 | 1.000 | 1.000 | 1.000 | 110 | 1.000 | 1.000 | 1.000 |
| Overall (micro-averaged) | 596 | 0.998 | 0.812 | 0.895 | 234 | 1.000 | 0.972 | 0.986 |
| Overall (macro-averaged) | 596 | 1.000 | 0.725 | 0.798 | 234 | 1.000 | 0.766 | 0.829 |

F1, F1 measure; P, precision; R, recall.

CUIs (see the top half of Fig. 1).

### 2.4. Phase II: identification of CUIs from EMRs

In the second phase, a series of text preprocessing modules were applied to the patient clinical notes retrieved from the EMR database before performing CUI identification (see the bottom half of Fig. 1). Because spelling errors are common in clinical notes, the text from the evaluation corpus was first preprocessed to correct misspelled words using Google's spell checker API. Specifically, a ranked list of recommended words was generated by Google's spell checker and the misspelled word was unconditionally replaced by the first word of the list. Next, acronyms and abbreviations were expanded by looking up a list of common clinical acronyms and abbreviations used locally (Supplemental Table 2). Non-ASCII characters were removed to avoid problems during MetaMap processing. MetaMap configured with the word sense disambiguation option was then used to detect mentions of the target concepts in the target list (Supplemental Table 1) by mapping the text from the evaluation corpus to the UMLS CUIs. The NegEx algorithm was used to detect negated concepts [32].

### 2.5. Phase III: a task-specific EMR interface for assessing IVT eligibility criteria

To present the results of matched IVT-relevant information to clinicians, a new user interface was developed in the third phase (Fig. 2). The interface provides a list of contraindications to IVT in the left panel. The upper right panel lists clinical notes sorted by date in descending order. A contraindication is highlighted in red color and marked with an asterisk if any mention of the related medical concepts is detected in clinical notes within the time frame for this contraindication. The corresponding clinical notes are also marked with an asterisk. The contents of the selected clinical note are shown in the lower right panel, in which the phrases containing the identified concepts in positive form are highlighted in red, whereas those in negated form are highlighted in green.

### 2.6. Experimental evaluation and statistical analysis

Two experiments were conducted. The first experiment examined the performance of automatic identification of the target concepts by the proposed algorithm. Manual annotations were used as the reference standard. Two stroke neurologists (SFS and LCH) independently annotated the unprocessed evaluation corpus and determined whether the target concepts were mentioned. All discrepancies were resolved by consensus. In performance evaluation, we considered two levels of granularity, i.e., phrase level and document level. For example, if a

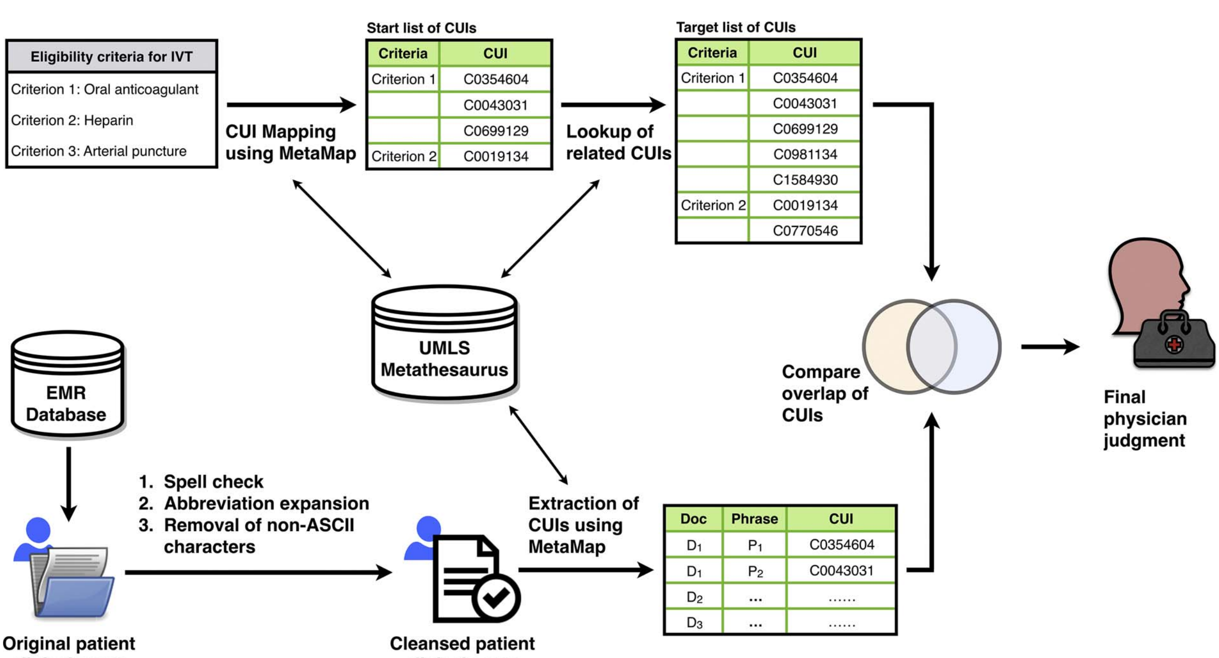


**Fig. 1.** The workflow of the information processing algorithm.

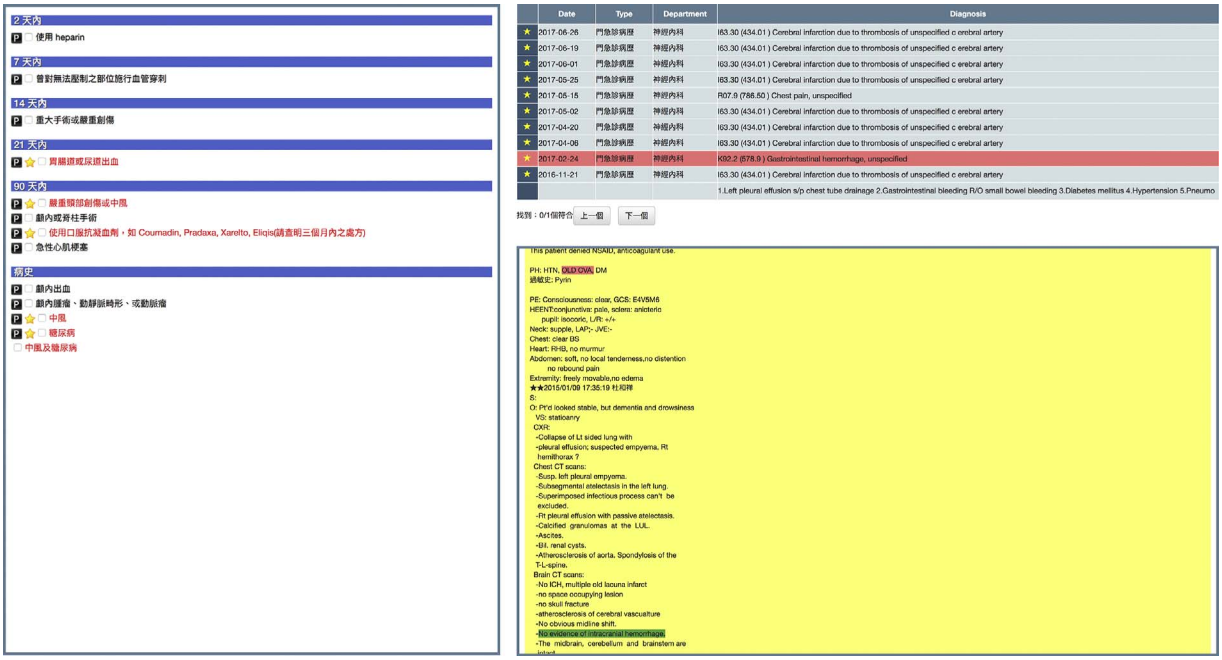


**Fig. 2.** The new task-specific EMR interface. Clinical notes with mentions of medical concepts related to contraindications to intravenous thrombolysis are marked with an asterisk. Concepts in positive form are highlighted in red whereas those in negated form are highlighted in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clinical note reads "He had an ACUTE ISCHEMIC STROKE two years ago. The CT showed a LACUNAR INFARCTION in the left internal capsule. He was left with right hemiparesis after the STROKE". The document-level annotation would be positive for the target concept of stroke, whereas the phrase-level annotation would list three phrases related to this target concept.

Precision (positive predictive value), recall (sensitivity), and F1 measure were used to evaluate the performance of automatic identification of each target concept against the reference standard at the phrase and document levels. We did not differentiate positive from negated concepts. For example, if a clinical note reads "no evidence of INTRACRANIAL HEMORRHAGE", this is still a positive mention of intracranial hemorrhage.

The second experiment recruited twelve clinicians from the staff on the acute stroke team of the study hospital to compare user performance between the task-specific and the current EMR interfaces. Participation was voluntary and compensated. Age, gender, years on acute stroke team, and years of experience with the current EMR were collected for each participant. The current EMR interface (Fig. 3) uses a tabbed view to provide access to clinical notes. List boxes in the top third of the view contain lists of clinical notes sorted by date, and the contents of clinical notes are displayed in the bottom two thirds of the view. Users can scroll through the lists and select what to read. Although the section headers in clinical notes are written in Chinese, the clinical notes themselves are written in English at the study hospital.

Four mock patient records were derived from the real patient records used in the first experiment. We ensured that each patient record had a similar number of clinical notes and amount of text (Supplemental Table 3). In addition, each patient had a different combination of contraindications to IVT to increase representativeness. Similarly, two stroke neurologists determined the contraindications for IVT in the mock patients, with differences adjudicated to reach a consensus reference answer.

The second experiment employed a 2-period crossover design (Fig. 4). Each user reviewed the four patient records, two with the current EMR interface, and two with the task-specific EMR interface, resulting in a total of 48 testing scenarios. Immediately before using the task-specific EMR interface, users had at most 30 min to practice with the new interface. The total number of users was set at 12 so that each mock patient appeared six times in testing scenarios with the current EMR interface and in those with the task-specific EMR interface (Fig. 4). Users had to check a list of 11 contraindications for IVT (Supplemental Table 1) for each mock patient. Their answers were labeled as correct or incorrect against the reference answer. Each correct answer was given one point. The total points for each testing scenario were divided by 11 to get an accuracy score (maximum = 1). Given the importance of efficiency in clinical practice of IVT, the time to completion for each testing scenario was recorded. After the experiment, users were asked to complete the System Usability Scale (SUS) [34] to evaluate the usability of the task-specific EMR interface.

Given the small sample size in the user experiment, non-parametric statistical analyses were performed. Because each user was exposed to each testing EMR interface twice, the accuracy scores and time to completion were collapsed by averaging for each EMR interface. Then the Wilcoxon signed-rank test was performed for comparison between two EMR interfaces because users were measured repeatedly [35]. As collapse of features results in loss of information and decreases the testing power, sensitivity analyses were performed without collapsing the data [35].

Two-tailed $P$ values < 0.05 were considered statistically significant. Statistical analyses were performed using Stata 15 (StataCorp, College Station, Texas).

## 3. Results

### 3.1. Experiment I: performance of automatic identification of IVT eligibility criteria

A corpus of 234 notes from 78 patients were retrieved from the EMR database. A total of 596 phrases were manually determined to be related to the target concepts in Table 1. The most common target concepts identified included stroke, diabetes mellitus, and oral anticoagulant. Table 1 reports the performance of the proposed algorithm. Among the 17 target concepts, the precision was always above 0.99, whereas the recall varied from 0.125 to 1. The F1 measures ranged between 0.222 and 1 at the phrase level and between 0.286 and 1 at the

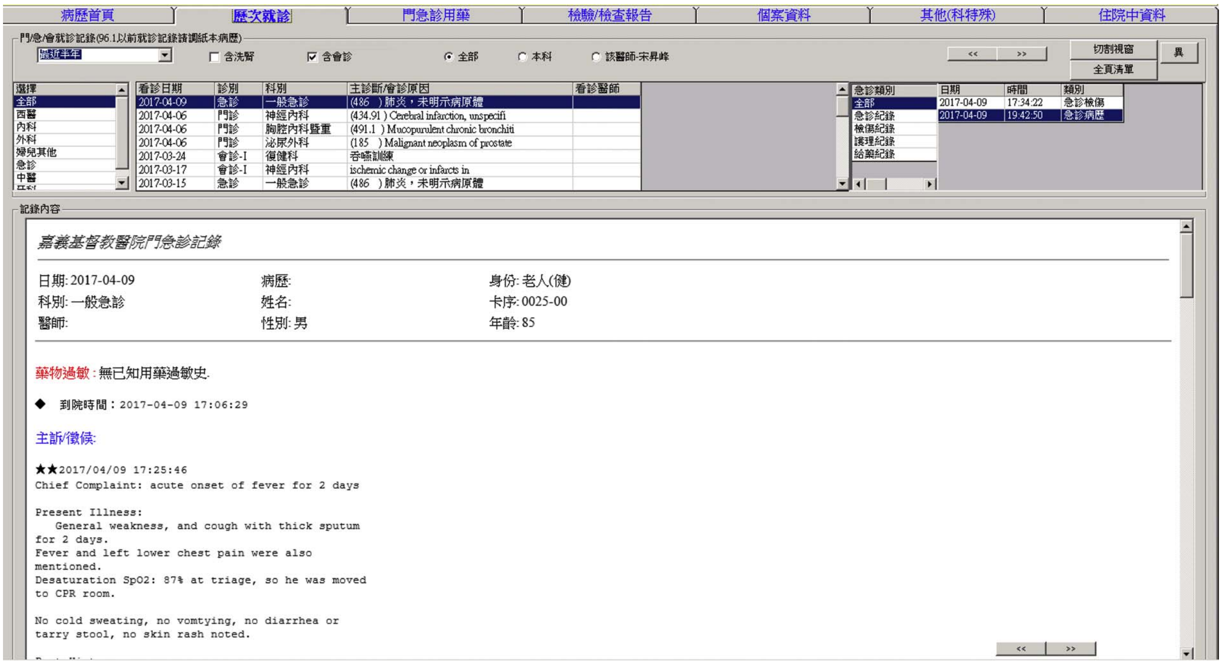


**Fig. 3.** The current EMR interface used in the study hospital.

document level. Lower F1 measures were achieved in identifying the concepts regarding major surgery, acute myocardial infarction, and intracranial neoplasm. Overall, the micro-averaged precisions, recalls, and F1 measures were 0.998, 0.812, and 0.895 at the phrase level and 1, 0.972, and 0.986 at the document level, respectively. The macro-averaged precisions, recalls, and F1 measures were 1, 0.725 and 0.798 at the phrase level and 1, 0.766, and 0.829 at the document level, respectively. These performance metrics indicate that the algorithm has a high level of overall classification accuracy.

### 3.2. Analysis of mapping errors

Of the 596 phrases, only one was identified as a false positive result. Its text included "Taiwan Stroke Society Guidelines", which was annotated by MetaMap with the concept of stroke (C0038454: STROKE). A total of 114 false negatives were recorded. The false negatives could be generalized into three types. In the first type (n = 100), concepts relevant to a target concept was identified by MetaMap but their CUIs were not included in the target list for that concept. For example, "right middle cerebral artery territory infarction" was mapped to "C0751849: Right Middle Cerebral Artery Infarction", which was not listed in the target list for the concept of stroke. The second type (n = 9) was due to missed term failures, in which a relevant term was simply not identified by MetaMap [36,37]. For example, "heparinization" indicates therapeutic administration of heparin. However, it was not identified by MetaMap. The third type (n = 5) was attributed to boundary failures, in which a single coherent term was incorrectly parsed into multiple terms [37]. For example, "gastric ulcers and bleeding" was manually annotated with the concept of gastrointestinal hemorrhage. Instead, it was recognized by MetaMap as two UMLS concepts, i.e., "C0038358: Ulcers, Gastric (Gastric ulcer)" and "C0019080: BLEEDING (Hemorrhage)".

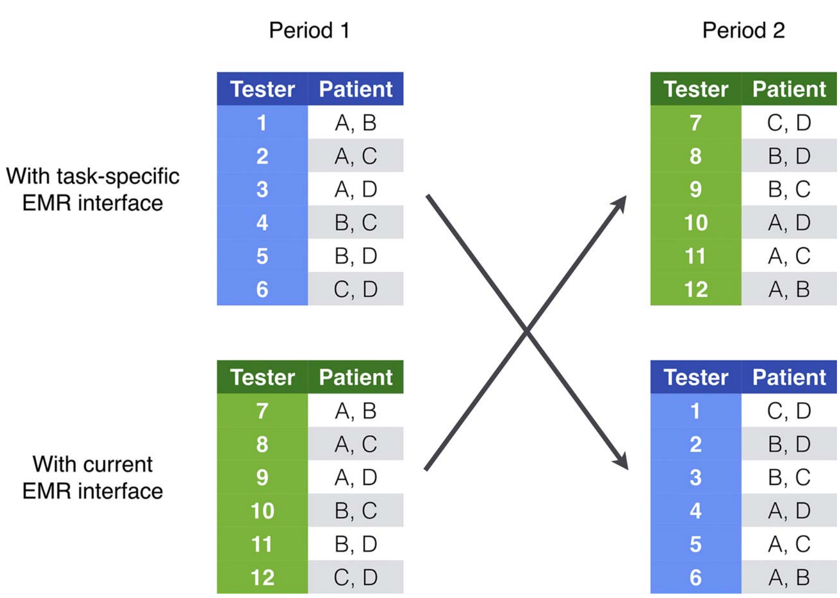


**Fig. 4.** Diagram of the user experiment design.

**Table 2**
Characteristics and performance of users in experiment II. Each user reviewed two patient records with the current EMR interface and two with the task-specific EMR interface. The time to completion and accuracy scores were averaged for each EMR interface.

| Users | Age, year | Gender | On acute stroke team, year | Experience with current EMR, year | SUS score | Current EMR interface | | Task-specific EMR interface | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Average time, min | Average accuracy score | Average time, min | Average accuracy score |
| #1 | 37 | F | 10 | 7 | 57.5 | 1.76 | 82% | 5.12 | 95% |
| #2 | 32 | F | 10 | 8 | 52.5 | 1.65 | 82% | 2.68 | 86% |
| #3 | 31 | F | 5 | 3 | 72.5 | 2.20 | 91% | 1.82 | 91% |
| #4 | 47 | M | 10 | 11 | 30.0 | 1.62 | 77% | 4.87 | 82% |
| #5 | 40 | F | 8 | 6 | 62.5 | 2.41 | 73% | 0.96 | 95% |
| #6 | 40 | M | 10 | 7 | 77.5 | 1.05 | 73% | 0.86 | 95% |
| #7 | 31 | M | 3 | 3 | 75.0 | 1.48 | 73% | 0.77 | 95% |
| #8 | 37 | M | 6 | 6 | 82.5 | 3.80 | 86% | 2.50 | 91% |
| #9 | 28 | F | 1 | 1 | 65.0 | 0.93 | 68% | 3.10 | 86% |
| #10 | 45 | F | 10 | 7 | 75.0 | 3.09 | 91% | 2.42 | 91% |
| #11 | 35 | F | 2 | 2 | 62.5 | 2.24 | 68% | 2.52 | 100% |
| #12 | 39 | F | 10 | 6 | 67.5 | 0.92 | 82% | 0.87 | 64% |

EMR, electronic medical record; SUS, System Usability Scale.

### 3.3. Experiment II: user experiment to assess the task-specific EMR interface

The goal of the user experiment was to examine the capabilities of the improved user interface powered by NLP techniques to improve time and accuracy of the clinical decision on eligibility for IVT. The twelve users comprised two neurologists, eight nurse practitioners, and two ED physicians. The four mock patients had one, three, three, and six contraindications to IVT, respectively. Each mock patient record contained 38 clinical notes and on average each note contained 420 word tokens (Supplemental Table 3). The time to completion for all 48 testing scenarios ranged from 0.58 to 5.97 min, with a median of 1.92 min. The accuracy scores ranged from 36% to 100%, with a median of 91%. None of the 24 testing scenarios using the current EMR interface got all correct answers on the 11 contraindications, whereas 8 (33%) of those using the task-specific EMR interface were completely correct.

Table 2 shows the characteristics of the users, the SUS score, the average time to completion, and the average accuracy score. No statistically significant difference ($p = 0.754$) was observed in the time to complete the checklist for IVT between the two interfaces (Table 3). However, participants using the task-specific EMR interface had a significantly higher average accuracy score ($p = 0.016$) than those using the current EMR interface (Table 3). These findings did not change in the sensitivity analyses ($p = 0.353$ for time and $p = 0.003$ for average accuracy score). The median SUS score was 66 points, with an interquartile range of 60–75 points, which means the usability of the task-specific interface was between “OK” and “good” [38].

## 4. Discussion

### 4.1. Main findings

We found that automatic identification of contraindications to IVT by using MetaMap and the UMLS Metathesaurus is a plausible new direction. For a timely stroke thrombolysis, it is crucial that the acute stroke team can extract key information regarding IVT eligibility from the EMRs within a short time. The results from the user experiment showed that using the current EMR interface, the staff on our acute stroke team could still extract this information quickly, probably because of familiarity in the navigation of the existing system. However, user errors in extracting IVT-relevant data from the EMRs were common. The new task-specific EMR interface helped reduce errors in assessing contraindications to IVT without sacrificing the precious time to offer care.

**Table 3**
Comparison of test results between the two EMR interfaces in experiment II. The time to completion and accuracy scores for the 12 participants were compared using the Wilcoxon signed-rank test.

| | Current EMR interface | Task-specific EMR interface | *P* |
|---|---|---|---|
| Time, min, median (IQR) | 1.70 (1.27–2.33) | 2.46 (0.91–2.89) | 0.754 |
| Accuracy score, median (IQR) | 80% (73%–84%) | 91% (86%–95%) | 0.016 |

EMR, electronic medical record; IQR, interquartile range.

### 4.2. Remedies for mapping errors

In this study, MetaMap's precision was high whereas its recall was suboptimal. Because the information processing algorithm was intended to identify IVT-relevant information and clinicians still have to review all the facts and make the final treatment decision, the algorithm should be viewed as a screening test. Therefore, it is preferable to have false positives rather than false negatives. A low recall suggests that the algorithm produced more false negatives and needed to be improved. As demonstrated by the failure analysis, the main reason was that although MetaMap effectively identified concepts relevant to a target concept, some of their CUIs were not included in the target list for that concept. This problem could be rectified by manually expanding the start list of target concepts based on expert knowledge. A possible strategy could be to implement an interactive mode [28,39] in the EMR interface to allow clinicians to add phrases relevant to each target concept, which in turn triggers the information processing algorithm to rebuild the start list and the target list. The system performance could thus be improved without intervention of the software developers.

Nevertheless, it may be difficult to improve the low recall for some of the target concepts such as major surgery. This is because clinicians generally document actual surgical procedures such as “cholecystectomy” rather than the literal phrase of “major surgery” in patient notes. Moreover, the definition of major surgery may vary across physicians, even though the guidelines explicitly state that major surgery within previous 14 days is a contraindication to IVT. Therefore, a list of concepts related to major surgery will be needed a priori. However, it will be unfeasible to generate an exhaustive list of surgical procedures which are considered major surgery. Future work may explore how structured data, such as the International Classification of Diseases Procedure Codes, can be used in this aspect.

### *4.3. EMR usability*

EMR systems have been widely adopted in Taiwan and a national EMR exchange system has been established to provide inter-institution EMR exchange for improving continuity in healthcare [40]. Patient information is now more readily accessible and transferable. However, with the rapid growth of large amounts of unstructured clinical documents, the so-called “patient information explosion”, EMR users are faced with various challenges, including information filtering, context-sensitive decision support, and data reliability [41]. A usable EMR interface should support clinical decisions by reducing user cognitive workload and improving work processes so that clinicians are able to complete tasks effectively and efficiently with high satisfaction. Usability problems may not only reduce efficiency but also compromise patient safety [42]. A survey on the usability requirements of EMR systems found that “ease of finding the required information on the screen” was the utmost requirement [43]. As seen in this study, errors in assessing contraindications to IVT were almost inevitable with the current EMR interface. By highlighting IVT-relevant information, the task-specific EMR interface significantly reduced errors in assessing contraindications and might thus ensure patient safety.

In addition to enhancing patient safety, prompt and accurate evaluation of IVT eligibility could potentially ease the workload of stroke physicians. To facilitate timely stroke thrombolysis, code stroke protocols are widely implemented in hospitals providing acute stroke care [29,44,45]. Code stroke protocols generally adopt various strategies to expedite the decision-making process of IVT, including pre-acquisition of medical history from EMRs and pre-notification of stroke physicians. However, many code stroke activations are futile because a significant proportion of stroke patients are not eligible for IVT even though they present to the ED in time [29,46]. Excessive code stroke activations have led to overload of on-call stroke physicians [29]. If the eligibility for IVT could have been assessed accurately with the task-specific interface by ED personnel before the activation of the code stroke protocol, such futile activations might have been avoided.

### *4.4. Clinical decision support for stroke*

Despite the significant benefits of using clinical decision support on healthcare [47], the existing electronic tools to support decision-making in IVT are rudimentary and suboptimal [48]. There are very few existing tools focusing on risk communication with patients and outcome prediction of IVT using validated equations [49–51]. For assessment of IVT eligibility, Sun et al. developed a decision support system that searched EMRs for contraindications to IVT based on structured data including International Classification of Diseases codes and medication codes, and presented IVT-relevant information on hand-held devices [52]. Nevertheless, previous studies have found that NLP techniques have advantages over methods using disease codes in identifying clinical cases at only slightly greater costs of time and effort [53,54]. It was advised that information extracted from NLP-processed text should be used in combination with structured data.

### *4.5. Limitations*

Certain limitations were present in this study. First, only 234 clinical notes from 78 patients were used to test the performance of automatic identification of IVT eligibility criteria. The small sample size may raise concerns regarding the interpretability of the results and the generalizability of findings. In addition, it has made some of the target concepts too rare to be tested effectively (Table 1). Although this may reflect the low incidence of these medical conditions in patients with stroke, obtaining information about these medical conditions is still crucial before a treatment decision can be made. Further work should collect and test clinical notes from patients with these medical conditions before deploying the new EMR interface. Second, the user experiment was not conducted in real-world practice settings. We had to use mock patient records to standardize the experiment among different testing scenarios in the user experiment. Further experiments in the real-world practice are needed to investigate the usefulness of the task-specific EMR interface. Third, we did not measure the time spent on processing patient clinical notes. A main weakness of MetaMap is the long processing time, resulting in its insufficiency for real-time annotation of large amounts of clinical notes [32]. An information technology infrastructure capable of processing large volume of notes should be ensured before the implementation of the task-specific EMR interface for production use [35]. Fourth, the user experiment compared an already familiar interface and an entirely new one. The participants were unlikely to master the new interface within limited time. Although no significant difference was found in the completion time between the two EMR interfaces, the completion time is more likely to be further reduced after the participants gain more experience with the new interface. Finally, we did not use structured data because the purpose of this study was to explore the feasibility of using NLP techniques on the assessment of IVT eligibility criteria.

## 5. Conclusion

With the widespread use of EMRs, the amount of unstructured free-text clinical documents is increasing. Although clinical notes contain rich patient information, they are less amenable to manipulation and utilization by clinical decision support tools compared to structured data. Our work contributes to the literature and healthcare practice in several ways. First, our feasibility study demonstrated that a task-specific EMR interface using NLP techniques may help extract and present meaningful information in a timely manner to facilitate the decision-making process of IVT in patients with AIS. Second, UMLS, despite being an authoritative source of health and biomedical vocabularies, is still limited in clinical use. Our approach adopted UMLS to produce a list of CUIs to be used as targets for mapping; however, the final target list still needs revision by clinicians to include additional relevant CUIs. But, once this target list is generated and reviewed, it does not need to be changed frequently. Therefore, it offers a good level of stability for the proposed automatic NLP algorithm to produce consistent results. In the end, the whole task-specific user interface together with the NLP enhanced system could improve the overall accuracy of patient selection for IVT. Third, our work provides some successful evidence for bridging three disciplines (healthcare, NLP, and technology) to provide a practical solution to a problem that has a pressing need for an acceptable solution.

Future work should include combining information from structured data and using feedback from domain experts to iteratively refine the relevancy of presented information, which will hopefully enable clinicians to improve acute stroke care.

## Funding

This research was supported in part by the Ministry of Science and Technology (grant number MOST 105-2314-B-705-001) and the Ditmanson Medical Foundation Chia-Yi Christian Hospital Research Program (R106-021).

## Competing interests

The authors have no competing interests to declare.

## Author contributions

All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: SFS and YHH. Acquisition of data: SFS, DPW, LCH and YHS. Analysis and interpretation of data: SFS, KC and YHH. Drafting of the manuscript: SFS and YHH. Critical revision of the manuscript for important intellectual content: All authors. Study supervision: YHH.

Summary Points

What was already known on the topic:

- Intravenous thrombolysis (IVT) is a standard treatment for acute ischemic stroke (AIS), but the use of IVT is limited to patients fulfilling the eligibility criteria.
- The effect of IVT on functional outcomes is time-dependent; patients with AIS should be treated within 4.5 h of stroke onset.
- The existing electronic tools to support decision making in IVT are rudimentary and suboptimal. There are very few existing tools focusing on risk communication with patients and outcome prediction of IVT. Only one tool was developed to search electronic medical records (EMRs) for contra-indications to IVT based on structured data.
- Natural language processing (NLP) techniques have received little attention in the management of stroke.

What knowledge this study adds:

- It is feasible to automatically identify key information regarding IVT eligibility from the EMRs using an off-the-shelf NLP tool, MetaMap, and the UMLS Metathesaurus.
- The user experiment showed that with the conventional EMR interface, clinicians could quickly extract IVT-relevant information from the EMRs, albeit with many errors.
- By highlighting IVT-relevant information, the task-specific EMR interface significantly reduced errors in assessing contraindications to IVT and might thus ensure patient safety and ease the workload of stroke physicians.
- NLP techniques can help extract key information regarding IVT eligibility from the EMRs within a short time and thus improve the quality of clinical decision making.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijmedinf.2018.02.005.

## References


[1] V.L. Feigin, B. Norrving, G.A. Mensah, Global burden of stroke, Circ. Res. 120 (2017) 439–448.
[2] B. Ovbiagele, M.N. Nguyen-Huynh, Stroke epidemiology: advancing our understanding of disease mechanism and therapy, Neurotherapeutics 8 (2011) 319–329.
[3] E.C. Jauch, J.L. Saver, H.P. Adams, A. Bruno, J.J.B. Connors, B.M. Demaerschalk, et al., Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association, Stroke 44 (2013) 870–947.
[4] The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, Tissue plasminogen activator for acute ischemic stroke, N. Engl. J. Med. 333 (1995) 1581–1587.
[5] W. Hacke, M. Kaste, E. Bluhmki, M. Brozman, A. Dávalos, D. Guidetti, et al., Thrombolysis with alteplase 3–4.5 hours after acute ischemic stroke, N. Engl. J. Med. 359 (2008) 1317–1329.
[6] K.R. Lees, E. Bluhmki, R. von Kummer, T.G. Brott, D. Toni, J.C. Grotta, et al., Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials, Lancet 375 (2010) 1695–1703.
[7] J.L. Saver, G.C. Fonarow, E.E. Smith, M.J. Reeves, M.V. Grau-Sepulveda, W. Pan, et al., Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke, JAMA 309 (2013) 2480–2488.
[8] A. Meretoja, D. Strbian, S. Mustanoja, T. Tatlisumak, P.J. Lindsberg, M. Kaste, Reducing in-hospital delay to 20 minutes in stroke thrombolysis, Neurology 79 (2012) 306–313.
[9] Y. Xian, E.E. Smith, X. Zhao, E.D. Peterson, D.M. Olson, A.F. Hernandez, et al., Strategies used by hospitals to improve speed of tissue-type plasminogen activator treatment in acute ischemic stroke, Stroke 45 (2014) 1387–1395.
[10] D. Strbian, T. Sairanen, A. Meretoja, J. Pitkäniemi, J. Putaala, O. Salonen, et al., Patient outcomes from symptomatic intracerebral hemorrhage after stroke thrombolysis, Neurology 77 (2011) 341–348.
[11] M.I. Weintraub, Thrombolysis (tissue plasminogen activator) in stroke: a medicolegal quagmire, Stroke 37 (2006) 1917–1922.
[12] W.N. Whiteley, K.B. Slot, P. Fernandes, P. Sandercock, J. Wardlaw, Risk factors for intracranial hemorrhage in acute ischemic stroke patients treated with recombinant tissue plasminogen activator: a systematic review and meta-analysis of 55 studies, Stroke 43 (2012) 2904–2909.
[13] H.-H. Hu, Taiwan Guidelines for the Management of Stroke 2008, Taiwan Stroke Society, Taipei, 2008.
[14] A.M. Lopez-Yunez, A. Bruno, L.S. Williams, E. Yilmaz, C. Zurrú, J. Biller, Protocol violations in community-based rTPA stroke treatment are associated with symptomatic intracerebral hemorrhage, Stroke 32 (2001) 12–16.
[15] D.M. Tisnado, J.L. Adams, H. Liu, C.L. Damberg, W.-P. Chen, F.A. Hu, et al., What is the concordance between the medical record and patient self-report as data sources for ambulatory care? Med. Care 44 (2006) 132–140.
[16] D.M. Kriegsman, B.W. Penninx, J.T. van Eijk, A.J. Boeke, D.J. Deeg, Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy, J. Clin. Epidemiol. 49 (1996) 1407–1417.
[17] H.H. Rau, C.Y. Hsu, Y.L. Lee, W. Chen, Developing electronic health records in Taiwan, IT Prof. 12 (2010) 17–25.
[18] P. Sharda, A.K. Das, V.L. Patel, Specifying design criteria for electronic medical record interface using cognitive framework, AMIA Annu. Symp. Proc. 2003 (2003) 594–598.
[19] O. Ben-Assuli, D. Sagi, M. Leshno, A. Ironi, A. Ziv, Improving diagnostic accuracy using EHR in emergency departments: a simulation-based study, J. Biomed. Inform. 55 (2015) 31–40.
[20] F.Y. Shih, M.H. Ma, S.C. Chen, H.P. Wang, C.C. Fang, R.S. Shyu, et al., ED overcrowding in Taiwan: facts and strategies, Am. J. Emerg. Med. 17 (1999) 198–202.
[21] A. Laxmisan, F. Hakimzada, O.R. Sayan, R.A. Green, J. Zhang, V.L. Patel, The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care, Int. J. Med. Inform. 76 (2007) 801–811.
[22] K.B. Wagholikar, K.L. MacLaughlin, M.R. Henry, R.A. Greenes, R.A. Hankey, H. Liu, et al., Clinical decision support with automated text processing for cervical cancer screening, J. Am. Med. Inform. Assoc. 19 (2012) 833–839.
[23] H. Salmasian, D.E. Freedberg, C. Friedman, Deriving comorbidities from medical records using natural language processing, J. Am. Med. Inform. Assoc. 20 (2013) e239–e242.
[24] R.J. Byrd, S.R. Steinhubl, J. Sun, S. Ebadollahi, W.F. Stewart, Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records, Int. J. Med. Inform. 83 (2014) 983–992.
[25] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, et al., Clinical information extraction applications: a literature review, J. Biomed. Inform. 77 (2017) 34–49.
[26] I. Spasić, Text mining of cancer-related information: review of current status and future directions, Int. J. Med. Inform. 83 (2014) 605–623.
[27] D.L. Mowery, B.E. Chapman, M. Conway, B.R. South, E. Madden, S. Keyhani, et al., Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis, J. Biomed. Semantics 7 (2016) 26.
[28] P. Giang, A. Williams, L. Argyros, Automated extraction of the Barthel Index from clinical texts, AMIA Annu. Symp. Proc. 2013 (2013) 486–495.
[29] S.-F. Sung, M.-C. Tseng, Code stroke: a mismatch between number of activation and number of thrombolysis, J. Formos. Med. Assoc. 113 (2014) 442–446.
[30] S.-F. Sung, Y.-C. Huang, C.-T. Ong, Y.-W. Chen, A parallel thrombolysis protocol with nurse practitioners As coordinators minimized door-to-needle time for acute ischemic stroke, Stroke Res. Treat. 2011 (2011) 198518–8.
[31] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, Proc. AMIA Symp. (2001) 17–21.
[32] A.R. Aronson, F.-M. Lang, An overview of MetaMap: historical perspective and recent advances, J. Am. Med. Inform. Assoc. 17 (2010) 229–236.
[33] K. Davis, C. Staes, J. Duncan, S. Igo, J.C. Facelli, Identification of pneumonia and influenza deaths using the Death Certificate Pipeline, BMC Med. Inform. Decis. Mak. 12 (2012) 37.
[34] J. Brooke, SUS – A quick and dirty usability scale, in: P.W. Jordan, B. Thomas, I.L. McClelland, B. Weerdmeester (Eds.), Usability Evaluation in Industry, CRC Press, London, 1996, pp. 189–194.
[35] J.S. Hirsch, J.S. Tanenbaum, S. Lipsky Gorman, C. Liu, E. Schmitz, D. Hashorva, et al., HARVEST, a longitudinal patient record summarizer, J. Am. Med. Inform. Assoc. 22 (2015) 263–274.
[36] G. Divita, T. Tse, L. Roth, Failure analysis of MetaMap transfer (MMTx), Stud. Health Technol. Inform. 107 (2004) 763–767.
[37] A. Park, A.L. Hartzler, J. Huh, D.W. McDonald, W. Pratt, Automatically detecting failures in natural language processing tools for online community text, J. Med. Internet Res. 17 (2015) e212.
[38] A. Bangor, P. Kortum, J. Miller, Determining what individual SUS scores mean: adding an adjective rating scale, J. Usability Stud. 4 (2009) 114–123.
[39] G. Trivedi, P. Pham, W.W. Chapman, R. Hwa, J. Wiebe, H. Hochheiser, NLPReViz: an interactive tool for natural language processing on clinical text, J. Am. Med. Inform. Assoc. 25 (2017) 81–87.
[40] Y.-C.J. Li, J.-C. Yen, W.-T. Chiu, W.-S. Jian, S. Syed-Abdul, M.-H. Hsu, Building a

national electronic medical record exchange system – experiences in Taiwan, Comput. Methods Progr. Biomed. 121 (2015) 14–20.
[41] E.S. Berner, J. Moss, Informatics challenges for the impending patient information explosion, J. Am. Med. Inform. Assoc. 12 (2005) 614–617.
[42] M.F. Walji, E. Kalenderian, M. Piotrowski, D. Tran, K.K. Kookal, O. Tokede, et al., Are three methods better than one? A comparative assessment of usability evaluation methods in an EHR, Int. J. Med. Inform. 83 (2014) 361–367.
[43] M. Farzandipour, Z. Meidani, H. Riazi, M. Sadeqi Jabali, Task-specific usability requirements of electronic medical records systems: lessons learned from a national survey of end-users, Inform. Health Soc. Care. 13 (2017) 1–20.
[44] A. Meretoja, L. Weir, M. Ugalde, N. Yassi, B. Yan, P. Hand, et al., Helsinki model cut stroke thrombolysis delays to 25 minutes in Melbourne in only 4 months, Neurology 81 (2013) 1071–1076.
[45] C.-H. Chen, S.-C. Tang, L.-K. Tsai, M.-J. Hsieh, S.-J. Yeh, K.-Y. Huang, et al., Stroke code improves intravenous thrombolysis administration in acute ischemic stroke, PLoS One 9 (2014) e104862.
[46] D. Kleindorfer, B. Kissela, A. Schneider, D. Woo, J. Khoury, R. Miller, et al., Eligibility for recombinant tissue plasminogen activator in acute ischemic stroke: a population-based study, Stroke 35 (2004) e27–e29.
[47] S.S. Jones, R.S. Rudin, T. Perry, P.G. Shekelle, Health information technology: an updated systematic review with a focus on meaningful use, Ann. Intern. Med. 160 (2014) 48–54.
[48] D. Flynn, G.A. Ford, L. Stobbart, H. Rodgers, M.J. Murtagh, R.G. Thomson, A review of decision support, risk communication and patient information tools for thrombolytic treatment in acute stroke: lessons for tool developers, BMC Health Serv. Res. 13 (2013) 225.
[49] D.M. Kent, H.P. Selker, R. Ruthazer, E. Bluhmki, W. Hacke, The stroke-thrombolytic predictive instrument: a predictive instrument for intravenous thrombolysis in acute ischemic stroke, Stroke 37 (2006) 2957–2962.
[50] V.L. Cunningham, The outcome wheel: a potential tool for shared decision-making in ischemic stroke thrombolysis, CJEM 10 (2008) 545–551.
[51] G. Saposnik, J. Fang, M.K. Kapral, J.V. Tu, M. Mamdani, P. Austin, et al., The iScore predicts effectiveness of thrombolytic therapy for acute ischemic stroke, Stroke 43 (2012) 1315–1322.
[52] M.-C. Sun, J.-A. Chan, A clinical decision support tool to screen health records for contraindications to stroke thrombolysis-a pilot study, BMC Med. Inform. Decis. Mak. 15 (2015) 105.
[53] L. Li, H.S. Chase, C.O. Patel, C. Friedman, C. Weng, Comparing ICD9-encoded diagnoses and NLP-processed discharge summaries for clinical trials pre-screening: a case study, AMIA Annu. Symp. Proc. 2008 (2008) 404–408.
[54] J. Friedlin, M. Overhage, M.A. Al-Haddad, J.A. Waters, J.J.R. Aguilar-Saavedra, J. Kesterson, et al., Comparing methods for identifying pancreatic cancer patients using electronic data sources, AMIA Annu. Symp. Proc. 2010 (2010) 237–241.