MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190254

Identifying Suicidal Adolescents from Mental Health Records Using Natural Language Processing

Sumithra Velupillai [a,b], Sophie Epstein [a,c], André Bittar [a], Thomas Stephenson [c], Rina Dutta [*,a,c], Johnny Downs [*,a,c],

[a] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK,
[b] School of Electrical Engineering and Computer Science, KTH, Stockholm, Sweden
[c] South London and Maudsley NHS Foundation Trust, London, UK

[*] contributed equally to this work

Abstract

Suicidal ideation is a risk factor for self-harm, completed suicide and can be indicative of mental health issues. Adolescents are a particularly vulnerable group, but few studies have examined suicidal behaviour prevalence in large cohorts. Electronic Health Records (EHRs) are a rich source of secondary health care data that could be used to estimate prevalence. Most EHR documentation related to suicide risk is written in free text, thus requiring Natural Language Processing (NLP) approaches. We adapted and evaluated a simple lexicon- and rule-based NLP approach to identify suicidal adolescents from a large EHR database. We developed a comprehensive manually annotated EHR reference standard and assessed NLP performance at both document and patient level on data from 200 patients (~5000 documents). We achieved promising results (>80% f1 score at both document and patient level). Simple NLP approaches can be successfully used to identify patients who exhibit suicidal risk behaviour, and our proposed approach could be useful for other populations and settings.

Keywords:

Natural Language Processing; Suicide; Electronic Health Records

Introduction

Suicidal ideation (SI) occurs in approximately 10% of adolescents in the general population [1,2]. It is a risk factor for self-harm and completed suicide, and is associated with a number of mental disorders. Amongst young people attending Child and Adolescent Mental Health Services (CAMHS), who are often receiving therapeutic support for mental health disorders, the prevalence is higher. SI is an important risk indicator that is used in clinical practice by CAMHS professionals. It is asked about in routine clinical assessments, particularly in situations where patients report low mood, anxiety, emotional regulation difficulties, trauma or distress following difficult life experiences.

Information about suicide risk including suicidal thoughts can be documented in structured risk assessment forms [3], or in free-text in Electronic Health Records (EHRs). Whilst admissions to general hospitals following suicide attempts would be routinely recorded with structured diagnostic codes, routine assessments of suicidal behaviour risk in clinical practice are predominantly recorded as free-text [4–6]. To capture and extract suicide-related information from text within EHRs, Natural Language Processing (NLP) approaches are essential. Common to many NLP solutions for extracting particular pieces of information from EHR text are: to 1) identify concepts of relevance (e.g. the term *suicide*) and 2) classify contextual attributes that semantically alter the meaning of the identified mention. Negation detection is particularly important for clinical constructs that are routinely documented as part of clinical assessments (e.g. *denies suicidal ideation*). Current state-of-the-art NLP methodologies that model such problems in the clinical domain rely on machine learning approaches or symbolic approaches (or a combination of both) [7,8].

However, extracting *patient*-level risks or classifications from (a collection of) documents presents some challenges. Each patient EHR might contain several documents, and each document might contain several relevant (and irrelevant) mentions of clinically important information. Particularly for complex clinical constructs such as suicidality, generating sufficiently large and high-quality datasets that could be used to develop machine-learning based models is costly due to the time and human effort required for manual data annotation. Inferring patient-level labels from individual mentions can be done by post-processing identified mentions using currently available clinical evidence, as was done in a recent study for asserting asthma status [9], or by other heuristic approaches.

Haerian et al. (2012) found that NLP approaches combined with ICD-9 codes had higher precision/positive predictive value (PPV) than using ICD-9 codes alone when evaluated on data from the New York Presbyterian Hospital/Columbia University Medical Center [4]. Similarly, Anderson et al. (2015) found that only 3% of patients who had an indication of suicidal ideation written in text had a corresponding ICD-9 code in a study performed on a large distributed health network of primary care organizations [5]. In a study to identify suicidal behaviour amongst pregnant women, similar findings are reported: using diagnostic codes alone reduced sensitivity considerably [10]. Rule- and machine-learning based approaches to extract suicide ideation and suicide attempt mentions from psychiatric EHR text have been developed with promising results [6], but were evaluated on the mention level, not the patient level.

We have previously developed a step-wise rule-based NLP approach that 1) identifies suicide-related mentions and filters out negated instances using lexicons (for concepts as well as negation terms) and rules based on syntactic information (constituency-based parse tree of sentences), 2) determines a document-level label based on heuristics (counting the number of positive and negative instances and assigning a document-level label based on a majority rule), and 3) determines a patient-level label based on document labels (patient positive for suicidal behaviour = the patient had one (or more) documents labelled as positive for suicidality). This approach was evaluated on a cohort of adolescents diagnosed with autism spectrum disorder (ASD), resulting in precision, recall, and f1 scores all > 0.85 at the document and patient level [11].

In the study presented here, we extend this approach. We had two main aims: 1) to generate a manually annotated reference standard of an adolescent cohort that was inclusive of all mental health conditions and 2) to apply and modify an existing NLP approach for mention-level extraction of specified clinical constructs [12] that only relied on target (relevant clinical concepts) and modifier (terms that alter the contextual meaning of the target, e.g. negation) lexicons and a sentence tokenizer; thus, minimizing the need for time-costly pre-processing steps such as syntactic parsing, and test its performance on the new cohort. Our focus was on finding *asserted* suicidality risk, filtering out negated cases. Further, we applied previously published target and modifier lexicons (in this case, negation terms), to assess their applicability on a new clinical use-case and dataset. Finally, we developed and evaluated these lexicons at both document and patient level. To our knowledge, this is the first study that focuses on a general adolescent patient cohort and that systematically analyses the applicability of published lexical resources on the problem of retrieving asserted suicide risk from EHR text.

The capacity to extract data on suicidality from free-text EHRs has many implications: it provides an estimate of the prevalence of this problem in a clinical CAMHS population; it generates a cohort of high-risk patients in which to study risk factors for self-harm and suicide; and it contributes to an emerging field of research using text from routinely collected health data and data-driven methods to help understand suicide risk at scale.

Methods

Data Source and Clinical Cohort

We used data from the Clinical Records Interactive Search (CRIS) database [13,14], which contains anonymized EHRs from the South London and Maudsley (SLaM) NHS Foundation Trust and has ethical approval for research use (Oxford REC C, reference 08/H0606/71+5) under an extensive governance model. The full cohort consisted of all patients aged 11-17 and in contact with CAMHS in SLaM between April 1st 2009 and March 31st 2016. All documents for these patients in this time period were extracted (documents include event notes, correspondence letters, and free-text sections from different types of semi-structured assessments), yielding a total of 1,601,422 documents (derived from 23,455 patients)[1]. No structured data was included, and access to patient outcomes was not covered by the ethical approval for this project.

Reference Standard Development

EHR Corpus

All documents for a sample of 200 patients were extracted for manual review. This subset was split into a training set (100 patients, 2883 documents), and a test set (100 patients, 2602 documents). The sample was randomly extracted from the patients who had a number of EHR documents within the 1st and 3rd quartiles (12,146 patients, 342,037 documents in total, cut-off points: minimum of 10, maximum of 61 per patient).

Annotation

A review of similar studies was performed to guide the definition of suicide-related information, and suicidal behaviour. In summary, any mentions that indicated the desire to kill oneself, end one's life, or wanting to be dead/to die (including having tried to kill oneself) were included. Mentions related to self-harm without suicidal intent or no explicit mention of suicidal intent were excluded. No assumptions on implicit information about intent (e.g. overdoses) were made and were thus not annotated. All documents for each patient were manually reviewed and annotated by a child psychiatrist (SE) for a document-level label (non-relevant, suicidal, non-suicidal, or uncertain). Each document labelled as suicidal, non-suicidal or uncertain was also annotated with mention-level annotations (positive, negated, uncertain or unrelated). The definition of a mention-level annotation was deliberately not constrained to specific textual units (e.g. word or paragraph), instead, any text segment that was indicative of the mention-level labels could be marked by the annotator to allow for expressivity and clinically relevant segments. Guidelines were iteratively developed to refine inclusion and exclusion criteria for each label. A small subset (100 documents, randomly extracted with a 2:1 ratio of not positive:positive) was annotated by a second clinical annotator (TS) to measure inter-annotator agreement (IAA). We used the Extensible Human Oracle Suite of Tools (eHost)[2] annotation tool.

Document- and Patient-Level Label Heuristics

Because the purpose of the clinical use-case for which this approach was developed was to generate a cohort of patients who had *ever* exhibited suicidal tendencies, priority was placed on detecting *affirmed* suicide-related information. Thus, if there was *at least one* affirmed mention of suicidality in a document, the document label was set to suicidal. Similarly, a patient-level label of suicidal was set if there was at least one document labelled as positive for suicidality (Figure 1).

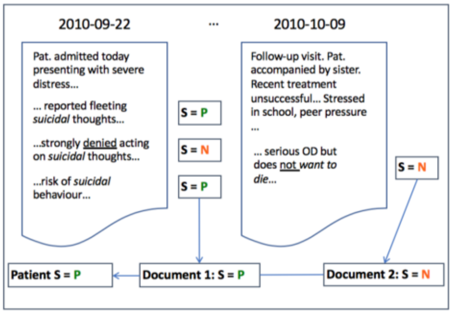

Figure 1 - Example illustrating document and patient-level label heuristics. A patient may have several EHR documents. This example illustrates two fictitious documents from two fictitious dates, where there are multiple suicide-related mentions in the first document (left), 2 positive (P) for suicidality (S) and one negated (N), which leads to a positive document-level label, and only one mention in the second document (right) which is negated, leading to a negated document-level label. Because one document was labelled as positive, the patient-level label is positive (Patient S = P)

Natural Language Processing Approach

We extended our previous approach by relying solely on pyConTextNLP [12] (version 0.6.0.0) that only requires a sentence tokenizer for pre-processing, and on two lexicons: one that defines relevant concepts/mentions (*target terms*, e.g. *suicidal*) and one that defines modifiers (e.g. negations) and so-called termination terms. Termination terms define if the scope of a target modifier should be triggered within a sentence. For instance, in a sentence like *'The patient **denies** wanting to die **but** has had suicidal thoughts.'*, the scope of the negation term

[1] As of 12 April 2018

[2] http://blulab.chpc.utah.edu/content/ehost-extensible-human-oracle-suite-tools

'denies' covers *'wanting to die'* and not '*suicidal'* due to the termination term ***but*** pyConTextNLP allows for matching with character-based regular expressions. We used spaCy[3] (version 2.0.9) for sentence tokenization. The work was performed in a Python 2.7.6 environment. The development and analysis of target and modifier term lexicons was the main focus, and was done in two main phases. For target terms, we reviewed previous similar studies [6,11,15] and experimented empirically with these to identify a good coverage lexicon. As a baseline, we used a minimal target lexicon only containing the term *suicide* and its variants (*suicid*)* to assess the impact of the addition of other target terms. For modifier terms, we used three previously published off-the-shelf lexicons: 1) negation terms used for detecting suicidality in adolescents with ASD ("AMIA2017") [11], 2) negation terms used to detect suicidal ideation ("SREP2018") [6], and 3) a collection of negation terms developed for a variety of clinical use-cases (not specifically suicidality) and languages ("MEDINFO2013") [16]. The off-the-shelf lexicon that yielded best results on the development set was then further iteratively adapted based on findings from manual error analyses. To assess the applicability of the lexicons on the development set, priority was given to optimizing performance on detecting *affirmed* suicidality at both document and patient level. The test set was held aside throughout the development phase.

Table 1 – Reference standard document label distributions

document label	development set	test set
non-relevant	2570	2393
suicidal	136	54
non-suicidal	46	33
uncertain	131	122
total	*2883*	*2602*

Table 2 – Reference standard patient label distributions

patient label	development set	test set
non-relevant	52	66
suicidal	37	26
non-suicidal	11	8
total	*100*	*100*

Evaluation

We measured inter-annotator agreement on a subset with Cohen's κ and accuracy. NLP performance was measured against the annotated reference standard development and test sets with precision (PPV), recall (sensitivity) and f1-score at a document- and patient-level, focusing on results for *affirmed* suicidality. Finally, qualitative error analysis on results was performed to inform future improvements.

Results

Reference Standard and Inter-Annotator Agreement

The overall distribution of the resulting annotations on a document and patient level is presented in Tables 1 and 2. Non-relevant documents are in the majority of both the development and test sets (89% and 92%, respectively). A 4.7% of the documents in the development set were positive for suicidality, and 2.1% in the test set. On a patient level, the distribution of patients labelled as suicidal was: 37% (development set) versus 26% (test set). Inter-annotator agreement on the subset of 100 documents was very high: κ 0.96, 98% accuracy.

Natural Language Processing Results

Adaptations of the target term lexicon involved adding terms like *end his/her life*, *kill him/herself*. Overall performance with these additions resulted in 15-17% f1-score improvement on document-level classification (baseline f1-score results using only *suicid** were 64-72% on the development set (56-68% on the test set). On a patient level, baseline results ranged between 70-79% on the development set (71-79% on the test set), compared to 83-86% using the extended target lexicons on the development set (82-83% on the test set).

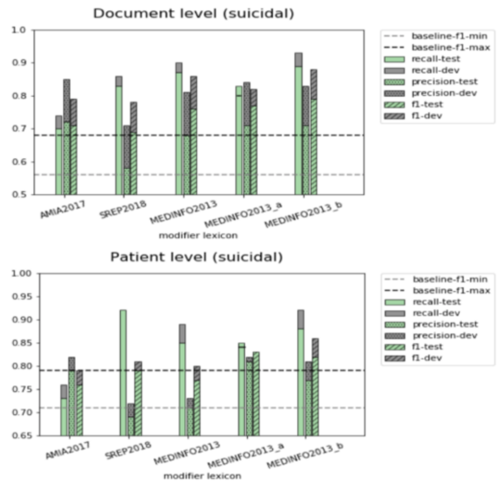

Figure 2 - Results (precision, recall and f1-score) on development (dev, grey) and test (green) sets for document (top) and patient (bottom) level classification (positive for suicidality). Each value on the x-axis represents a modifier lexicon configuration: 1) previous studies: AMIA2017 [11]; SREP2018 [6] and MEDINFO2013 [16] or 2) modifications to the MEDINFO2013 lexicon based on iterative error analysis on the dev set (MEDINFO2013_a and _b). The dashed lines represent f1-scores using the baseline target lexicon (suicid only) on the test set: grey line shows lowest results (min), black line highest results (max)*

Results for the different configurations of modifier lexicons are shown in Figure 2 (document-level on top, patient-level bottom). The precision results on the document-level ranged from 71% (using the SREP2018 lexicon) to 85% (using the AMIA2017 lexicon); recall results ranged from 74% (AMIA2017 lexicon) to 90% (MEDINFO2013 lexicon); f1-score results from 78% (SREP2018) to 86% (MEDINFO2013) without any further adaptation on the development set. The MEDINFO2013 lexicon resulted in best overall off-the-shelf f1-score results and was used for additional adaptation. Errors produced by the MEDINFO2013 lexicon were analysed and iteratively revised based on the development set by adding missing negation terms and revising conjunction terms. We present results for two adapted lexicons: MEDINFO2013_a and _b. In both lexicons, the following negation terms were added: *doesn't*, *doesnt* (sic), *never*, *none*. The main difference between the two adaptations is the list of conjunction/termination terms. In MEDINFO2013_b, 24 conjunction terms, such as *unless* and *whereas,* were added. Furthermore, some conjunction terms that were included in the original MEDINFO2013 lexicon were excluded in both adapted versions, like *patient*, *recent*, *today*.

[3] https://spacy.io/

Document-level results on the blind test set were overall lower compared to the development set (58-72% precision, 70-87% recall and 69-75% f1-score) when using the off-the-shelf modifier lexicons. The adapted lexicons also resulted in lower overall performance compared to the development set, except for recall using MEDINFO2013_a, which improved from 80% to 83%, and MEDINFO2013_b which improved from 88% to 89% (but resulted in lower f1-score).

On a patient level, results for the off-the-shelf lexicons ranged from 72-82% precision, 76-92% recall and 79-81% f1-score on the development set. The adapted lexicons resulted in more balanced, and, on average, higher precision and recall results (81-82%, 84-92% respectively) and slightly higher f1-scores (83-86%). On the blind test set, best f1-score results were obtained with the two adapted lexicons (82% for MEDINFO_a and 83% for MEDINFO_b).

On the test set, the lowest baseline f1-score document-level results were produced by the AMIA2017 lexicon (56%), the highest with the adapted lexicons MEDINFO2013_a and _b (68%). However, on patient-level results the highest baseline result was 79% (MEDINFO2013_b) but results were similar for MEDINFO2013 and MEDINFO2013_a (78%), the lowest results were produced with the AMIA2017 lexicon (71%).

Discussion

We have developed a new comprehensive manually annotated reference standard of EHRs from an adolescent mental health patient cohort. Most documents do not contain any suicide-related information (~90%), only 2-5% documents contain affirmed suicide-related information. On a patient level prevalence is higher (26-37%). This is slightly higher than was found in our previous work [11], but the definition of what constituted affirmed suicide-related information was not identical. Inter-annotator agreement on a small subset was high, indicating that the task is well-defined. Previous studies have shown that using diagnostic codes alone to identify risk of suicidal behaviour is insufficient [4,5,10]. We performed a preliminary assessment of this on our development data and found similar results: only 6 of 37 patients (16.2%) would have been identified as at risk of suicidal behaviour if using information from structured risk assessment forms only.

Our adaptation of pyConTextNLP resulted in similar or higher overall results as compared with our previous NLP approach [11] when applied on the development set, with the added benefit of more efficient processing time (approx. 10 times faster) mainly due to not requiring time-costly pre-processing steps such as syntactic analysis. This shows that a lexical, surface-level approach is sufficient to deal with syntactic phenomena such as negation in these contexts. Off-the-shelf modifier lexicons unsurprisingly yielded varying results – adapting these lexicons based on findings from iterative error analysis on the development set resulted in improvements also on the blind test set.

For our clinical use-case, accurate patient-level classifications are the priority, not document-level. In general, high recall was easier to achieve than high precision. This is probably due to the NLP approach failing to correctly classify certain types of negations (e.g. semi-structured forms, general advice paragraphs) and third party mentions (e.g. relatives), which generate false positives and lower precision, without effecting recall. Precision is of greater importance for this use-case, as false positives are a bigger concern than false negatives.

Extending the target lexicon with additional terms improved document-level results more than patient-level results. This indicates that the term *suicide* and its variants really is the key term overall on patient-level assertions (i.e., it is sufficient for one document to have one affirmed mention of *suicid**, and this term is by far the most common suicide-related term overall). However, if it is important for the end use-case to e.g. identify *when* the first mention of affirmed suicide risk is documented, optimal performance on the document level would be essential. To our knowledge, our study is the first of its kind to extensively analyse the relation between document- and patient-level assertions using this particular type of NLP approach. Our findings could be valuable for further studies in this area, e.g., distributions of suicide-related information in affirmed, as well as negated or other, semantic modifications overall in EHR notes, to inform post-processing heuristics or other data-driven representations of this clinical construct. All lexicons, guidelines and scripts are made available online[4].

Our study has some limitations. The majority of the documents were only annotated by one clinician. An extension of our work would involve double-annotating a larger subset for further inter-annotator agreement analysis. Because the agreement was high on our subsample, we would expect to get high agreement also on a larger sample. We have only focused on applying negation detection on this data and with this NLP approach. We plan to further study the uncertain and unrelated annotations and develop our lexicons to capture these semantic modifiers as well as filtering out third party mentions and references to the past. The off-the-shelf lexicons were not developed specifically for the NLP system we used in this study and in particular were not designed with the same approach for finding the scope of modifiers by defining conjunction terms. We have not studied the individual effects of these terms, which would be an interesting future area to study. In addition, we plan to extend the analysis of lexical terms with other external resources such as the UMLS and SNOMED-CT. Clinical text is often written in non-standard grammatical structures (e.g. long sentences, abbreviations, lack of subjects and/or predicates) which motivates further analysis in the pre-processing steps (sentence and word tokenization) as well as document sections. Furthermore, we have prioritized results on extracting asserted suicide-related information. It might be the case that improving results on the negation detection part would lead to improved results on the assertion detection part, which is something we will study further. Finally, lexicon- and rule-based approaches always suffer from poor generalizability in the sense that spelling variants and new terms need to be added manually. We plan to extend our comparison with other approaches such as Metamap, and also to apply more data-driven approaches for finding relevant terms e.g. by developing word- and sentence embedding models on larger datasets to automatically generate extended representations. However, because of the low prevalence problem on the document level, this approach will need careful study design to ensure that sufficient data samples are used. One approach, similar to Downs et al. [11], would be to filter documents using specific suicide-related target terms and use these documents for generating semantic representations. The applicability of this approach on EHR data from other institutions would also need further investigation.

By adapting and developing a simple NLP tool to identify text relating to affirmed mentions of suicidality, we have been able to identify a cohort of patients who are at higher risk of future adverse outcomes including self-harm. This cohort could be studied for a number of important clinical outcomes. Initially, we are planning a study which will explore socio-demographic,

[4] https://github.com/KCL-Health-NLP/camhs_pycontext_adaptation

clinical and educational risk factors for hospital presentations with self-harm in this high-risk population. This could provide additional information to clinicians on which to base assessments of risk for self-harm and could therefore have an important impact on clinical management.

Conclusions

We have developed a comprehensive manually annotated EHR corpus with rich suicide-related information from a general adolescent cohort. Our adaptation of a rule-based NLP approach and systematic analysis of existing lexicons applied to this data is a contribution to this understudied field for several reasons: we show that simple NLP approaches can have promising results; we confirm previous published results that most suicide-related information is written in free-text; we illustrate the challenges in defining patient-level labels based on mention- and document-level extraction for this low prevalence problem. Our adapted approach could be used in different populations and settings to define cohorts for similar use-cases. We look forward to seeing further developments of NLP approaches for this important problem.

Acknowledgements

Funding support: SV: Swedish Research Council (2015-00359), Marie Skłodowska Curie Actions, Cofund, Project INCA 600398. SE and the project: an MQ Data Science award and The Psychiatry Research Trust (grants held by RD). RD: a Clinician Scientist Fellowship (project e-HOST-IT) from the Health Foundation in partnership with the Academy of Medical Sciences, also funds AB. JD: Medical Research Council (MRC) Clinical Research Training Fellowship (MR/L017105/1) and Psychiatry Research Trust Peggy Pollak Research Fellowship in Developmental Psychiatry.

References

[1] M.K. Nock, J.G. Green, I. Hwang, K.A. McLaughlin, N.A. Sampson, A.M. Zaslavsky, and R.C. Kessler, Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescents Sup, *JAMA Psychiatry* **70** (2013) 300–310.

[2] B. Mars, J. Heron, E.D. Klonsky, P. Moran, R.C. O'Connor, K. Tilling, P. Wilkinson, and D. Gunnell, What distinguishes adolescents with suicidal thoughts from those who have attempted suicide? A population-based birth cohort study, *J Child Psychol Psychiatry* (2018).

[3] C.J. Hawley, B. Littlechild, T. Sivakumaran, H. Sender, T.M. Gale, and K.J. Wilson, Structure and content of risk assessment proformas in mental healthcare, *J Ment Health* **15** (2006) 437–448.

[4] K. Haerian, H. Salmasian, and C. Friedman, Methods for identifying suicide or suicidal ideation in EHRs, in: *AMIA Proc*, 2012: pp. 1244–1253.

[5] H.D. Anderson, W.D. Pace, E. Brandt, R.D. Nielsen, R.R. Allen, A.M. Libby, D.R. West, and R.J. Valuck, Monitoring suicidal patients in primary care using electronic health records, *J Am Board Fam Med* **28** (2015) 65–71.

[6] A.C. Fernandes, R. Dutta, S. Velupillai, J. Sanyal, R. Stewart, and D. Chandran, Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing, *Sci Rep* **8** (2018) 7426.

[7] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, Clinical information extraction applications: A literature review, *J Biomed Inform* **77** (2018) 34–49.

[8] S. Wu, T. Miller, J. Masanz, M. Coarr, S. Halgrim, D. Carrell, and C. Clark, Negation's not solved: generalizability versus optimizability in clinical natural language processing, *PloS One* **9** (2014) e112774.

[9] H. Kaur, S. Sohn, C.-I. Wi, E. Ryu, M.A. Park, K. Bachman, H. Kita, I. Croghan, J.A. Castro-Rodriguez, G.A. Voge, H. Liu, and Y.J. Juhn, Automated chart review utilizing natural language processing algorithm for asthma predictive index, *BMC Pulm Med* **18** (2018) 34.

[10] Q.-Y. Zhong, E.W. Karlson, B. Gelaye, S. Finan, P. Avillach, J.W. Smoller, T. Cai, and M.A. Williams, Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing, *BMC Med Inform Decis Mak* **18** (2018).

[11] J. Downs, S. Velupillai, G. George, R. Holden, M. Kikoler, H. Dean, A. Fernandes, and R. Dutta, Detection of Suicidality in Adolescents with Autism Spectrum Disorders: Developing a Natural Language Processing Approach for Use in Electronic Health Records, *AMIA Proc* **2017** (2017) 641–649.

[12] B.E. Chapman, S. Lee, H.P. Kang, and W.W. Chapman, Document-Level Classification of CT Pulmonary Angiography Reports Based on an Extension of the ConText Algorithm, *J Biomed Inf* **44** (2011).

[13] G. Perera, M. Broadbent, F. Callard, C.-K. Chang, J. Downs, R. Dutta, A. Fernandes, R.D. Hayes, M. Henderson, R. Jackson, A. Jewell, G. Kadra, R. Little, M. Pritchard, H. S., A. Tulloch, and R. Stewart, Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open* **6** (2016).

[14] J. Downs, R. Gilbert, R.D. Hayes, M. Hotopf, and T. Ford, Linking health and education data to plan and evaluate services for children, *Arch Dis Child* **102** (2017) 599.

[15] C. Polling, A. Tulloch, S. Banerjee, S. Cross, R. Dutta, D.M. Wood, P.I. Dargan, and M. Hotopf, Using routine clinical and administrative data to produce a dataset of attendances at Emergency Departments following self-harm., *BMC Emerg Med* **15** (2015) 15.

[16] W. Chapman, D. Hillert, S. Velupillai, M. Kvist, M. Skeppstedt, B. Chapman, M. Conway, M. Tharp, D. Mowery, and L. Deleger, Extending the NegEx Lexicon for Multiple Languages, *Stud Health Technol Inform* **192** (2013) 677–681.

Address for correspondence

Sumithra Velupillai, sumithra.velupillai@kcl.ac.uk